How **data and AI** shape financial services

JOURNAL OF FINANCIAL SERVICES

PUBLISHED BY: **PROJECTIVE GROUP INSTITUTE**

02
2026

Projective INSTITUTE

JOURNAL OF FINANCIAL SERVICES

How **data and AI** shape financial services

Contents

10 **Data mesh as a strategic foundation for AI in financial services: from concept to scaled adoption**
Kristof Meganck, Chief Data Officer, BNP Paribas Fortis
Bart Claeys, Belgium Data Practice Lead, Projective Group

18 **Quantifying sustainability for structured credit products: designing an NLP-driven, hybrid ESG scoring approach for CMBS and structured credit products**
Budha Bhattacharya, Head of Systematic Research, Lombard Odier Investment Management, and Industrial Professor of Banking and Finance, Institute of Finance and Technology, University College London (UCL)
Francesca Medda, Professor of Applied Economics and Finance, and Director, Institute of Finance and Technology, University College London (UCL)

36 **The data integrity imperative: from predictive models to AI agents in financial services**
Richard P. Padbury, Director of Research and Ecosystems, BMO Financial Group
Ryan B. Duffy, U.S. Chief Data and Analytics Officer, BMO Financial Group
Kristin L. Milchanowski, Chief Artificial Intelligence (AI) and Data Officer, BMO Financial Group

48 **Data culture: the silent hero of data value creation**
Gary Paul, Partner and U.K. Data Practice Lead, Projective Group
Keeley Miles, Chief Financial Officer, FCE Bank PLC

56 **Data Governance: the tide that lifts all data initiatives**
Shahina Khan, EMEA Chief Data and Analytics Officer, SMBC Group

66 **The leadership of cyber resilience: from controls to choices**
Martijn Dekker, Professor of Business and Cybersecurity, University of Amsterdam, and CISO, ABN AMRO Bank N.V.

74 **Building data resilience: adapting to modern threats and evolving technologies in a financial markets infrastructure**
Laure Molinier, Director, Group Operational Resilience, Euroclear

84 **Is ESG reporting data fit for purpose?**
Tensie Whelan, Distinguished Professor of Practice Emerita, Business & Society, Stern School of Business, New York University

94 **Securing the future of Payments**
Johan Gerber, Executive Vice President and Head of Security Solutions, Mastercard
Kate Pohl, Global Sales Lead, Projective Group

106 **Big tech in finance: how to ensure a level playing field in a data-driven economy**
Judith Arnal, Senior Research Fellow, Centre for European Policy Studies (CEPS), Elcano Royal Institute, and the European Credit Research Institute (ECRI)

120 **Data risk: a strategic enterprise challenge that requires a holistic board-level offensive**
James Halcomb, Chief Research and Development Officer, EDM Association
Scott Beange, Head of Data Management and Cyber Strategy, Projective Group

130 **Secure data centric decision making**
Vijay Varadharajan, Emeritus Professor, The University of Newcastle, and former Microsoft Chair Professor of Innovation, Macquarie University Sydney, Australia

140 **Beyond the turning point: re-forging financial institutions for the age of agentic AI**
Bill Oates, Cloud and AI Capability Lead, Projective Group

150 **The synthetic empathy framework: merging AI precision with psychological insight for customer retention**
Xuan Zhang, Ph.D. Student in Computer Science, New Jersey Institute of Technology
Guiling Wang, Distinguished Professor of Computer Science, New Jersey Institute of Technology

164 **The case for material ESG data in corporate finance**
Rodrigo Tavares, Invited Full Professor, NOVA School of Business and Economics (Nova SBE)

172 **Tokenization and financial market inefficiencies**
Itai Agur, Senior Economist, Macro-Financial Division, Research Department, International Monetary Fund (IMF)
Germán Villegas-Bauer, Economist, Macro-Financial Division, Research Department, International Monetary Fund (IMF)
Tommaso Mancini-Griffoli, Assistant Director and Chief of Payments, Currencies, and Infrastructure Division, Monetary and Capital Markets Department, International Monetary Fund (IMF)
Maria Soledad Martinez Peria, Assistant Director and Chief of Macro-Financial Division, Research Department, International Monetary Fund (IMF)
Brandon Joel Tan, Economist, Payments, Currencies, and Infrastructure Division, Monetary and Capital Markets Department, International Monetary Fund (IMF)

192 **Efficient underwriting using agentic AI**
Mohammad Asif Ali, Technical Lead, PNC Financial Services Group, Inc.

210 **Agentic AI in banking: between narratives and statistics**
Udo Milkau, Digital Counsellor, Frankfurt, Germany

224 **Economic perspectives on digital infrastructure**
Chris A Richardson, Founder, PerceptionNexus Analytics LLC., and Executive-in-Residence, Department of Finance, University of Miami, Herbert Business School

232 **Wealth management in Switzerland and Liechtenstein 2025: structural trends, performance dynamics, and strategic imperatives**
Christoph Künzle, Senior Lecturer, ZHAW School of Management and Law
Claude Baumann, Executive Chairman, FIN21 AG
Sarina Feldmann, Research Associate, FIN21 AG

242 **How AI is redefining transaction banking: from data to intelligence**
Stephen Peters, Founder, Averus.ai
Guy Moons, Founder, Deeguisa Consulting

Dear reader,

Welcome to the second edition of the Projective Group Institute's Journal of Financial Services.

We publish this edition at a moment when data and Artificial Intelligence are reshaping the financial services sector in ways that extend far beyond technology. They are exposing structural weaknesses, challenging long-established operating models and prompting institutions to confront questions that can no longer be deferred.

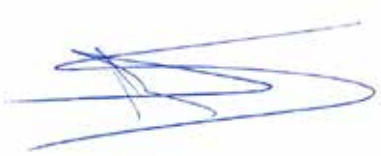
The industry's approach to data has also evolved. The first Chief Data Officer roles appeared in the early 2000s, signalling that data required dedicated senior oversight. Responsibilities then expanded from governance and compliance to broader business enablement. Today, the rise of the Chief Data & Artificial Intelligence Officer (CDAIO) reflects a clear shift. Data and AI are now recognised as one integrated area of accountability within many institutions.

Despite this progression, many organisations continue to face fundamental data challenges. Questions such as what data exists, where it is held, who uses it and whether that use is appropriate often lack clear answers. These issues were once manageable within fragmented systems or through manual controls, but the rapid emergence of AI has changed this. Modern analytical tools, able to interrogate vast datasets instantly, have revealed both the strengths and the fragilities of data foundations with unprecedented clarity.

At the same time, regulatory expectations have increased. The Digital Operational Resilience Act (DORA) strengthens requirements around operational resilience and the governance of critical systems and data, while the EU AI Act sets new standards for data quality, traceability and oversight of AI-enabled processes. Together, they underline that data governance has become a strategic imperative rather than a technical consideration.

This second edition of the journal examines what these shifts mean in practice. Our contributors, drawn from industry, academia and regulatory thought leadership, explore how data and AI are influencing decision-making, risk management and operational resilience. They highlight emerging progress, persistent challenges and the capabilities institutions will need to adopt AI safely, responsibly and at scale.

I would like to thank all contributors for their thoughtful analysis and the expertise reflected in their work. I hope this edition offers clarity at a time when the industry is actively seeking it and I welcome your reflections as we continue this discussion.



Stefan Dierckx, Founder & CEO, Projective Group

How data and AI shape financial services

Data mesh as a strategic foundation for AI in financial services: from concept to scaled adoption

Kristof Meganck,
Chief Data Officer,
BNP Paribas Fortis

Bart Claeys,
Belgium Data Practice Lead,
Projective Group

ABSTRACT

The rapid acceleration of Artificial Intelligence (AI) adoption in financial services has shifted the focus from experimentation to industrialization. Banks are making tangible progress, with AI increasingly embedded in client-facing, operational, and compliance-related processes. At the same time, moving from successful use-cases to sustained, enterprise-wide adoption remains a complex, ongoing journey. As AI capabilities mature, underlying data foundations - rather than algorithms alone - emerge as the decisive factor shaping scalability, trust, and regulatory readiness. Drawing on industry research and practitioner experience from a large European bank, this paper examines how challenges related to data quality, ownership, lineage, accessibility, and governance intensify as AI moves into production. It highlights why traditional centralized and hybrid data models struggle to support this next phase, and how data mesh principles offer a pragmatic operating model shift. By treating data as a product, aligning ownership with business domains, and embedding federated governance, data mesh helps create the structural conditions required for scalable and compliant AI. The central conclusion is that progress in AI adoption is inseparable from progress in data maturity, making data management a strategic concern for financial institutions navigating the next phase of AI-driven transformation.

1. Introduction

The rapid evolution of Artificial Intelligence (AI), and more recently Generative AI (GenAI), has triggered an unprecedented wave of interest and investment across industries. Global investment in AI is increasing at double-digit rates, surpassing U.S.$250 billion in 2024 and expected to grow substantially in the coming years. Financial services remain one of the most active sectors, driven by intense competitive pressures, regulatory scrutiny, and rising customer expectations. Boards, executive committees, and regulators alike are now grappling with the same fundamental question: how can AI be harnessed to create sustainable value without undermining trust, compliance, or operational resilience?

These figures reflect both optimism and urgency. Billions are being allocated to AI initiatives each year, with banks among the most active investors. Yet, despite this momentum, a striking pattern is emerging. While experimentation is widespread and proof-of-concepts abound, fewer than one in five organizations report having successfully scaled AI across multiple business functions, underscoring a persistent gap between experimentation and industrialization.

This gap between ambition and realization is often attributed to model risk, ethical concerns, or regulatory uncertainty. These factors matter, but they are rarely the decisive constraint: the ability to manage data at scale, across domains, and with sufficient trust to support industrialized AI is. In practice, many AI initiatives fail not because the algorithms are insufficiently powerful, but because the underlying data landscape is fragmented, opaque, and misaligned with how banks operate.

Banks are, by nature, highly data-intensive organizations. Decades of product innovation, regulatory layering, mergers, and system upgrades have resulted in complex data ecosystems spanning hundreds of platforms and countless interfaces. These environments were not designed with AI consumption in mind. As a result, efforts to deploy AI models at scale frequently collide with issues of data quality, ownership, lineage, and accountability - challenges that intensify as regulatory scrutiny increases.

It is in this context that data management must be reconsidered, not as a purely technical discipline, but as an operating model decision. Traditional, centralized approaches to data governance and platform design are increasingly struggling to keep pace with the speed, autonomy, and contextual understanding required for modern AI use-cases. For many banks, this realization is driving a shift toward data mesh principles: a model that emphasizes domain ownership, data as a product, and federated governance.

Data management has become the decisive factor in making AI production-ready and scalable in banking. Drawing on industry research and practical experience, it argues that data mesh is not an architectural trend, but a pragmatic response to the structural realities of large financial institutions - and a necessary foundation for turning AI investment into lasting business value.

2. AI value in banking is well understood - scaling is not - yet!

At **BNP Paribas Fortis**, the potential of AI has already moved beyond experimentation in several concrete areas of the organization. Today, approximately one hundred AI use-cases are operating in production within the Belgian entity, while close to nine hundred have been deployed across the wider Group. Many of these initiatives are concentrated in client-facing and compliance-adjacent activities, where productivity gains are tangible and operational risks can be tightly managed.

In BNP's front office functions, for example, conversational AI solutions (BNP's chatbot called "Samy") are being used to better capture the intent behind client questions, moving beyond rigid keyword-based interactions and improving both response quality and client experience. Speech-to-text and natural language processing (NLP) technologies are also applied to advisory and sales conversations, supporting more systematic regulatory (MiFid) controls while significantly reducing the manual effort required for post-interaction documentation. Automated generation of call summaries and meeting minutes has begun to relieve front-office teams of administrative burden, allowing them to focus more fully on client engagement.

AI is also increasingly embedded in compliance processes such as know-your-customer (KYC) recertification. By assisting in the analysis of transactional behavior and client profiles, AI-supported

approaches help accelerate recertification cycles, improve consistency across reviews, and reduce preparation time for both business and control functions.

Taken together, these use-cases represent a meaningful source of value creation. At BNP Paribas Fortis, AI initiatives translate into tens of millions of euros in recurrent value, with the cumulative impact reaching several hundred million euros at Group level. Beyond direct financial benefits, the productivity gains associated with deployed AI solutions are comparable to the workload of several hundred full-time equivalents, enabling teams to absorb additional activities and redirect capacity toward higher-value work.

At the same time, this experience highlights an important structural pattern. Even where AI delivers measurable value and operates at scale, successful deployments remain largely confined to well-defined processes and clearly bounded data domains. Extending these gains horizontally across the enterprise proves significantly more complex, and it is at this point that data foundations, operating models, and governance structures become decisive.

3. Data challenges intensify as AI moves toward production

AI systems are fundamentally dependent on data - not just in volume, but in structure, consistency, and context. Industry studies consistently show that most of the effort in AI initiatives - often as much as 70–80% - is consumed by data preparation and remediation rather than model development. While early experimentation can often tolerate imperfect datasets, production environments cannot. As models are operationalized, expectations around explainability, reproducibility, and auditability increase dramatically.

Several data-related challenges tend to surface at this stage. First, data quality and consistency become critical. AI models trained on incomplete, outdated, or inconsistent data produce unreliable outputs, eroding trust among users and stakeholders. In regulated environments, such issues quickly escalate from technical concerns to compliance risks.

Second, data lineage and transparency are essential. Banks must be able to explain where data originates from, how it has been transformed, and how it is used in decision-making processes. This is particularly important for AI systems influencing credit decisions, fraud detection, or customer interactions.

Third, data accessibility and timeliness often fall short of AI requirements. Even when data exists, it may be locked within systems that are difficult to integrate, subject to restrictive access controls, or updated too infrequently to support real-time use cases.

Finally, accountability for data is frequently unclear. In many organizations, responsibility for data quality and fitness for purpose is diffuse, leading to gaps between business expectations and technical execution.

These challenges are not new. What AI changes is their impact. Where traditional analytics might degrade gracefully under imperfect conditions, AI systems tend to amplify data deficiencies, making them visible - and costly - at scale.

4. Legacy systems and the limits of centralized data models

Most banks operate within technology landscapes shaped by decades of incremental change. Core banking platforms, risk systems, customer databases, and regulatory reporting engines have evolved independently, often optimized for stability rather than interoperability.

To manage this complexity, many institutions have historically pursued centralized data architectures: enterprise data warehouses, data lakes, or shared analytics platforms designed to consolidate information and enforce uniform governance. While these approaches have delivered value, they also exhibit structural limitations when confronted with modern AI demands.

Centralized models struggle to scale organizationally. As data volumes and use-cases grow, bottlenecks emerge around ingestion, transformation, and governance. Central teams become overloaded, while business domains feel disconnected from the data they depend on.

Moreover, centralized control often obscures domain context. Data is extracted from operational systems and transformed without sufficient understanding of its business meaning, leading to misinterpretation and misuse downstream. For AI systems, which rely heavily on contextual signals, this loss of meaning can be particularly damaging.

Finally, centralized approaches tend to be slow to adapt. AI use-cases evolve rapidly, requiring frequent adjustments to data pipelines and features. In rigid architectures, each change introduces coordination overhead, increasing time-to-market and reducing experimentation velocity.

These structural limitations are not theoretical; at BNP Paribas Fortis, many of the challenges commonly associated with large, legacy data environments are encountered in day-to-day operations. Transparency remains a persistent issue: teams are often unaware of what data already exists, and potential value remains untapped due to limited visibility and discoverability. Even when relevant data is known to exist, accessibility can be highly time-consuming, with lengthy discovery-to-access cycles and unclear points of contact slowing delivery and frustrating business users.

Governance and tooling further compound these challenges. A disconnect frequently emerges between business and technical teams, reinforced by fragmented toolsets and the need for specialized skills that limit broader adoption. At the same time, questions of ownership and accountability remain central. Strong expectations around correct data usage coexist with the absence of a single source of truth and consistent definitions, increasing the effort required to establish trust in data. Compliance considerations add another layer of complexity, as uncertainty around permissible data sharing and downstream usage can inhibit reuse, even where value potential is clear.

These issues are reflected in the bank's current operating model. Like many large financial institutions, BNP Paribas Fortis operates a hybrid data landscape, combining a centralized data platform primarily serving business intelligence and reporting needs with more localized data stores embedded within specific business domains. While this approach has delivered stability and supported historical reporting requirements, it has also introduced structural constraints. Central teams can become bottlenecks, extending time-to-market; data is duplicated through point-to-point integrations, eroding lineage and quality; and information often traverses multiple layers before it can be used to create value. Investments, moreover, are frequently optimized for single business purposes, resulting in platforms that must be built, maintained, and governed in isolation.

Taken together, this environment illustrates why traditional centralized or hybrid data models increasingly struggle to support the speed, reuse, and trust required for scalable AI initiatives.

These constraints have led many banks to question whether traditional data management models are compatible with the pace and complexity of AI-driven transformation.

5. Reframing data management: from platforms to products

Addressing these challenges requires more than incremental optimization. It calls for a shift in how data is conceptualized and governed within the organization.

Increasingly, banks are exploring data as-a-product thinking, where datasets are treated as first-class assets designed to serve specific consumers - including AI models - with clear quality expectations, ownership, and lifecycle management. This reframing moves data management closer to the business domains that generate and understand the data, rather than isolating it within central teams.

Under this paradigm, data quality is not enforced solely through centralized controls, but through domain accountability, as mentioned before. Those closest to the source are best positioned to understand its semantics, constraints, and appropriate use. This alignment becomes particularly powerful when AI models consume data across multiple domains, each contributing specialized signals.

Importantly, treating data as a product does not imply abandoning governance or standards. On the contrary, it requires explicit contracts, shared definitions, and common interoperability principles. What changes is the focus of responsibility and the balance between autonomy and coordination.

At BNP Paribas Fortis, these considerations led to a deliberate shift in data management approach towards the end of 2024. In response to persistent

challenges around transparency, ownership, reuse, and time-to-market, the bank chose to adopt "data mesh" principles, explicitly moving accountability for data closer to the business tribes. In this model, responsibility for data products is assigned to business-aligned domains - covering both products and channels - rather than being concentrated within a central function.

Practically, this shift reframes the role of the central data organization. Rather than acting as the primary producer of data assets, it focuses on enabling domains by providing a self-service data platform and a set of standardized capabilities that allow teams to build, expose, and consume data products in a consistent and governed manner. This includes shared infrastructure for data access, interoperability, and governance, designed to lower the barrier for domain teams to make their data available for analytics and AI use-cases.

This change was driven by a clear assessment of trade-offs. Without redistributing ownership and reducing dependence on centralized delivery, the organization risked slowing the rollout and scaling of AI initiatives, limiting its ability to respond to innovation, and perpetuating data duplication, reconciliation effort, and legacy integration patterns. By contrast, treating data as a product within domains creates the conditions for reuse and synergy across the organization, while maintaining the control and traceability required in a regulated environment.

In this sense, the move toward a data mesh-inspired operating model is less about adopting a new architectural paradigm than about aligning data ownership, incentives, and capabilities with where value is created. Fueling AI adoption, accelerating innovation, and realizing synergies across domains become outcomes of this alignment, rather than objectives pursued through additional layers of centralization.

6. Data mesh as an enabler for scalable AI

While still unevenly adopted, data mesh principles are increasingly referenced in large enterprises as a response to scaling challenges in data-intensive and regulated environments. Not as another architectural trend, but as a fundamentally different way of aligning data ownership, governance, and delivery with business value. At its core, it combines four principles: domain-oriented data ownership, data-as-a-product, self-serve data infrastructure, and federated computational governance.

For AI initiatives, this model offers several advantages.

First, domain-oriented ownership ensures that data feeding AI models is grounded in business reality. Features and labels are defined by those who understand their implications, reducing the risk of semantic drift.

Second, data products provide clear interfaces and quality guarantees. AI teams can consume data with greater confidence, knowing its provenance, refresh frequency, and limitations.

Third, self-serve infrastructure lowers the barrier to experimentation while maintaining consistency. Central teams provide reusable capabilities - ingestion frameworks, catalogues, lineage, permissions, and quality monitoring - enabling domains to deliver without reinventing infrastructure.

Finally, federated governance balances regulatory compliance with agility. Common policies are enforced through automation rather than manual oversight, allowing AI systems to scale without sacrificing control.

At BNP Paribas Fortis, the application of data mesh principles is taking shape through a growing portfolio of domain-owned data products, supported by explicit governance and architectural choices. Early proof-of-concept data products have focused on foundational datasets with high reuse potential, such as current account reference data and the relationships between clients and the financial products they hold. These data products are designed to serve multiple consumers - including analytics, reporting, and AI use-cases - through well-defined interfaces and quality expectations.

Building on these initial examples, the roadmap for the coming phases extends into areas where data reuse and trust are particularly critical. Planned data products include those supporting KYC and AML processes, human resources analytics, insurance-related data sourced through partnerships as well as transaction datasets required for judicial enquiries and regulatory obligations such as subject access requests. Together, these examples illustrate how data products are prioritized not only based on immediate demand, but also on their potential to serve multiple regulatory, operational, and AI-driven use-cases over time.

Governance plays a central role in ensuring that this decentralization does not erode control. Before a data product is approved for build and exposure, it is reviewed by a dedicated Data Product Board. This body assesses the underlying business need, expected reusability, financial rationale, and the contractual elements associated with the product, including service levels and usage conditions. In this way, governance shifts towards upfront enablement, embedding accountability and trust into the lifecycle of each data product.

These principles are reinforced through deliberate architectural decisions. For each business domain, an operational data store is established to support operational and near-real-time use-cases, while also serving as a structured input layer for analytical data products. This approach helps offload legacy core systems, supports different latency requirements, and creates a clear separation between operational consumption and analytical reuse. In cases where real-time access is essential, data can be consumed directly from domain-level operational stores, while analytical and batch-oriented use-cases rely on curated data products.

Taken together, these examples demonstrate how data mesh principles translate into concrete design choices. Domain ownership, product thinking, federated governance, and architecture are aligned to support AI use-cases that are both scalable and compliant - without reintroducing the bottlenecks of fully centralized data models.

7. Making AI production-ready: operational implications

Adopting a data mesh-inspired approach has implications beyond architecture. It reshapes roles, incentives, and ways of working.

Experience increasingly shows that the most difficult aspects of scaling AI are not technical, but organizational. At BNP Paribas Fortis, the shift toward a data mesh-inspired operating model has reinforced this insight. While technology and architecture are necessary enablers, the primary challenge lies in changing how teams perceive ownership, accountability, and value creation around data.

For business-aligned domains to take on responsibility for data products, the rationale must be clear. Teams are asked to allocate part of their budget and capacity to building and maintaining data products - often in competition with more immediately visible initiatives such as launching new banking products or channel features. Similarly, local reporting and analytics teams must be convinced that shared data products will ultimately enable faster and more reliable answers to business questions than the highly flexible, but fragmented, “shadow IT” solutions many have relied on to date.

Recognizing these dynamics, a dedicated change management track was established as an integral part of the initiative. Its focus is not only on communication, but on actively identifying sources of resistance across the organization and addressing them through engagement, transparency, and concrete examples of emerging value. This emphasis reflects an understanding that data mesh is fundamentally a transformation of ways of working, rather than a purely technical rollout.

Senior leadership commitment plays a critical role in sustaining this transformation. In the early phases, the financial business case for data products is rarely visible in isolation; value materializes only once products can be combined, reused, and aggregated across domains. Maintaining momentum, therefore, requires consistent C-level sponsorship and a willingness to enforce agreed governance and architectural principles over time, even when short-term pressures compete for attention.

At BNP Paribas Fortis, this journey is still in its early stages. While the architectural, technical, governance, and operating model foundations have been carefully prepared, the first data products are only now being built. As with any ambitious transformation, obstacles and adjustments are expected in the months and years ahead. What matters is not the absence of friction, but the ability to learn, adapt, and remain disciplined as the organization moves from intent to sustained execution.

8. Conclusion: data maturity as a prerequisite for AI maturity

AI will continue to evolve, and new techniques will undoubtedly expand what is technically possible. For banks, however, sustainable advantage will not come from chasing the latest model, but from building the conditions under which AI can be safely and effectively scaled.

Data management sits at the center of this challenge. Institutions that invest in clear ownership, high-quality data products, and governance models aligned with their operational realities will be better positioned to translate AI ambition into durable outcomes.

Data mesh offers a pragmatic blueprint for achieving this. It does not replace platforms, governance, or central capabilities - it reorganizes them. It distributes responsibility without losing control. It allows transformation to proceed incrementally rather than through risky, multi-year replacements of legacy systems.

For institutions facing rising expectations from customers, competitors, and regulators, data mesh provides the structural backbone for AI to realize its full potential. It is not an end state, but a strategic path toward a more agile, trusted, and value-driven data ecosystem.

How data and AI shape financial services

Quantifying sustainability for structured credit products: designing an NLP-driven, **hybrid ESG scoring approach** for CMBS and structured credit products

Budha Bhattacharya,
Head of Systematic Research, Lombard Odier Investment Management, and Industrial Professor of Banking and Finance, Institute of Finance and Technology, University College London (UCL)

Francesca Medda,
Professor of Applied Economics and Finance, and Director, Institute of Finance and Technology, University College London (UCL)

ABSTRACT

Environmental, social, and governance (ESG) integration in structured credit is becoming a practical necessity, yet most established ESG scoring methods were designed for single-issuer instruments and often break down when applied to securitizations. This article proposes a hybrid, natural language processing (NLP)-driven ESG scoring methodology tailored to structured credit - particularly commercial mortgage backed securities (CMBS) - where risk is distributed across multiple entities (originators, underwriters, servicers, trustees, and borrowers) and where critical evidence is embedded in large, unstructured transaction documents. The approach combines (i) advanced NLP for entity and role extraction from CMBS documentation, (ii) structured ESG metrics from reputable quantitative datasets, and (iii) news-based sentiment and controversy signals to improve responsiveness and reduce overreliance on self-disclosed information. The methodology also introduces hierarchical (role-based) weighting to reflect the fact that different parties exert varying influence over ESG outcomes and risk transmission. A Freddie Mac Green Bond CMBS case study illustrates how the approach can be applied in practice using document extraction, identifier linkage, and illustrative (hypothetical) quantitative and sentiment scores to demonstrate score decomposition and interpretability.

1. **Data and materials disclaimer**: any screenshots, figures, or excerpts from third-party materials (including, but not limited to, publicly available documents or websites of Freddie Mac and other market participants) are reproduced solely for illustrative and explanatory purposes in the context of academic discussion and analysis. Such materials are used to demonstrate methodological concepts and do not imply endorsement, affiliation, or approval by the respective rights holders. All trademarks, service marks, and copyrighted materials remain the property of their respective owners. Where applicable, the source of such materials is acknowledged. No claim is made regarding the accuracy, completeness, or current validity of third-party content beyond its illustrative use in this article.

1. Introduction: why ESG breaks down when it comes to structured credit products

ESG scoring rose quickly in equities and corporate credit because those markets have a relatively straightforward unit of analysis: a single operating issuer with a defined disclosure boundary. In this article, we argue that this framing does not transfer cleanly to structured credit. In CMBS and other securitizations, the investor is exposed to a structure - often backed by pools of assets - where risk is distributed across different parties and contractual mechanisms. The result is that applying a single "issuer ESG score" can obscure where the real sustainability risks sit and who can influence outcomes.

More broadly, ESG scores and ratings are known to diverge significantly across providers due to differences in methodologies, scope, and data inputs, which already limits their usefulness for investment decision-making even before the additional complexity of structured credit instruments is considered [Berg et al. (2022), Kotsantonis and Serafeim (2019)].

Firstly, the **entity structure is multi-layered**. Structured credit products involve multiple roles - originators, servicers, underwriters, trustees, borrowers, and investors - each contributing differently to ESG risk exposure. In a CMBS transaction, for example, originators shape underwriting quality and the initial ESG characteristics of underlying loans, servicers manage the ongoing administration of cash flows and compliance, and trustees oversee compliance and administrative functions. Because these roles are distinct, ESG risks can be granular (e.g., tied to a specific servicer's governance practices) and also systemic (e.g., driven by market-wide data gaps and inconsistent disclosure).

Secondly, **incentives can be misaligned** across the structure. Stakeholders may have different priorities - some may focus on loan volume or short-term financial outcomes, while others lack enforcement mechanisms to drive ESG improvements. This creates a practical challenge: an ESG issue at one node in the structure can propagate through the deal in ways that traditional ESG ratings - built for single issuers - do not model. We also argue that waterfall and payment-hierarchy mechanics in structured products can complicate how ESG-linked risks ultimately affect investor cash flows.

Thirdly, the **relevant evidence is often unstructured and difficult to extract at scale**. A key reason ESG breaks down in structured products is operational: much of what investors need to know is embedded in dense documentation - prospectuses, agreements, and reporting packages. This article frames CMBS documentation as voluminous and unstructured, making systematic extraction of ESG insights difficult using traditional methods. This creates a structural bottleneck: investors either simplify (and risk missing key signals) or rely on manual analysis that does not scale.

Fourthly, **data availability is uneven, especially for private entities**. A number of parties in structured finance operate in private markets with limited disclosure obligations. This reduces data coverage and forces reliance on proxies or indirect metrics. In practice, this creates two failure modes: (i) missing data for essential entities, and (ii) mis-mapping entities (e.g., confusing a subsidiary with a parent), which can introduce false precision.

Fifth, **both "purely qualitative" and "purely quantitative" ESG approaches have gaps in this setting**. We review qualitative (often expert-judgement driven and subject to inconsistency across providers) and quantitative frameworks (more replicable but sometimes too rigid or slow to capture fast-moving controversies) and suggest that these weaknesses become more pronounced in structured products because their structures require role-specific attribution and because risks evolve over time as news, controversies, and regulatory developments emerge.

Finally, **greenwashing and disclosure bias are amplified risks when ESG is measured solely from self-reported data**. In this article, we discuss how traditional ESG approaches often rely heavily on corporate disclosures, which can be incomplete or biased. This matters in structured finance because the investor may rely on limited information about private counterparties and on transaction-level disclosures that vary across deals. This article, therefore, motivates the use of external signals - particularly news-based sentiment and controversy monitoring - as a necessary complement to structured ESG metrics.

In short, this article's diagnosis is that structured credit ESG fails when treated as a single-issuer rating problem. In securitizations, ESG is fundamentally an attribution and evidence problem: the investor needs to know which entities matter, what roles they play, what the documents say, and what is changing over time.

Table 1 summarizes the key conceptual differences between issuer-level ESG frameworks and the requirements of structured credit products, as discussed throughout the article. It highlights why role-based attribution, document-level evidence, and ongoing monitoring are necessary to produce explainable and decision-useful ESG assessments for securitized instruments.

2. What makes CMBS/ structured products different from single-issuer credit

Structured and securitized products pose distinct challenges for ESG integration, as they involve multiple counterparties, layered contractual arrangements, and asset-level risk exposures that are not well captured by issuer-centric ESG frameworks [PRI (2021)].

In this article, we use CMBS as the anchor structured product to explain why conventional ESG approaches do not transfer cleanly from corporate bonds. In corporate credit, a bondholder's risk is

Table 1: Where traditional issuer-centric ESG approaches break down and what structured credit ESG requires.

Dimension	Traditional issuer-centric ESG	What structured credit requires
Unit of analysis	Single operating issuer	Transaction-level view spanning multiple entities
Entity structure	One balance sheet, one governance structure	Network of originators, underwriters, servicers, trustees, and borrowers
Attribution of ESG risk	Aggregated at the issuer level	Role-based attribution reflecting influence and control
Source of ESG evidence	Corporate disclosures and reports	Legal transaction documents (prospectus, servicing agreements, and reports)
Treatment of private entities	Often excluded or proxied	Explicit handling of private counterparties and SPVs
Data structure	Primarily structured datasets	A combination of structured data and unstructured text
Responsiveness	Periodic, slow-moving updates	Continuous monitoring with news and controversy signals
Handling of greenwashing risk	Reliance on self-reported disclosures	External validation via sentiment and controversy analysis
Transparency and explainability	Often opaque composite scores	Decomposable scores with entity- and role-level drivers
Mapping confidence	Implicit or assumed	Explicit identifier linkage and confidence scoring

primarily tied to an issuer's operations, governance, and disclosures, whereas in a CMBS, the investor is exposed to a securitization backed by a pool of commercial mortgages and structured by contractual arrangements among multiple parties. This changes the ESG problem in several specific ways.

- **The unit of risk is a structure, not an operating company**: in many structured deals, the legal issuer is a trust or special purpose vehicle (SPV) created for the transaction. The operational ESG footprint investors care about is, therefore, not the "issuer" as such but the combination of underlying collateral characteristics and the behavior and governance of the deal's counterparties. We argue, therefore, that sustainability assessment must shift from "issuer scoring" to "structure and role scoring."
- **CMBS involves role-based influence over ESG outcome**s: in CMBS, ESG risk is distributed across originators, underwriters, servicers, trustees, and borrowers, and cannot be represented by a single issuer-level score. A structured credit product often has a number of underlying actors. Originators influence underwriting and the ESG characteristics embedded at origination; underwriters affect disclosure quality and how ESG risks are communicated at issuance; servicers influence ongoing administration, compliance, and risk management; and trustees perform oversight and administrative functions and typically exert less direct influence over ESG outcomes. This unequal influence is why the methodology introduces hierarchical (role-based) weighting rather than treating all parties equally. These challenges are also reflected in rating agency commentary, which emphasizes the complexity of incorporating ESG considerations into CMBS analysis and ongoing surveillance of structured credit transactions [DBRS Morningstar (2023)].
- **CMBS evidence is embedded in transaction documentation**: unlike corporate credit, where ESG data can be sourced from standard corporate reporting and widely used datasets, structured products require investors to engage with prospectuses and other dense documentation. In CMBS, the sustainability profile of the transaction depends not only on the behavior of transaction counterparties, but also on the characteristics and performance of the underlying commercial mortgage collateral. In this article, we treat this as central: extracting ESG-relevant information from unstructured documents is not an optional add-on, it is a practical prerequisite for a scalable approach in this asset class. Prior research indicates that climate-related risks can translate into measurable impacts on commercial mortgage performance and, by extension, the dynamics of securitization structures [Holtermans (2024)].
- **Data coverage and transparency differ materially from public issuer markets**: in this article, we stress that some entities in CMBS structures may be private or have limited disclosure obligations, complicating ESG measurement. Even where structured ESG datasets exist, the investor still has to resolve a mapping problem: the entity name in documents must be matched to a stable identifier, and often to a parent entity with better coverage.
- **CMBS is inherently dynamic and requires monitoring**: in this article, we frame structured products as instruments whose risk profile can evolve: the relevance of controversies, the visibility of ESG risks, and the interpretation of counterparties' practices can change over time. A one-time ESG label at issuance is, therefore, not sufficient for ongoing risk management; an investor usable framework needs a mechanism for updates and responsiveness.
- **The investor's question is "what is exposed, who controls it, and what is changing?"**: this article's CMBS framing implies a different due diligence posture than corporate bonds. Instead of only asking whether an issuer has good ESG performance, the investor must ask: which parties and roles create the sustainability risk, which data sources support the assessment, what contractual or governance mechanisms influence behavior, and how can the assessment be updated as new information arrives? This is why this article's methodology is deliberately hybrid and evidence-based: it is designed around the realities of CMBS structures, the limits of disclosure, and the need for ongoing interpretability.

The schematic Figure 1 illustrates the key parties involved in a CMBS transaction and how ESG-related risks originate and propagate through the structure. ESG risk is introduced through underwriting standards, disclosure practices, and underlying collateral characteristics, and is subsequently

managed through servicing and overseen by trustees and administrators. Figure 1 highlights why ESG assessment in structured credit requires role-based attribution rather than issuer-centric scoring.

3. Proposed hybrid framework

The central contribution of this article is a hybrid ESG scoring framework for structured credit that combines three components: structured quantitative ESG metrics, NLP-driven extraction from unstructured deal documents, and news-based sentiment/controversy signals. Reliance on structured ESG data alone can be problematic, as such metrics are often disclosure-driven, backward-looking, and insufficiently granular for complex financial instruments [Kotsantonis and Serafeim (2019)]. This article motivates this as a response to the limitations of existing approaches: qualitative ratings can be subjective and inconsistent; purely quantitative approaches can oversimplify relationships and lag real-time developments; and both often fail to capture the role-based nature of structured products.

3.1 The design intent

The framework aims to preserve the rigor and replicability of structured data while adding two features that structured credit investors need in practice:

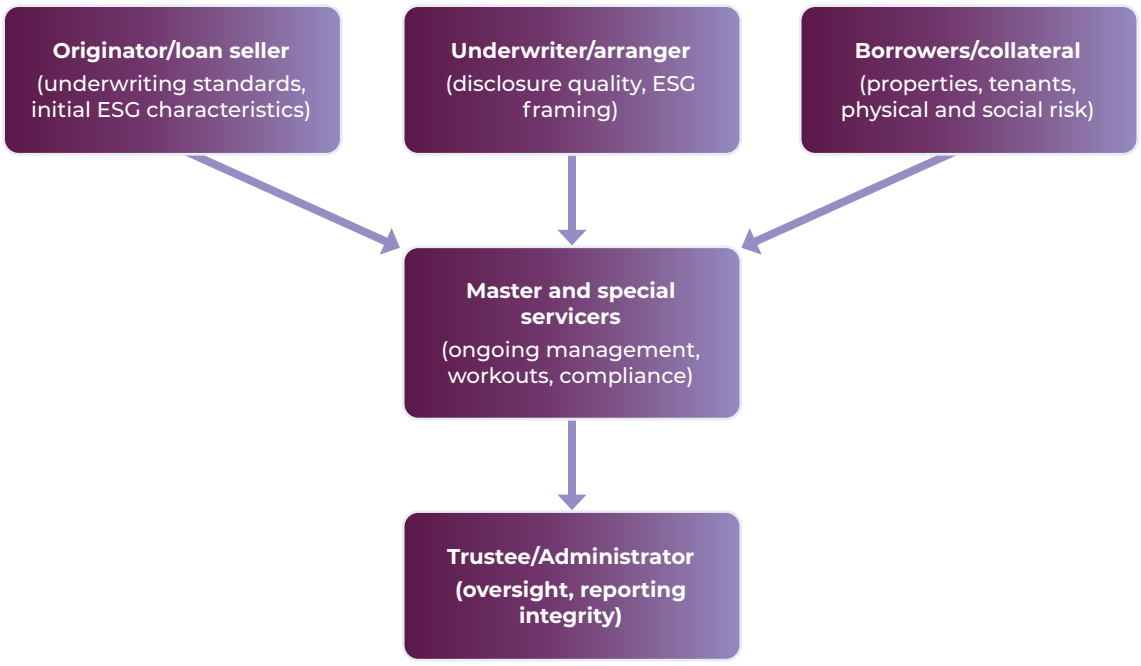
- **Role-specific attribution**: identifying which entities are involved in the deal and what roles they play, so the ESG assessment aligns with influence and risk transmission.
- **Dynamic responsiveness**: incorporating real-time controversy and sentiment signals so the score can reflect emerging developments rather than waiting for periodic dataset updates.

This article also emphasizes transparency: a score should be explainable, with traceability into the underlying components, rather than being an opaque composite.

3.1.1 Component 1: structured ESG data (the "quantitative baseline")

We frame structured ESG datasets (e.g., from major providers) as essential because they provide consistent Environmental, Social, and Governance metrics that can be benchmarked and compared across entities. These metrics help address the

Figure 1: CMBS conceptual ecosystem and propagation of ESG risk



inconsistency problem found in purely qualitative assessments. We also note, however, that structured ESG data is often built on corporate disclosures and can, therefore, be vulnerable to disclosure bias and greenwashing. In structured credit, it also may not cover private entities well. This is why structured ESG data is used as a baseline rather than as a complete solution.

3.1.2 Component 2: document intelligence via NLP (the "transaction evidence layer")

Recent research demonstrates that NLP can be used effectively to extract and structure ESG-relevant information from large volumes of financial and corporate disclosures [Schimanski et al. 2024)]. We argue in this article that unstructured CMBS documentation contains critical transaction-level information that traditional ESG scoring often ignores: which entities participate, what roles they play, and how responsibilities are allocated. The methodology, therefore, uses NLP to extract entities and classify their roles from CMBS documents systematically. This extraction enables hierarchical weighting and improves interpretability.

In this article, we propose a combined NLP approach, using:

- **Zero-shot classification** to label entities/roles without requiring large training datasets (valuable in niche domains like structured finance).
- **Transformer models (BERT/RoBERTa)** to capture context in dense financial text and improve the accuracy of entity recognition and role classification.
- **Large language models (LLMs)** to support nuanced contextual interpretation and role classification, where dense text makes relationships harder to interpret.

Importantly, in this article, we frame this combination as pragmatic: each model type contributes distinct strengths, and weights or confidence levels can be determined based on performance benchmarks.

3.1.3 Component 3: news-based sentiment and controversy (the "synthetic intelligence overlay")

To address greenwashing risk and to make ESG assessment responsive to emerging events, this article integrates news-based sentiment analysis with controversy monitoring. It references prior work and commercial approaches that use NLP to quantify ESG signals from news flows. It acknowledges limitations such as media bias and noise but argues that sentiment signals add valuable timeliness and can complement structured data.

News-based sentiment analysis has been shown to provide timely signals of ESG-related risks and controversies, complementing structured ESG metrics that update less frequently [Du et al. (2024)].

Conceptually, this article proposes that structured ESG metrics should carry the majority of weight, while sentiment/controversy should contribute a smaller but meaningful share. In the literature review, this article discusses an approximately **75% structured/25% sentiment**-style balance as an empirically motivated range (with the sentiment component intended to be complementary rather than dominant).

3.2 Why "hybrid" matters specifically for structured credit

This article's hybrid approach is not just "more data", it is designed to solve structured credit's core ESG problems:

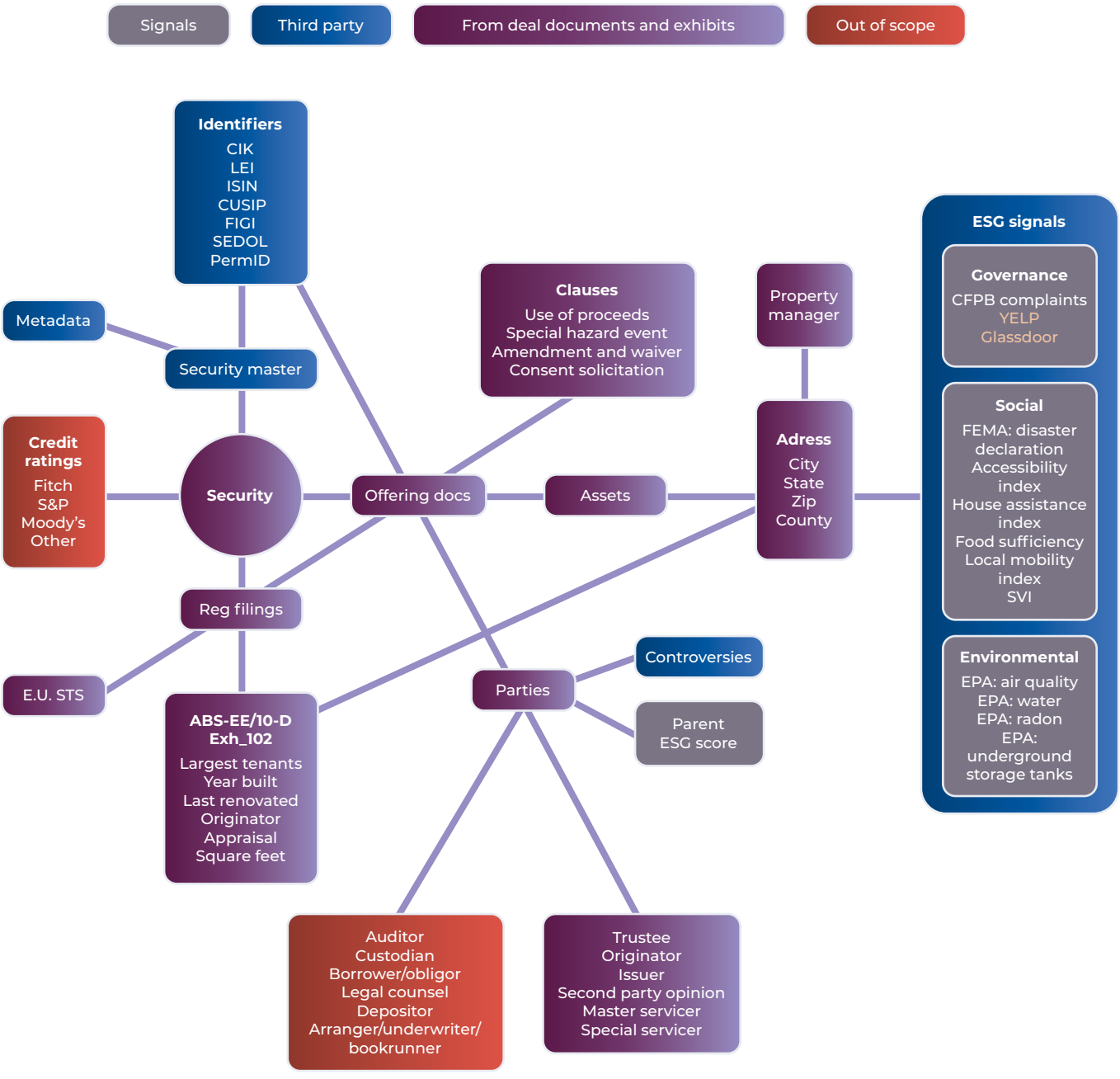
- **Attribution**: document intelligence identifies the entity network and roles, enabling role-based weighting.
- **Coverage and comparability**: structured ESG data provides consistent baseline metrics where coverage exists.
- **Timeliness**: sentiment/controversy captures real-time developments and reduces overreliance on self-disclosure.

The result is positioned as a more nuanced, investor-usable ESG score that is explainable and adaptable.

Figure 2 explicitly illustrates the end-to-end hybrid NLP-ESG scoring architecture. This schematic illustrates the conceptual workflow for deriving a transaction-level ESG score for structured credit products. Unstructured deal documentation is processed through an NLP pipeline to extract entities and classify their roles, which are then weighted hierarchically and combined with structured ESG metrics and news-based sentiment signals to produce a final hybrid ESG score. The figure is conceptual and intended for explanatory purposes.

Figure 2: End-to-end hybrid NLP-ESG scoring architecture

Figure 3: CMBS conceptual data map



4. Data and document methodology

This article frames the practical challenge in structured credit as a pipeline problem: how to consistently obtain, interpret, and integrate information from multiple sources - especially unstructured documents - without relying on deal-by-deal manual work.

4.1 Data inputs: structured and unstructured

The methodology integrates two broad data families:

- **Structured ESG data** from reputable datasets is used to compute quantitative ESG measures for entities. Below, we will describe Bloomberg ESG Disclosure Scores datasets as a key source (see Data Sources section), including separate governance and environmental/social datasets.
- **Unstructured CMBS documentation** is used as the primary source for identifying transaction entities and roles. In this article, we specifically use the **Freddie Mac Green Bond CMBS prospectus** in the case study and positions prospectuses and similar documents as information-dense sources in which entity roles are described but not readily machine-readable.

Our approach is, therefore, to treat documents as a core dataset rather than as supplementary context.

4.2 Document intelligence: what is extracted

Extraction objectives are practical and targeted. From CMBS documents, the methodology aims to identify:

- **Entities** involved in the structure (e.g., originators, underwriters, trustees, and servicers).
- **Roles** and responsibilities of those entities (used later for hierarchical weighting).
- Supporting information needed to link entities to datasets and to explain scoring.

The methodology section also proposes that entity extraction can be supported by an "entity extraction score" that combines signals from multiple NLP models (zero-shot, transformers, and LLMs) with empirically set weights.

4.3 How extraction is performed (high level)

We describe the use of multiple NLP approaches to classify and extract entities from unstructured text:

- **Zero-shot classification** is used to identify entities and assign roles without requiring extensive labelled training data. This is considered necessary because structured credit is a niche, and labelled datasets may be scarce.
- **Transformer models (BERT/RoBERTa)** are used because they are strong at capturing contextual relationships in dense financial language.
- **LLMs** are used for contextual understanding and nuanced role classification.

We further emphasize a training and testing process (noted in a figure reference) to justify model selection and maintain extraction reliability.

4.4 Conceptual data flow and transparency

In this article, we explicitly reference a "CMBS conceptual data map" (Figure 3) that illustrates the end-to-end flow from data sources through NLP processing, structured data integration, and scoring outputs. While the summary here does not reproduce the figure, the operational implication is essential: the methodology is designed as a repeatable pipeline with clearly defined processing stages.

4.5 What this article does not specify

The article describes the conceptual pipeline and model types, but it does not provide full implementation-level details, such as specific software infrastructure choices, exact training datasets, or the production deployment architecture. Where operational teams adopt this approach, those details would need to be designed and governed in the implementation stage.

5. Entity mapping, identifiers, parent linkage, and data quality governance

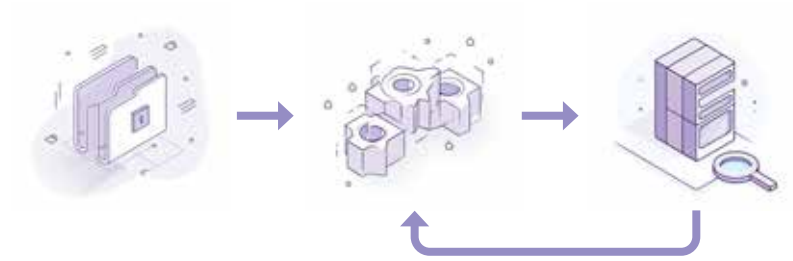
After extraction, entities exist as names in text. To combine document-extracted entities with structured ESG datasets and sentiment data, we highlight the need to link entities to stable identifiers and, where appropriate, to parent entities. This mapping layer is particularly important in structured credit because the same organization may appear under different legal entity names, and structured ESG data coverage may be stronger at the parent level.

5.1 Why mapping matters in structured credit

We frame mapping as necessary for three reasons:

1. **Joining data source**s: structured ESG datasets and sentiment datasets require stable identifiers; document text alone is not sufficient.
2. **Avoiding duplication and confusion**: without parent linkage, a single group could be counted multiple times under different subsidiaries or divisions.
3. **Improving coverage**: private or thinly covered entities may not have complete ESG data, while parent entities may.

Figure 4: Training and testing process



Document annotation

- Annotator will manually go through a subset of documents and identify sections that specify the characteristics of the bond
- Annotators will use the ICMA bag of words and assurance providers list to classify a bond
- For example, to identify if a bond is green, the title of the bond may specify the word "Green bonds"

Modelling

- Data scientists train the classification model based on documents that have been annotated
- Interaction with SMEs in case of questions or clarifications

Review of results

- Results are included in the model extractions
- SMEs and data scientists review the output and identify where the classification model is correct and wrong and compute achieved accuracy
- Feedback to the modelling stage for model improvement

We can compute the accuracy by comparing the annotations to the model extractions:

The accuracy is the number of correctly extracted answers over the total number of answers

5.2 Identifiers referenced in this article

In the Freddie Mac case study narrative, we explicitly reference linking extracted entities to identifiers such as:

- **ISIN** (security identifier).
- **LEI** (entity identifier).
- **PermID** (entity identifier used for linkage in some data ecosystems).

We position this linkage as part of "customizing the concept" from generic extraction toward an operational approach: extract entities, link to identifiers, retrieve ESG data, and tag entities intelligently.

5.3 Parent linkage and governance logic

Parent mapping is a practical necessity for reliably retrieving structured ESG scores, especially when counterparties appear as subsidiaries, divisions, or specific legal entities in the documentation. It also aligns with the broader methodological goal: role-based weighting is only meaningful if entities are consistently identified and linked across datasets.

5.4 Data quality and confidence

While we do not provide a complete data governance framework, we do emphasize transparency and explainability. In practice, this implies that entity mapping should be auditable: the system should preserve how an entity was mapped and which dataset values were used. We also discuss combining model outputs with weights determined by accuracy and robustness benchmarks, which is consistent with using confidence measures to govern extraction and mapping.

5.5 What this article does not specify

This article does not specify a complete mapping hierarchy (e.g., all possible identifier types, deterministic rules, or reconciliation procedures across multiple entity masters). It also does not specify how mapping exceptions should be handled operationally (e.g., escalation workflow). These would be natural extensions of the implementation consistent with this article's emphasis on transparency.

Table 2 summarizes the key steps involved in mapping entities extracted from structured credit documentation to stable identifiers and parent organizations to retrieve structured ESG data. The checklist reflects the entity extraction and mapping logic described in this article and illustrated in the Freddie Mac Green Bond CMBS case study, and highlights the importance of transparency and validation in ESG scoring.

6. Scoring approach (role-based hierarchy, explainability, monitoring)

This article's scoring approach is designed to address a structured credit-specific problem: ESG risk is distributed across entities with different roles, and ESG signals combine slow-moving structured metrics with fast-moving controversy information. The scoring methodology, therefore, has three defining features: hierarchical weighting, quantitative ESG integration, and sentiment-based dynamic adjustment, all supported by transparency.

6.1 Role-based hierarchical weighting

This article introduces hierarchical weighting to reflect the fact that different roles have varying ESG relevance and risk exposure. It provides role logic that aligns with structured credit intuition:

- **Originators (highest weight)**: because they directly influence the quality and ESG characteristics of underlying loans.
- **Underwriters**: because they affect ESG-related disclosures and the investor's initial ability to assess risk.
- **Servicers**: because they manage ongoing ESG risk and compliance across the life of the security.
- **Trustees (lowest weight)**: because their roles are primarily administrative and have less direct ESG impact.

We also provide an illustrative example of role-based weights, such as setting the originator weight higher than the servicer and trustee (e.g., originator = 1, servicer = 0.8, trustee = 0.6), as a way to formalize the hierarchy. We further note that industry-specific ESG weighting can be incorporated using benchmarks.

Table 2: Identifier and mapping checklist for structured credit ESG assessment

Step	Description	Example from the article
Entity name extracted	Legal entity name identified from deal documentation	"Midland Loan Services, a division of PNC Bank, National Association"
Identifier(s) assigned	Stable identifiers linked to the extracted entity (e.g., ISIN, LEI, PermID)	PermID linked to Midland Loan Services
Parent entity mapped	Ultimate or relevant parent entity identified for ESG data retrieval	PNC Financial Services Group
Structured ESG data retrieved	Quantitative ESG metrics retrieved using identifier/parent linkage	ESG metrics for PNC (hypothetical placeholders)
Confidence/validation	Validation of mapping and data linkage (manual review or model confidence)	Mapping confirmed via document context
Source evidence retained	Document excerpt or reference supporting entity-role mapping	Prospectus section describing the servicing role

6.2 Quantitative ESG integration (structured metrics)

This article integrates structured ESG data into the methodology as a replicable quantitative foundation. It describes quantitative ESG scoring across Environmental, Social, and Governance components and notes that factor weights can reflect relevance and investor preferences, with normalization against standards to support comparability. It later identifies Bloomberg ESG Disclosure Scores datasets as sources for governance and environmental/social metrics. In other words, the quantitative score is grounded in structured datasets, while the document intelligence layer ensures the correct entities and roles are evaluated.

6.3 News-based sentiment and controversy integration

To improve responsiveness and mitigate the limitations of static, self-reported ESG data, this article integrates news-based sentiment analysis. It describes sentiment scoring in terms of:

- Polarity (negative to positive).
- Volume-based weighting (reflecting significance/visibility of news).
- The number of news items analyzed.

The goal is to capture real-time ESG-related events and controversies that can materially change risk perception before they appear in periodic disclosures.

6.4 Combining quantitative and sentiment components

We propose a weighted integration of quantitative and sentiment-based scores, typically favoring structured ESG data. We discuss a parameterization that commonly assigns structured ESG the dominant weight (for example, we cite an α avoring structured ESG and report a typical value of 0.75). The case study tables use a 75% quantitative/25% sentiment split to demonstrate how component contributions add to a final hybrid score.

In this summary, the exact mathematical expression is omitted intentionally to keep the focus on executive interpretation rather than formula mechanics. The aim is to ensure that the combination is transparent and empirically justified, rather than arbitrary.

Table 3: Role hierarchy and indicative weights

Entity Name	Role	Hierarchical weight
Freddie Mac	Originator	Highest
J.P. Morgan	Lead underwriter	High
Wells Fargo	Trustee	Low
Midland Loan Services	Master servicer	Medium
Berkadia	Special servicer	Medium

Note: This generic illustrative table explicitly shows hierarchical weights based on defined entity roles. Real structured ESG scores would be derived from Bloomberg or equivalent sources.

6.5 Temporal dynamics: smoothing for stability and monitoring

This article explicitly introduces temporal smoothing using an exponential moving average approach to reduce short-term volatility while remaining responsive to ESG developments. This reflects a practical reality: sentiment signals can be noisy, and investors need a stable yet monitorable score. It frames this as a balance between stability and responsiveness.

6.6 Explainability and traceability

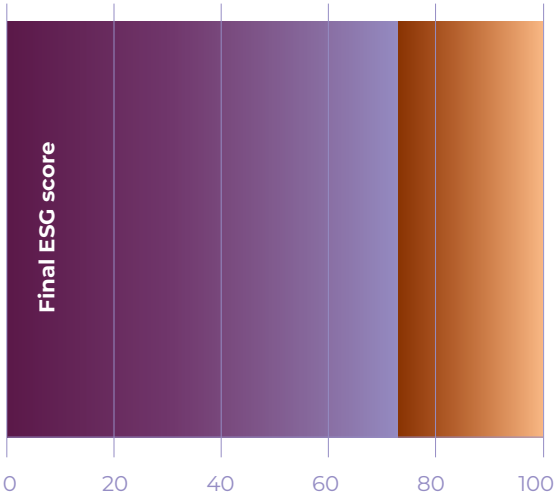
A key emphasis in this article is transparency and explainability: stakeholders should understand how components contribute to the final score. In an investment context, this implies outputs such as:

- Contribution breakdown (quantitative versus sentiment).
- Entity/role level drivers.
- Clear linkage between extracted entities and the score inputs.

This article argues that this explainability is essential for trust and decision-making in structured finance environments.

Figure 5 illustrates how the final ESG score for a structured credit transaction is composed of a dominant structured ESG baseline, derived from quantitative ESG metrics at the relevant entity and parent level, and a smaller news-based sentiment and controversy overlay designed to capture dynamic developments and mitigate reliance on self-reported disclosures. Values shown are hypothetical placeholders based on the Freddie Mac Green Bond CMBS illustration in the article.

Figure 5: Score decomposition view



7. Case study walk-through (Freddie Mac Green Bond CMBS) - only what the article states

We use Freddie Mac's Green Bond CMBS documentation as a case study to demonstrate the feasibility of the proposed approach. The case study presents a practical illustration of how document-based entity extraction and hybrid scoring can work in a real, structured credit context.

The Freddie Mac Green Bond CMBS program provides a suitable illustrative example, as it is explicitly designed to support energy-efficient and environmentally sustainable multifamily housing projects (Freddie Mac, 2020).

The purpose of this example is to demonstrate methodological feasibility and interpretability, rather than to provide calibrated or validated ESG scores.

7.1 What document is analyzed and why

The **Freddie Mac's Green Bond CMBS prospectus** is used as the primary document in this case study predominantly because it explicitly supports energy-efficient and environmentally sustainable multifamily housing projects. It also contains detailed descriptions of involved entities - originators, underwriters, trustees, and servicers - and their roles and responsibilities, making it a suitable testbed for extraction and hierarchical scoring.

7.2 NLP-based entity and role extraction

Using the methodology described earlier, entities and roles are extracted systematically from the prospectus. The extraction process combines:

- Zero-shot classification (for flexible role categorization).
- Transformer models (BERT/RoBERTa) for refinement and role differentiation.
- LLMs for contextual nuance.

We also include an illustrative entity role table that demonstrates hierarchical weights for a set of named entities and roles. In that illustrative example, the extracted entities include: Freddie Mac (Originator), J.P. Morgan (Lead Underwriter), Wells Fargo (Trustee), Midland Loan Services (Master Servicer), and Berkadia (Special Servicer), with qualitative hierarchical weight levels (e.g., highest/high/medium/low).

We also note that, for the specific case study, the concept can be further customized to link extracted entities to identifiers and to support ESG data retrieval and intelligent tagging.

7.3 Identifier linkage and enrichment (as described)

In this article, we explicitly describe linking extracted entities to identifiers such as PermID, ISIN, and LEI to enable ESG data retrieval and structured integration. We present an illustrative mapping table for a hypothetical security identifier (shown as "FM12345689" in the table) and list several transaction parties and their parent parties.

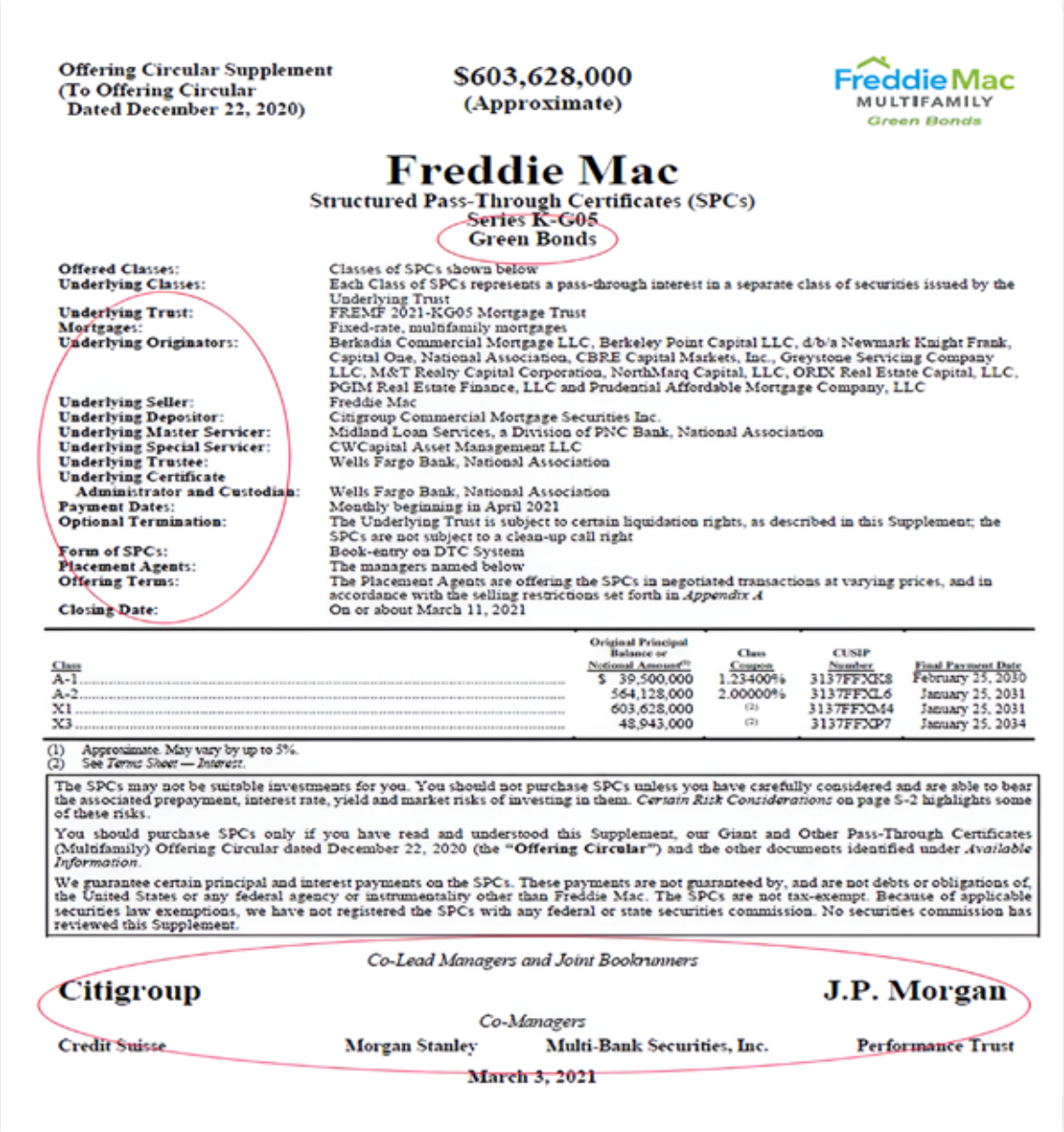
In the more detailed illustrative party list, the entities/roles include:

- **Underlying issuer**: Freddie Mac
- **Underlying trust**: FREMF 2021 KG05 Mortgage Trust (parent party noted as Freddie Mac)
- **Underlying originators (examples listed)**: PGIM Real Estate Finance, Berkeley Point Capital LLC, Newmark Knight Frank, and Capital One
- **Underlying seller**: Freddie Mac
- **Underlying depositor**: Citigroup Commercial Mortgage Securities Inc. (parent party shown as "Citi")
- **Underlying master servicer**: Midland Loan Services, a Division of PNC Bank, National Association (parent party shown as "PNC Bank")
- **Underlying special servicer**: CWCapital Asset Management LLC (parent party shown as "CWCAM")
- **Underlying trustee/certificate administrator and custodian**: Wells Fargo Bank, National Association (parent party shown as "Wells Fargo")

7.4 Application of the hybrid ESG scoring method

Hybrid scoring methodology integrates structured ESG scores and sentiment analysis to generate

Figure 6: Illustrative example of a Freddie Mac Green Bond



Offering Circular Supplement
(To Offering Circular Dated December 22, 2020)

$603,628,000
(Approximate)

Freddie Mac MULTIFAMILY Green Bonds

Freddie Mac

Structured Pass-Through Certificates (SPCs)
Series K-G05
Green Bonds

Offered Classes:	Classes of SPCs shown below
Underlying Classes:	Each Class of SPCs represents a pass-through interest in a separate class of securities issued by the Underlying Trust
Underlying Trust:	FREMF 2021-KG05 Mortgage Trust
Mortgages:	Fixed-rate, multifamily mortgages
Underlying Originators:	Berkadia Commercial Mortgage LLC, Berkeley Point Capital LLC, d/b/a Newmark Knight Frank, Capital One, National Association, CBRE Capital Markets, Inc., Greystone Servicing Company LLC, M&T Realty Capital Corporation, NorthMarq Capital, LLC, ORIX Real Estate Capital, LLC, PGIM Real Estate Finance, LLC and Prudential Affordable Mortgage Company, LLC
Underlying Seller:	Freddie Mac
Underlying Depositor:	Citigroup Commercial Mortgage Securities Inc.
Underlying Master Servicer:	Midland Loan Services, a Division of PNC Bank, National Association
Underlying Special Servicer:	CWCapital Asset Management LLC
Underlying Trustee:	Wells Fargo Bank, National Association
Underlying Certificate Administrator and Custodian:	Wells Fargo Bank, National Association
Payment Dates:	Monthly beginning in April 2021
Optional Termination:	The Underlying Trust is subject to certain liquidation rights, as described in this Supplement; the SPCs are not subject to a clean-up call right
Form of SPCs:	Book-entry on DTC System
Placement Agents:	The managers named below
Offering Terms:	The Placement Agents are offering the SPCs in negotiated transactions at varying prices, and in accordance with the selling restrictions set forth in *Appendix A*
Closing Date:	On or about March 11, 2021

Class	Original Principal Balance or Notional Amount(1)	Class Coupon	CUSIP Number	Final Payment Date
A-1	$ 39,500,000	1.23400%	3137FFXK8	February 25, 2030
A-2	564,128,000	2.00000%	3137FFXL6	January 25, 2031
X1	603,628,000	(2)	3137FFXM4	January 25, 2031
X3	48,943,000	(2)	3137FFXP7	January 25, 2034

(1) Approximate. May vary by up to 5%.
(2) See *Terms Sheet — Interest*.

The SPCs may not be suitable investments for you. You should not purchase SPCs unless you have carefully considered and are able to bear the associated prepayment, interest rate, yield and market risks of investing in them. *Certain Risk Considerations* on page S-2 highlights some of these risks.

You should purchase SPCs only if you have read and understood this Supplement, our Giant and Other Pass-Through Certificates (Multifamily) Offering Circular dated December 22, 2020 (the "**Offering Circular**") and the other documents identified under *Available Information*.

We guarantee certain principal and interest payments on the SPCs. These payments are not guaranteed by, and are not debts or obligations of, the United States or any federal agency or instrumentality other than Freddie Mac. The SPCs are not tax-exempt. Because of applicable securities law exemptions, we have not registered the SPCs with any federal or state securities commission. No securities commission has reviewed this Supplement.

Co-Lead Managers and Joint Bookrunners

Citigroup **J.P. Morgan**

Co-Managers

Credit Suisse **Morgan Stanley** **Multi-Bank Securities, Inc.** **Performance Trust**

March 3, 2021

comprehensive ESG scores for identified entities. Importantly, in the case study demonstration:

- **Structured ESG scores are hypothetical placeholders** pending retrieval from Bloomberg (or equivalent).
- **Sentiment analysis is applied to publicly available news datasets** to assess controversies and ESG-related events.
- The provided scores are **explicitly hypothetical and clearly marked as placeholders** for illustrative purposes.

The scoring tables demonstrate how a final hybrid score is formed using a 75% quantitative/25% sentiment contribution split. Table 4 provides illustrative scores for entities including Freddie Mac, J.P. Morgan, Wells Fargo, Midland Loan Services, and Berkadia, and shows how the quantitative and sentiment components contribute to the final score.

Table 4: Illustrative entity scores and decomposition (quantitative versus sentiment)

Entity Name	Structured ESG score	Sentiment ESG score	Final hybrid ESG score (75% quantitative + 25% sentiment)
Freddie Mac	85	90	86.25
J.P. Morgan	78	70	76.00
Wells Fargo	72	65	70.25
Midland Loan Services	80	85	81.25
Berkadia	75	88	78.25

Note: Scores provided above are explicitly hypothetical and clearly marked as placeholders for illustrative purposes.
All numerical values are illustrative placeholders and do not represent live or proprietary ESG scores.

7.5 Results interpretation (as described)

The results discussion is qualitative and focused on feasibility: it states that the hybrid approach balances structured quantitative ESG data with dynamic sentiment inputs, producing a comprehensive and responsive ESG assessment tailored to CMBS. The key takeaway is that the framework can (i) extract the relevant entity network from documentation, (ii) apply role-based weighting, and (iii) transparently integrate quantitative and sentiment signals.

7.6 What the case study does not claim

We do not present the case study as a fully calibrated, production-grade validation with audited input datasets; instead, it presents it as an empirical demonstration of method feasibility using illustrative, placeholder scores. It also does not specify predictive performance metrics (e.g., forecasting defaults) for the case study output; the emphasis is on methodology design, transparency, and applicability.

8. Practical implementation considerations (operating model, controls, cadence, limitations)

While this article is primarily methodological, it also discusses practical implications, limitations, and future research directions that are directly relevant to implementing the approach in an investment setting.

8.1 Stakeholder value and operating implications

We have identified multiple stakeholder benefits:

- **Investors**: improved ability to assess ESG risks in structured products with more granularity and less reliance on issuer-level proxies, supporting better-informed decisions and potentially reducing exposure to unforeseen ESG risks.
- **Issuers**: clearer insight into ESG performance drivers, informing deal structuring and potentially improving investor appeal and pricing efficiency.
- **Regulators**: a transparent framework that may align with evolving ESG requirements and could inform future disclosure standards.

- **Rating agencies**: a practical model for integrating NLP entity extraction and sentiment into ESG assessments to improve timeliness and utility.
- **Market efficiency**: improved pricing and recognition of ESG risks due to more transparent and responsive scoring.

8.2 Governance and controls (what must be true for investor-grade use)

In this article, transparency and explainability are repeatedly emphasized. Translating that into governance terms implies a set of controls that an operating model would need:

- **Traceability**: the ability to explain how each component contributes to the overall score.
- **Model justification**: the choice of NLP models and weights is framed as empirically justified by performance benchmarks in finance NLP literature.
- **Monitoring**: the inclusion of temporal smoothing reflects a desire for stable monitoring rather than noisy reactivity.

We do not specify a complete model risk management framework (e.g., formal validation checklists, audit pack formats, or approval committees), but the design principles point toward those requirements.

8.3 Update cadence and lifecycle monitoring

We frame structured product ESG assessment as dynamic and introduce temporal smoothing to manage volatility. In practical terms, that implies:

- Structured ESG datasets update on their own cadence (often periodic), while
- Sentiment and controversy signals can update more frequently, requiring governance to prevent noise from dominating.

We do not specify a concrete refresh schedule (e.g., daily vs. weekly), but do highlight the need for ongoing updates throughout the product lifecycle.

8.4 Limitations and future research directions

We list several limitations and opportunities for future work:

- **Data availability and quality**: the methodology relies on the completeness and reliability of ESG data, which can be especially challenging for private entities.
- **Broader validation**: empirical validation across additional structured finance products (RMBS, ABS, CLOs) are required to establish generalizability and predictive validity.
- **Dynamic weight calibration**: future research could calibrate hierarchical weights dynamically using advanced ML to reflect changing market conditions and materiality.
- **Sentiment refinement**: sentiment signals can be affected by media bias and noise; there is a need for refining and diversifying sources and techniques.
- **Regulatory alignment**: ongoing alignment with evolving ESG regulations and reporting standards across jurisdictions.
- **Asset level integration**: incorporating granular asset-level data (e.g., energy efficiency ratings, flood risk assessments) to strengthen assessments.
- **Temporal dynamics**: extending lifecycle monitoring and updating for structured products.
- **AI ethics and bias mitigation**: explicit attention to ethical implications and bias detection/mitigation in NLP and ML methods.
- **Cross-asset comparability**: development of a standardized framework to compare ESG across different structured product types.

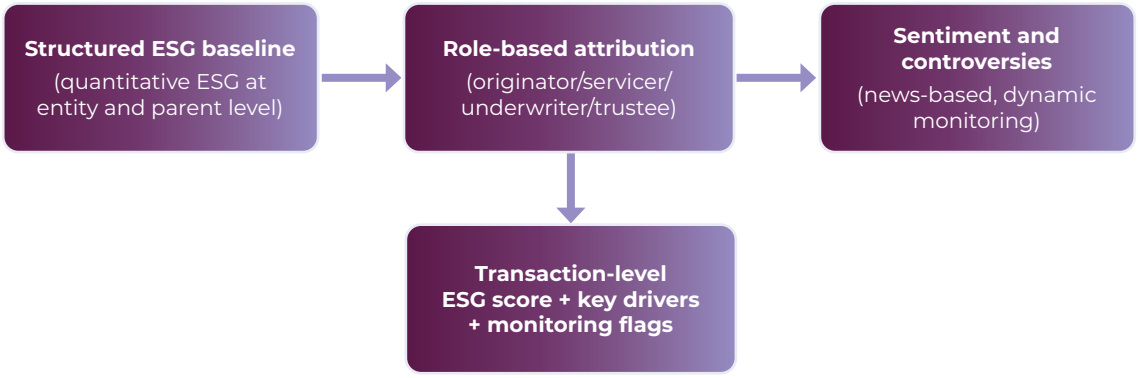
9. Future outlook

This article positions its approach as an advancement in sustainable structured finance analytics, arguing that continued refinement, validation, and broader application can support a shift toward greater sustainability, transparency, and investor confidence in structured credit markets.

10. Conclusion

This article addresses a core challenge in sustainable finance: structured credit products - particularly CMBS - do not fit neatly into traditional ESG scoring frameworks designed for single-issuer securities. In structured products, ESG risk is distributed across multiple parties with different roles, incentives, and degrees of influence. At the same time, the most decision-relevant evidence often

Figure 7: Monitoring dashboard concept - illustrative



sits in dense, unstructured transaction documents, while structured ESG datasets can be incomplete (especially for private entities) and vulnerable to disclosure bias.

To close this gap, this article proposes a hybrid ESG scoring methodology tailored to structured credit. It combines three complementary layers: structured quantitative ESG metrics for replicability and comparability, NLP driven document intelligence to extract transaction entities and roles and enable role based hierarchical weighting, and news-based sentiment and controversy signals to improve timeliness and reduce reliance on self reported information. The scoring logic is designed to be transparent - showing how quantitative and sentiment components contribute to the overall output - and it introduces temporal smoothing to balance responsiveness with stability.

A Freddie Mac Green Bond CMBS case study demonstrates the approach in practice using the bond prospectus as an extraction source, linking extracted entities to identifiers such as ISIN/LEI/PermID, and illustrating hybrid score decomposition using explicitly hypothetical placeholder scores. While the case study is framed as feasibility rather than full production validation, it shows how the method can transform unstructured documentation into an attributable, monitorable ESG view.

This article concludes that more rigorous validation, improved data coverage, refined sentiment techniques, and deeper asset level integration are key next steps. Overall, it positions hybrid, document-aware ESG scoring as a practical path to more accurate, responsive, and explainable sustainability assessments in structured credit - supporting better investor decision making and aligning with evolving regulatory expectations.

References

Berg, F., J. F. Koelbel, and R. Rigobon, 2022, "Aggregate confusion: the divergence of ESG ratings," Review of Finance 26:6, 1315-1344

Bua, G., D. Kapp, F. Ramella, and L. Rognone, 2024, "Transition versus physical climate risk pricing in European financial markets: a text-based approach," European Journal of Finance 30:17, 2076-2110

DBRS Morningstar, 2023, "ESG factors in global CMBS," commentary report, DBRS Morningstar, https://tinyurl.com/9awn4pyn

Du, F., K. Yu, and J. Xiao, 2024, "A text-based measure of ESG risk exposure," Procedia Computer Science 238, 240-247

Holtermans, R., 2024, "Climate risk and commercial mortgage delinquency," Journal of Real Estate Finance and Economics 64:4, 994-1037

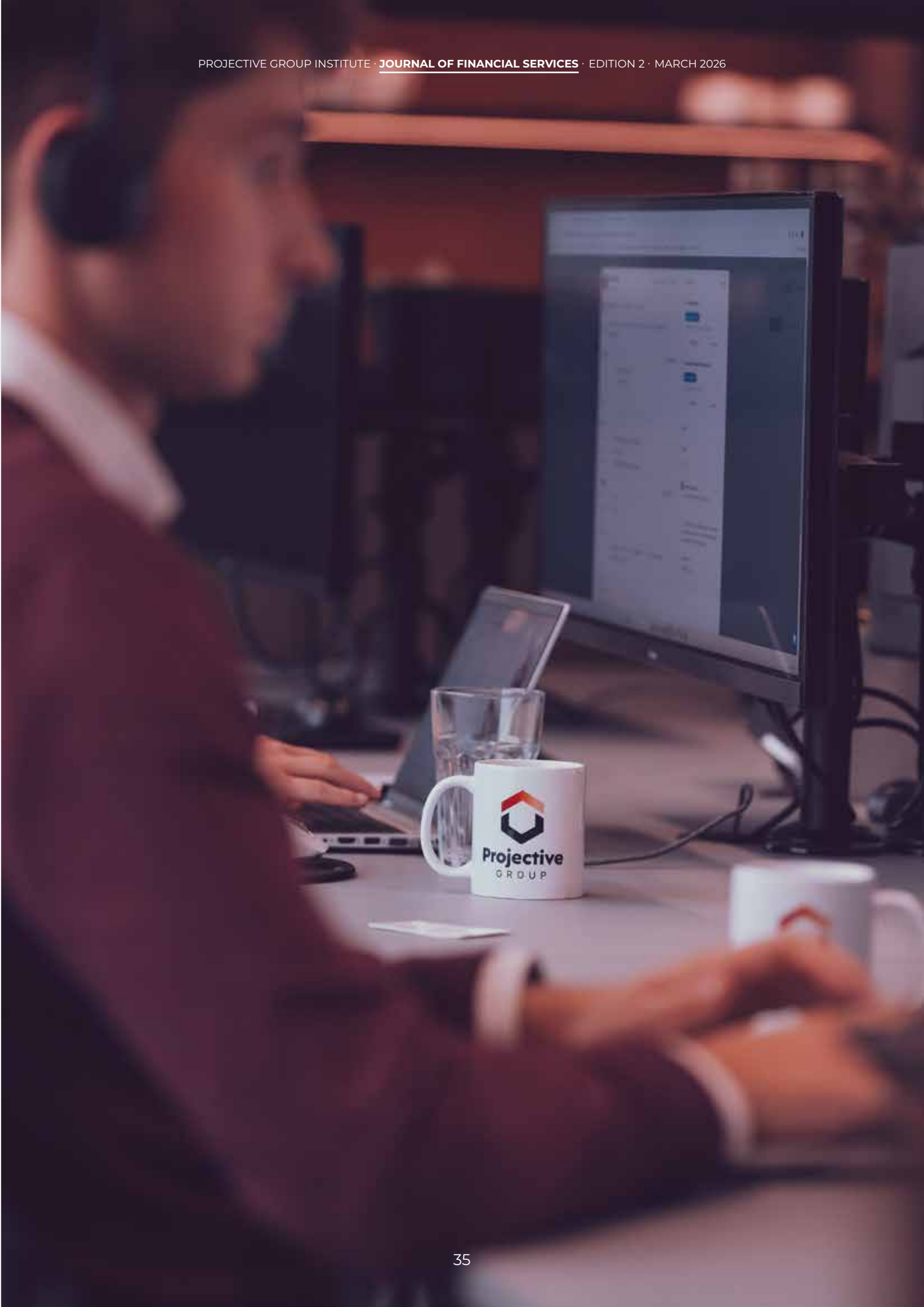
Kotsantonis, S., and G. Serafeim, 2019, "Four things no one will tell you about ESG data," Journal of Applied Corporate Finance 31:2, 50-58

Ouazad, A., and M. E. Kahn, 2019, "Securitisation dynamics in the aftermath of natural disasters," NBER working paper no. 26322, National Bureau of Economic Research

PRI, 2021, "ESG incorporation in securitised products: the challenges ahead," Principles for Responsible Investment

Schimanski, T., A. Reding, N. Reding, J. Bingler, M. Kraus, and M. Leippold, 2024, "Bridging the gap in ESG measurement: using NLP to quantify environmental, social, and governance communication," Finance Research Letters 61, 104979

Schmidt, A. B., 2023, "News-based ESG ratings for optimal portfolios: SASB versus SDG," SSRN working paper


Projective
GROUP

How data and AI shape financial services

The **data integrity imperative**: from predictive models to AI agents in financial services

Richard P. Padbury,
Director of Research and Ecosystems,
BMO Financial Group

Ryan B. Duffy,
U.S. Chief Data and Analytics Officer,
BMO Financial Group

Kristin L. Milchanowski,
Chief Artificial Intelligence (AI) and Data Officer, BMO Financial Group

ABSTRACT

In financial services, the transition from Machine Learning to goal-oriented autonomous AI Agents promises transformative improvements in speed, scale, and cost. Yet reliability at production scale is constrained by challenges rooted in the data foundation. Notably, hallucinations and decision errors are not generated from algorithmic model randomness, but from data drift, weak lineage, and stale or ungoverned information and context. This paper asserts that dependable AI is fundamentally a data-layer challenge and examines these challenges in the context of the banking sector. Substantive progress in Generative Artificial Intelligence (GenAI) results from strengthening a unified, governed, and observable data plane that embeds continuous data-quality telemetry, automated source-to-agent lineage, near real-time schema-change detection, and policy-as-metadata to enforce privacy and purpose-bound access. The architecture follows FAIR principles and supervisory expectations (e.g., SR 11-7/OCC 2011-12, BCBS 239, and GDPR/PIPEDA), and describes multi-agent integrity patterns (i.e., builder/checker) and self-healing pipelines that prevent silent error propagation. When combined, these controls and capabilities elevate the horizontal foundation to regulatory-grade standards and enable workflow-embedded agents in vertical functions (e.g., onboarding, lending, AML, and Treasury) to plan, act, and adapt with confidence while remaining human-centered in their design and application.

1. Introduction: from models to autonomous intelligence: The Evolution in Banking AI

The trajectory of Artificial Intelligence (AI) in financial services is progressively the integration of human expertise and machine capabilities to enhance routine decision making. This evolution began with the industrialization of Machine Learning (ML), advanced through the creative capabilities of Generative Artificial Intelligence (GenAI) and is now progressing towards emerging AI Agent systems that operate as goal-oriented assistants within the enterprise. These systems are designed to augment human judgment, work within established controls and governance, and ensure that accountability, transparency, and auditability remain central throughout the decision lifecycle.

1.1. Defining the arc of autonomy

To govern this transition effectively, leadership must maintain absolute clarity on the distinctions between these technological phases. ML remains the cornerstone of predictive analytics, applying statistical patterns in historical datasets to power customer insights, credit scoring, and initial risk assessments. GenAI shifted the focus toward language, content, and interface transformation, enabling hyper-personalized engagement and the automation of document intensive workflows.

The industry is now entering the era of AI Agents, systems designed to build on this integration by collaborating with humans to initiate and orchestrate decisions in real-time under defined governance frameworks. Looking ahead, these agents are expected to evolve into goal-oriented, self-improving systems that operate with embedded human oversight and accountability. Unlike previous generations of automation, agentic systems can reason, plan, and execute complex, multi-step tasks across operating groups, while maintaining transparency, auditability, and alignment with regulatory standards.

1.2. The exponential value proposition

A growing body of evidence suggests early GenAI initiatives have fallen short of expectations [Challapally et al. (2025)]. The core issue is that many organizations have concentrated on horizontal applications; broad, organization-wide tools that are easy to deploy but typically deliver only modest, diffuse value. Meanwhile, the more transformative opportunities lie in vertical, function-specific applications, yet these efforts rarely progress beyond pilot phases due to challenges such as fragmented data, legacy systems, integration complexity, and organizational inertia [Challapally et al. (2025), Sukharevsky et al. (2025)]. AI Agents are viewed as a solution to overcome this paradox [Sukharevsky et al. (2025)].

This technological arc is projected to deliver exponential gains in performance and return on investment (RoI). Early implementations of AI Agents indicate potential productivity gains. As highlighted in Table 2, in one example, a research firm boosted data quality and achieved enhanced

Table 1: From models to autonomous intelligence: the evolution of AI

AI category	Core competency	Primary interaction	Outcome
Machine learning	Prediction	Data analysis	Default forecasting; fraud detection
Generative AI	Content generation	Prompt-response	Personalized marketing; virtual concierge
AI Agents	Orchestration	Goal execution	Trade reconciliation; real-time AML checks
Agentic AI	Agency	Goal-driven autonomy	Adaptive treasury management; end-to-end loan lifecycle

Table 2: AI Agent use cases

Strategic outcome	Performance improvement	Economic impact
Enhanced market insights	Productivity gains of ~60%	Annual savings >$3 million [Challapally et al. (2025)]
Know your customer (KYC)	File-closure rates improved ~67%	KYC-related costs reduced by ~20% [Seibert et al. (2025)]
Fraud prevention	2-4x more suspicious activity identified versus legacy system	~50% fewer false positives [May (2023)]
Faster credit decision	30-60% increase in productivity	~30% faster credit turnaround [Challapally et al. (2025)]

market insights leading to productivity gains of 60% and annual savings of over U.S.$3 million [Challapally et al. (2025)]. In another example, a retail bank reported achieving significant productivity gains and a 30% faster credit decision process by deploying AI Agents to automate data collection and draft credit-risk memos, shifting relationship managers from manual writing to strategic oversight [Sukharevsky et al. (2025)].

Industry data shows that 93% of financial institutions expect AI to improve profitability over the next five years, with Citibank projecting a 9% boost (equivalent to U.S.$170 billion) in global banking profits by 2028 [Citi (2024)]. Realizing these gains requires more than deploying specialized agents in complex workflows like lending and customer onboarding; it requires an enterprise wide data infrastructure that enables those agents to access clean, governed, and cross-domain context in real time.

These requirements lead directly to the next realization: horizontal data foundations are not a downstream concern, but the primary enabler of vertical agent performance.

2. Critical dependency on horizontal foundations

The fundamental realization for the C-suite is that exponential gains are impossible if the data foundation remains siloed. This is true across the full AI arc: traditional ML requires consistent, integrated features to avoid leakage and bias; GenAI depends on current, governed retrieval to ground answers; AI Agents need real-time context from multiple systems to decide and act; and agentic intelligence adds planning, monitoring, and self-correction. This amplifies the demand for a unified, cross-domain data plane.

Agents require an architecture that spans business units and systems, because they must pull context from varied sources simultaneously to make accurate decisions. This progression is intrinsically linked to customer centric outcomes, enabling personalization at scale, greater speed of service, and stronger protection through timely fraud defense. Crucially, the move from ML to GenAI to AI Agents to Agentic AI is additive: it extends capability rather than replacing prior methods, and at every stage the independent data layer is the non-negotiable prerequisite.

This unified architecture is not only a technical imperative but also a governance foundation. Responsible innovation requires that human-in-the-loop controls, model validation, and compliance reviews are embedded into the same operational fabric that powers AI systems, from ML models to GenAI assistants to autonomous Agents.

Thriving in this paradigm demands a fundamental shift in mindset: AI should not be treated as a mere software enhancement but as a digitally empowered capability that must operate under uncompromising governance and transparency.

2.1. Data: the horizontal foundation for intelligent agents

With this dependency established, the data foundation must meet specific technical requirements to ensure the reliability and accuracy of intelligent agents. In an agentic world, data is no longer a static repository for reporting; it is a real-time, contextual fuel that enables autonomous systems to perceive their environment, reason over it, and act.

Supporting this requires a shared foundation with specific architectural qualities. ML pipelines need reconciled features drawn from consistent sources; Gen AI models require access to governed, up-to-date content; agents must retrieve and combine context from many systems in parallel; and agentic frameworks must be able to monitor, explain, and correct their own outputs. These capabilities depend on a single, governed data layer that exposes freshness, lineage, and compatibility across the enterprise, ensuring that every model and every agent operate from the same truth set.

2.2. Requirements for agent reliability

To operate effectively across the banking enterprise, agents require integration that breaks the "silo" mentality and provides real-time, cross-business context. This ensures, for example, that an agent managing a customer's mortgage application has the same situational understanding as the agent managing their wealth portfolio or credit-card limits. In practice, this demands common entity resolution, shared semantic definitions, and source-to-agent lineage so decisions can be traced, challenged, and audited.

Furthermore, privacy and personal information protection must be enforced at every touchpoint. Once data is in motion within a multi-agent system, the traditional perimeter-based security model becomes insufficient. Privacy controls must be embedded directly in the data pipelines, minimizing access by default, enforcing purpose-bound use, and ensuring agents only retrieve the minimum necessary information for their specific task. These protections should travel with the data (policy-as-metadata), enabling auditable and compliant autonomy across ML, GenAI, AI Agents, and agentic workflows.

But in many institutions, fragmented platforms and tool-specific entitlements prevent uniform enforcement. This gap defines the limitations of legacy architecture and directly constrains the reliability, scalability, and governance of intelligent agents.

3. Limitations of legacy architectures

Many incumbent institutions lack the integrated and governed data foundation required for agent reliability. Heterogeneous platforms, brittle integrations, and manual processes produce inconsistent and stale records. Agents consuming this input propagate those defects through downstream decisions at scale. Common limitations include:

- **Poor lineage and discoverability**: without a clear map of where data originated and how it has been transformed, agents may pull from contradictory or outdated "sources of truth", leading to inconsistent outputs.
- **Lack of reusability**: data formatted for a specific reporting tool is often unusable for an autonomous agent requiring high-granularity context.
- **Governance gaps**: manual data cleaning efforts are unsustainable at the scale required for real-time agents. This leads to risk and compliance issues, including bias, opacity in reasoning, and frequent breaches.

These limitations often lead to agentic drift; the divergence of agent behavior from intended objectives due to misaligned or stale context, insufficient policy constraints, or degraded telemetry. Mitigation requires gated execution, lineage-based reconstruction, policy-as-metadata, and human-in-the-loop escalation embedded within established governance frameworks.

3.1. Principles and standards for the new data era

To overcome legacy limitations, the data foundation must be anchored in proven frameworks and supervisory expectations. This ensures agent decisions are traceable, explainable, and defensible. For

Table 3: Regulatory standards and implications for AI Agents

Regulatory Framework	Core data requirement	Implication for agents
OSFI E-23 [OSFI (2025)]	Model integrity and data accuracy	Agents must be validated at every stage of the lifecycle
SR 11-7 [Fed (2011)]/ OCC 2011-12 [OCC (2011)]	Independent model validation and "effective challenge"	Decision pathways must be explainable and auditable
PIPEDA/GDPR	Privacy-by-design and purpose limitation	Data minimization must be automated in pipelines
BCBS 239 [BCBS (2013)]	Risk data aggregation and reporting accuracy	Real-time lineage is essential for compliance

example, the FAIR principles (Findable, Accessible, Interoperable, Reusable) provide a practical compass for making enterprise data machine-actionable and reusable at scale, rather than trapped in silos or bespoke formats [Owens et al. (2025)].

A critical, but often under-communicated, dimension of the data strategy is semantic consistency, achieved through ontologies and taxonomies. In banking, ontologies provide a formal, machine-readable way to define entities (clients, legal structures, accounts, and instruments) and the relationships among them (ownership, control, transaction flows, and exposure/risk linkages). This semantic layer ensures that agents interpret the same facts in the same way across business lines. For example, the industry-standard Financial Industry Business Ontology (FIBO) captures relationships between a client, their legal entities, associated transactions, and relevant obligations; deploying such semantics helps agents avoid logic errors caused by inconsistent labels, duplicated identifiers, or ambiguous entity mappings [EDM Council (2017)].

Why are machine actionable data design and semantic consistency important? Leveraging frameworks, like FAIR, is not simply a technical choice to boost performance; in many sectors, they also serve as a critical foundation for meeting regulatory requirements. Regulators increasingly expect model decisions to be supported by data quality, accuracy, lineage, and effective challenge throughout the model lifecycle (e.g., OSFI E-23 in Canada; SR 11-7/OCC 2011-12 in the U.S.). These expectations apply as much to agentic systems as to traditional models: if agents materially influence credit, fraud or anti-money laundering (AML), for example, the data feeding those agents must be governed and auditable to regulator-grade standards.

Bottom line: data frameworks (e.g., FAIR) and regulatory expectations define how data must be managed; ontologies/taxonomies define what that data means. Together, they enable agents to consume current, contextual, and traceable information to produce outputs that remain explainable and auditable.

4. Building the trusted, resilient data layer for agents

Having established the constraints of legacy architectures, the next step is constructing the trusted data foundation required to reduce hallucinations and support reliable agent behavior at scale. This foundation elevates data from a passive asset to a lifecycle governed resource aligned with policy and supervisory standards and instrumented for continuous observability.

4.1 Managing data as a governed asset

In the agentic era, data management must move from periodic audits to continuous governance. Practically, this requires architecture that integrates data across business units and supplies cross domain context to agents. The same controls used for financial reporting (lineage, validation, and reconciliation) must apply to autonomous decisions: agents should only act on validated, current, and traceable inputs. A resilient data foundation is enabled by three capabilities:

- **Automated discovery and structuring**: use AI assisted scanning, cataloging, and classification to keep data assets findable and accurately described, reducing manual stewardship load and preventing "unknown inputs" from reaching agents.
- **Explainable decision pathways**: maintain source to agent lineage and attach data citations to agent outputs so that every recommendation or action can be reconstructed and challenged ("effective challenge," per SR 11-7).
- **Operational resilience**: apply stress testing and traffic shaping to ensure the data foundation remains performant under high velocity agent workload. Design safe defaults when confidence, freshness, or provenance signals fall below thresholds.

Establishing governed data is only the first step; sustaining its reliability demands continuous visibility into data flows and transformations. Telemetry and lineage provide that visibility, forming the backbone for mitigating risks such as hallucinations.

4.2. The role of telemetry and lineage in hallucination mitigation

Reducing hallucinations requires continuous data quality monitoring, so that agents do not fill gaps left by bad or missing inputs. When knowledge bases are stale or schema changes silently misalign fields, a generative agent can confidently produce an incorrect answer. The remedy is to treat data signals as first-class citizens in agent pipelines by surfacing freshness, provenance, and compatibility checks, and gating execution on those signals.

Empirically, agent reliability depends more on the freshness and integrity of retrieved context than on model size or decoding settings. Structured, schema-bound tasks (e.g., SQL generation) tend to remain stable under deterministic configurations, whereas retrieval-augmented workflows common in customer support and compliance exhibit higher output variance when recency, provenance, or schema alignment slip. The most effective lever for improving agent performance is, therefore, to ensure the context window is populated with accurate, current, and provenance-checked data, which is precisely what continuous telemetry and source-to-agent lineage provide.

To operationalize this, institutions should implement real-time telemetry of freshness, completeness, timeliness, duplication, and anomaly signals across key datasets; end-to-end lineage from source through transformations and distribution to the agent, enabling reconstruction, rollback, or quarantine of bad inputs; near real-time schema-change detection with automatic contract checks and compatibility tests to prevent silent breakage; explicit recency service level agreements (SLAs) and expiry rules for retrieval, especially for Retrieval-Augmented-Generation (RAG)-backed knowledge and policy content; semantic drift monitoring to track population stability and concept coverage for critical fields (e.g., via population stability index (PSI) and ontology-based checks); and guarded execution using builder/checker multi-agent patterns with human-in-the-loop escalation when provenance, confidence, or policy constraints are violated.

By embedding telemetry, lineage, schema safety, and semantic checks into the horizontal data plane, institutions convert hallucination from an opaque

Table 4: Mitigation strategies to reduce hallucinations

Technical requirement	Link to hallucination	Mitigation strategy
Schema consistency	Upstream structural changes break downstream mapping	Real-time alerting tied to schema registries
Data freshness	Stale policy documents cause agents to cite old rules	Update frequency SLAs for RAG knowledge bases
Lineage traceability	Ungoverned transforms introduce silent errors	Automated SQL query parsing for real-time mapping
Semantic coherence	Changes in language/context degrade logic	Population Stability Index (PSI) monitoring

symptom into a detectable, diagnosable, and remediable condition before it propagates to customer decisions. With these signals exposed and enforced, AI Agents can act as active data stewards flagging stale or misaligned inputs, attaching citations and version metadata to outputs, proposing corrective actions (e.g., schema contract updates or KB refresh), opening data-quality tickets, and pausing or routing workflows when data conditions fall below defined thresholds. This proactive role sets the stage for the next evolution: AI Agents as active data stewards.

5. AI Agents as active data stewards

The complexity of the modern data estate has outpaced the capacity of manual processes. The next frontier in financial services operations is the use of AI Agents as active components that support data governance, transitioning data pipelines from static conduits to adaptive, self-healing systems [Kirubakaran et al. (2025), Priyanka et al. (2025)].

5.1. The outlook for autonomous stewardship

In this emerging model, AI Agents autonomously curate, clean, and document data sources under human oversight and compliance guardrails. These "data quality agents" continuously scan for anomalies, such as duplicate records, missing fields, or formatting errors and apply corrective actions based on learned business rules. This approach has the potential to reduce manual data cleaning efforts while improving the fidelity of data feeding production systems, enabling domain experts to focus more time on analysis and decision-making rather than routine remediation. Beyond individual agents, data infrastructures may evolve into adaptive, self-healing pipelines. These pipelines detect schema mismatches or sudden drops in data volume, cross-check historical patterns, and automatically remediate issues (e.g., job re-runs or updated Extract-Transform-Load (ETL) mappings) without waiting for human intervention.

5.2. Multi-agent integrity frameworks

One of the most effective ways to ensure the integrity of autonomous stewardship is through multi-agent systems with complementary and sometimes conflicting objectives. For example, a "builder agent" may prioritize rapid execution of a data task, while an architect or "checker agent" validates outputs against integrity rules and known facts. If a builder agent attempts to "cut corners" or fabricate values, the architect agent flags the inconsistency, resolving errors before they propagate downstream. This internal tension mirrors best practices in human engineering teams and provides a robust safeguard against hallucinations at scale.

The strategic unlock is twofold: accelerating speed and scale in banking operations, and strengthening resilience through robust risk, privacy, and regulatory controls. Hallucinations that stem from upstream data drift are no longer seen as inevitable failures but as diagnostic signals that stewardship agents must address.

These advances in autonomous stewardship and multi-agent integrity frameworks establish the foundation for BMO's vision: a resilient, trusted data ecosystem enabling reliable, scalable AI. The next section examines how this vision is operationalized.

6. Case study: BMO Financial Group (BMO)

As a North American financial institution, BMO's approach to AI is grounded in a digitally enabled, future-ready culture that prioritizes talent development, responsible innovation, and a Digital First orientation to technology. The strategy centers on reliable data foundations, accountable human-AI collaboration, and transparent governance, principles applicable beyond any single institution.

6.1. Enterprise enablement

External assessments and public disclosures underscore BMO's leadership in advancing AI talent development and organizational enablement, with a strong emphasis on broad based education and capability building [BMO (2025a)]. This commitment aligns with global best practices, including the NIST Artificial Intelligence Risk Management Framework, which defines AI systems as socio technical constructs shaped by human behavior, organizational processes, and societal dynamics [NIST (2023)]. Recognizing this interplay signals the need for multidisciplinary governance and risk management, integrating technical rigor with human-centered oversight to achieve trustworthy AI at scale.

Table 5: A conceptual multi-agent integrity framework

Stewardship agent type	Primary function	Business benefit
Cleansing agent	To autonomously correct duplicates/ formatting	Faster time-to-insight
Lineage agent	To auto-generate metadata and field mappings	Enhanced compliance and discoverability
Policy agent	To enforce access controls in real-time	Instant mitigation of privacy risks
Healing agent	To re-run failed jobs and adapt to schema shifts as they occur	Reduced data downtime and manual ticketing
Checker agent	To validate outputs against integrity rules and facts prior to promotion	Increased trust in outputs and integrity of information

6.1.1 Implication for institutions

AI maturity is accelerated when governance fluency and practical skills are distributed across the enterprise rather than concentrated in specialist teams. Organizations may benefit from pairing broad AI literacy programs (e.g., "AI for all"), champion networks, and responsible-adoption guardrails with cross-functional participation from risk, privacy, legal, technology, and business units. This co-ownership model helps safeguards adapt with workloads and regulatory expectations, reinforcing resilience and trust while enabling customer value.

6.2. Practical implementation: Lumi and the Network Platform

BMO's implementations emphasize employee augmentation, context integrity, and embedded governance. The following examples are used to illustrate intentional design choices in the development and deployment of the capabilities rather than promote specific tools:

- **Lumi Assistant (employee facing)**: a generative assistant focused on retrieval of policies and procedures to improve time to-answer. Client data are intentionally out of scope; responses carry versioning and citation metadata to support traceability. Reported outcomes relate to service quality and operational efficiency, with attention to colleague experience [BMO (2025b)].
- **One client relationship network platform (commercial)**: advanced analytics and AI unify relationship signals to support proactive engagement. Recognitions have highlighted analytics depth and connected insights across treasury and payments contexts, illustrating the role of networked data models in commercial banking [Globe Banner (2024)].
- **My financial progress (retail)**: a goal planning experience (with Conquest Planning) that adapts recommendations as client circumstances change. The design emphasizes transparency of scenario logic to align personalization with explainability [BMO (2025c)].

Across these programs, the recurring design patterns are employee augmentation, scope-controlled retrieval, and horizontal connectivity under responsible adoption guardrails. In parallel, BMO has stated an intent to measure human-AI collaboration (experience and outcomes) rather than rely exclusively on technical accuracy metrics [Milchanowski (2025)].

6.2.1 Implication for institutions

Reliable AI at scale requires a governed horizontal data foundation with provenance telemetry, complemented by measurement of human–AI interaction (experience, trust, outcomes) as primary indicators. Accuracy remains necessary but is insufficient on its own to ensure operational reliability.

6.3. Integrating governance and collaboration for sustainability

Scaling agentic systems responsibly calls for actions that anticipate evolving regulatory expec-

tations and embed resilience within AI operations. The principles below are framework agnostic and intended to support durable, future-ready capabilities across ML, GenAI, AI Agents, and agentic architectures:

- **Elevate the human–AI partnership**: establish a Human–AI Interaction Scorecard measuring adoption, task quality, safe interventions, and sentiment alongside business Key Performance Indicators (KPIs) [Milchanowski (2026)]. Define thresholds and escalation paths so trust and experience are actively managed. Position human oversight as a central mechanism of responsible innovation.
- **Institutionalize provenance and auditability**: automate versioning for datasets, prompts, and agent actions. Maintain immutable logs to support reproducibility and explainability, enabling effective challenge processes by risk and business teams. Transparent recordkeeping strengthens regulatory confidence.
- **Operationalize horizontal data signals**: surface freshness, lineage, and compatibility in a unified data foundation. Introduce dynamic gating and safe defaults when signals weaken; define recency SLAs with human handoffs. Ensuring agents act on current, contextual, and governed inputs is essential to mitigate systemic risk.
- **Embed privacy by design and purpose bound access**: default to minimal access; apply redaction and guardrails as needed; and restrict retrieval to the stated task. Carry policies with data so protections remain portable across channels. These practices align with emerging privacy mandates and support trust.
- **Institutionalize co owned governance**: formalize cross functional stewardship, data, risk, compliance, business, and technology with clear responsibility assignments, practitioner enablement, and rehearsed runbooks. Treat governance as a living capability that evolves with workloads and regulatory landscapes.

6.3.1 Implication for institutions

Organizations that center the human-AI partnership and combine it with horizontal data signals, portable privacy, and reproducible auditability are positioned to scale agentic systems with purpose under rigorous oversight. This approach advances customer value while safeguarding resilience, transparency, and regulatory alignment.

These strategic imperatives position BMO and similar institutions to advance reliable, scalable AI under rigorous oversight.

7. Conclusion: The Strategic Path Forward

The transition from predictive models to agentic intelligence is a strategic shift in how financial institutions manage risk and deliver value. Production experience indicates a simple reality: even sophisticated models cannot overcome a fragmented or stale data foundation. Hallucinations are diagnostic signals of upstream misalignment in data foundations.

For the C-suite and senior leadership, the mandate is clear: prioritize data layer observability and governance over incremental model tuning. This requires:

- **Investment in observability**: prioritize continuous data quality telemetry and source to agent lineage to detect and remediate drift in near real-time.
- **Institutionalize regulatory rigor**: align controls with supervisory expectations for lifecycle validation, effective challenge, privacy, and auditability. Treat compliance as a catalyst for scalable reliability rather than a constraint.
- **Enable autonomous stewardship**: design AI agents to act as data stewards, shifting manual cleanup to self-healing, policy aware pipelines that quarantine or defer actions when signals are weak.
- **Elevate talent**: build an AI-fluent workforce across domains (risk, privacy, operations, and engineering) that can challenge, calibrate, and govern autonomous systems in production.

A resilient, trusted horizontal data foundation reduces hallucinations at scale and unlocks the speed, scope, and auditability required to compete in a digital-first economy. Organizations that treat data integrity and observability not as checkboxes, but as preconditions for reliable autonomy, will set the pace in Agentic AI.

References

BCBS, 2013, "Principles for effective risk data aggregation and risk reporting (BCBS 239)," Basel Committee on Banking Supervision, January, https://tinyurl.com/yc27uxp5

BMO, 2025a, "BMO ranked #1 globally in AI talent development," BMO News & Insights Blog, BMO Financial Group, https://tinyurl.com/46t3thbc

BMO, 2025b, "Meet Lumi Assistant: BMO's award winning AI tool empowering Team BMO," BMO News Insights, BMO Financial Group, June 17, https://tinyurl.com/ezdphnuj

BMO, 2025c, "BMO's My Financial Progress digital platform empowers Canadians with personalized goal planning experience," BMO Newsroom, BMO Financial Group, July 10, https://tinyurl.com/3wnh3spb

Challapally, A., C. Pease, R. Raskar, and P. Chari, 2025, "The GenAI divide: State of AI in business 2025," MIT Project NANDA, https://tinyurl.com/4e57cwm4

Citi, 2024, "AI in finance: bot, bank and beyond," Citi GPS: Global Perspectives & Solutions, June, https://tinyurl.com/4svkb77y

EDM Council, 2017, "FIBO: Financial Industry Business Ontology," https://tinyurl.com/4stjsd7d

Fed, 2011, "SR 11 7: Model risk management," Board of Governors of the Federal reserve System, April 4, https://tinyurl.com/2d8yrcw5

Globe Banner, 2024, "BMO's treasury solutions earn four major awards highlighting digital innovation," December 11, https://tinyurl.com/mrywtrem

Kirubakaran, A. M., A. Parthasarathy, N. Saksena, R. S. Bodala, A. Deshpande, S. Malempati, S. Carimireddy, and A. Mazumder, 2025, "Governing cloud data pipelines with agentic AI" working paper

May, R. D., 2023, "How HSBC fights money launderers with artificial intelligence," Google Cloud, November 29, https://tinyurl.com/2s32een3

Milchanowski, K., 2025, "BMO Financial Group: bridging intelligence - measuring how well humans and AI interact," Opinion, Newsweek, September 16, https://tinyurl.com/yptd5da5

Milchanowski, K. L., 2026, Return on intelligence: a strategic enterprise playbook for scalable AI agents, first edition, Routledge

NIST, 2023, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," National Institute of Standards and Technology, U.S. Department of Commerce, January, https://tinyurl.com/ybs9e325

OCC, 2011, "Supervisory guidance on model risk management," Bulletin 2011 12, Office of the Comptroller of the Currency, April 4, https://tinyurl.com/49t8m45j

OSFI, 2025, "Guideline E 23—Model risk management (effective May 1, 2027)," Office of the Superintendent of Financial Institutions, September 11, https://tinyurl.com/2s3j6xh2

Owens, E., B. Sheehan, M. Mullins, M. Cunneen, and B. O'Connell, 2025, "On the transposition of FAIR data principles to financial services: an adapted FAAIR guideline," Accounting, Finance & Governance Review 34

Priyanka, K., S. Priyadharshini, N. A. Vignesh, A. Hameed, R. Reddy K, and Praveenkumar C., 2025, "Self-healing data pipelines: reinforcement learning for real-time fault detection and recovery," presented at the 2025 International Conference on Metaverse and Current Trends in Computing (ICMCTC)

Seibert, H., S. Riemer, B. Gehra, J. Sobott, L. Gerold, F. Hildebrand, N. Gittfried, and G. Lienke, 2025, "A faster path to scaling GenAI in banking compliance," Boston Consulting Group, https://tinyurl.com/5eamvmy8

Sukharevsky, A., D. Kerr, K. Hjartar, L. Hämäläinen, S. Bout, V. Di Leo, and G. Dagorret, 2025, "Seizing the agentic AI advantage," McKinsey & Co., https://tinyurl.com/47x45d42

How data and AI shape financial services

Data culture: the silent hero of data value creation

Gary Paul,
Partner and U.K. Data Practice Lead,
Projective Group

Keeley Miles,
Chief Financial Officer, FCE Bank PLC

ABSTRACT

This article argues that data culture constitutes a foundational organizational capability - one that AI cannot replace. It explores the concept of data culture in depth, examines its strategic importance within financial services, analyses the barriers to its development, and proposes practical mechanisms for cultivating and sustaining it. Drawing on academic literature, regulatory guidance, and industry practice, the article demonstrates that data culture is central to effective decision-making, innovation, trust, and long-term resilience.

1. Introduction

Over the past two decades, data has become a defining resource of organizational competitiveness. Advances in digital technologies, cloud computing, advanced analytics, and artificial intelligence (AI) have transformed how organizations generate, store, and process information. In the financial services sector in particular, data-intensive activities underpin every core function, including credit assessment, risk management, fraud detection, regulatory reporting, Financial Analysis and Planning, customer engagement, and capital allocation.

Yet, despite unprecedented levels of investment in data platforms and AI-enabled technologies, many financial institutions struggle to translate these investments into sustained performance improvements. Research consistently demonstrates that technology alone is insufficient to generate value; rather, organizational context, leadership, and culture play decisive roles in determining outcomes. This gap between technological capability and realized value is forcing organizations to search beyond the hype with increased focus on **data culture** as an enabler for value creation.

Data culture refers to the shared values, norms, practices, and behaviors that shape how data is perceived, trusted, and used within an organization. While AI can automate analysis and enhance predictive capability, it cannot substitute for the human and organizational factors required to interpret insights, exercise judgement, and act responsibly. In regulated and risk-sensitive sectors such as financial services, these cultural dimensions are particularly salient.

AI cannot replicate or replace the perspective that exists in an organization about how serious it is about being “data driven”. Top-down mandates from executives to embrace data and AI and realize value is common, yet the very same people fail to adopt it themselves, creating doubt and leaving staff to question if everyone is on the same journey.

Organizations that successfully build a data culture are more adaptable, resilient, and innovative in the face of disruption. Conversely, those that neglect it risk falling behind, even if they have access to vast amounts of information. More data is not necessarily an indicator of success, organized relevant data that can be easily understood is the secret to a successful data culture within an organization.

2. Conceptual foundations of data culture

2.1 Culture as an organizational construct

Organizational culture has long been recognized as a critical determinant of behavior and performance. Schein (2017) defines culture as a pattern of shared assumptions learned by a group as it solves problems of external adaptation and internal integration. These assumptions shape how members perceive reality, think, and act. Importantly, culture operates beneath the surface, influencing behavior even in the presence of formal rules or strategies.

Data culture can, therefore, be understood as a specific manifestation of organizational culture, one that governs how data is valued, interpreted, and used. Unlike technical capabilities, which can be acquired through investment, cultural capabilities emerge through sustained leadership behavior, shared learning, and reinforcement over time.

2.2 Defining data culture

Data culture may be defined as the organizational environment in which data is consistently regarded as a strategic asset, trusted as a basis for decision-making, and integrated into everyday practices across functions and hierarchies. In financial services, this definition necessarily incorporates additional dimensions, including regulatory compliance, ethical responsibility, and risk discipline.

A strong data culture does not imply blind reliance on quantitative outputs. Rather, it reflects an environment in which data-informed judgement is valued, where evidence, context, and professional expertise are integrated rather than treated as competing sources of authority.

2.3 Core dimensions of data culture

The academic and practitioner literature identifies several recurring dimensions of data culture:

- **Data accessibility**: employees must be able to access relevant data efficiently and securely. Restricted access reinforces dependency on specialists and undermines democratization.

- **Trust and data quality**: trust is foundational. If data is perceived as inaccurate or inconsistent, it will be ignored, regardless of analytical sophistication.
- **Data literacy**: individuals must possess the skills to interpret data, understand assumptions, communicate insights effectively and know how to turn that into actionable outcomes.
- **Decision rights and accountability**: clarity regarding who uses data, how it informs decisions, and how outcomes are evaluated.
- **Collaboration and knowledge sharing**: cultural norms that encourage cross-functional insight sharing rather than data hoarding.
- **Ethics and governance**: norms and controls that ensure responsible, transparent, and fair use of data.

Ultimately, data culture is about aligning people, processes, and technologies to treat data not just as a byproduct of operations but as a strategic asset.

3. The strategic importance of data culture in financial services

3.1 Improving decision-making quality

Financial institutions operate in environments where decisions carry significant financial, regulatory, and reputational consequences. A strong data culture enables organizations to move beyond intuition or legacy practices and adopt evidence-based proactive and strategic decision-making.

For example, a retail bank seeking to refine its lending strategy may traditionally rely on historical credit policies and expert judgement. Without a strong data culture, decisions may be influenced by organizational hierarchy or outdated assumptions. By contrast, a data-driven approach enables the integration of transactional data, customer behavior, macroeconomic indicators, and alternative data sources to enhance credit risk assessment and portfolio performance.

3.2 Enabling innovation in products and services

Data serves as a foundational input for innovation in financial services, supporting the development of personalized products, dynamic pricing models, and digital customer experiences. A strong data culture ensures that innovation is embedded across business lines rather than isolated within innovation labs or analytics teams.

For instance, insurers increasingly use behavioral and telematics data to design usage-based insurance products. Institutions with mature data cultures are better positioned to operationalize these insights responsibly, aligning innovation with regulatory and ethical expectations.

3.3 Enhancing agility and organizational resilience

The financial services sector is extremely sensitive to economic shocks, market volatility, and regulatory change. Institutions with strong data cultures are better equipped to detect early warning signals, stress-test scenarios, and respond proactively to emerging risks.

During periods of market disruption or economic downturn, banks with well-integrated data capabilities have demonstrated greater agility in adjusting credit policies, liquidity management strategies, and customer support measures. In contrast, institutions with fragmented data environments often struggle to respond in a timely and coordinated manner.

3.4 Building trust, transparency, and accountability

Trust is fundamental to financial services. Internally, trust in data enables employees to rely on analytics for decision-making. Externally, transparent, and responsible data practices are critical for maintaining confidence among regulators, customers, and investors.

For example, banks with strong data cultures are better positioned to provide clear, explainable insights into credit decisions, capital adequacy, and risk exposure. They are also able to more efficiently respond to regulatory changes and create a more responsive and transparent relationship. A solid data culture underpinned with excellent quality data, continues to be a key focus of all regulators globally. Similarly, insurers and asset managers that

use data transparently in pricing and performance reporting enhance stakeholder confidence and regulatory credibility.

3.5 Sustaining competitive advantage

In an increasingly digitized financial ecosystem, competitive advantage is intricately linked to the ability to harness data effectively. Leading financial institutions differentiate themselves through superior risk management, personalized customer engagement, and operational efficiency, all underpinned by strong data cultures.

Institutions that fail to embed data culture risk being outperformed by competitors that leverage advanced analytics, real-time insights, and AI-enabled decision-making to respond rapidly to customer and market demands.

3.6 Decision-making under risk and uncertainty

Financial services decision-making is characterized by uncertainty, complexity, and asymmetric information. Decisions regarding credit, underwriting, investment, and liquidity management involve trade-offs between risk and return, often under time pressure.

Studies demonstrate that data-driven decision-making is associated with superior performance outcomes, including productivity, profitability, and risk-adjusted returns. However, these benefits are contingent upon cultural readiness. Where data is mistrusted, poorly understood, or subordinated to hierarchy, its impact is muted and advantage quickly eroded with untimely or mis-informed human intervention.

A strong data culture enables financial institutions to challenge assumptions, test scenarios, and evaluate outcomes systematically. Importantly, it also encourages post-decision learning, allowing organizations to refine models and policies over time. The need to continually evolve and develop the data culture is as critical as having one itself; inertia is a silent killer, and such is the pace of advancement in data that a strong organization can lose competitive advantage quickly.

3.7 Risk management and regulatory compliance

Risk management lies at the heart of financial services. Regulatory frameworks such as the Basel Committee's "Principles for risk data aggregation and risk reporting" explicitly recognize the importance of data quality, governance, and organizational accountability [BCBS 2013)].

Institutions with weak data cultures often struggle to meet regulatory expectations, not due to lack of data, but due to fragmented ownership, inconsistent definitions, and limited understanding of data lineage. Every organization should understand the best source of truth and how the data flows through their organization to create trust and consistency. Having data that can be relied upon as the best source of truth should be easy to do, but many organizations still struggle with prioritizing this against other competing organizational priorities. Institutions with strong data cultures make priority calls and investments in these areas, meaning they are better positioned to produce accurate, timely, and explainable risk information, thereby enhancing supervisory confidence.

3.8 Innovation within regulatory constraints

Innovation in financial services is inherently constrained by regulation, risk appetite, and ethical considerations. Data culture plays a critical role in enabling responsible innovation by ensuring that analytical experimentation is aligned with governance frameworks.

For example, the development of AI-driven credit models requires not only technical expertise, but also cultural acceptance of model validation, explainability, and bias mitigation. Institutions with mature data cultures are better equipped to balance innovation with compliance, whereas those with weak cultures often experience tensions between innovation teams and risk functions.

3.9 Trust as a competitive asset

Trust is central to the functioning of financial systems. Customers entrust institutions with sensitive data and financial resources, while regulators rely on accurate reporting and responsible conduct. Data culture directly influences the credibility of an institution's actions and disclosures.

Transparent data practices, explainable decision-making, and ethical use of analytics contribute to institutional trust. Conversely, data misuse or opaque algorithms can rapidly erode confidence, leading to reputational damage and regulatory intervention.

4. Data culture and AI: complementarity, not substitution

4.1 The limits of technological determinism

A recurring misconception in both academic and practitioner discourse is that AI adoption will automatically result in better decisions [Schwaller (2025)]. This reflects a form of technological determinism that underestimates the role of human judgement and organizational context.

AI systems are trained on historical data, reflect embedded assumptions, and require interpretation. Without a strong data culture, organizations risk over-reliance on model outputs or, conversely, rejection of insights that challenge established beliefs.

4.2 Explainability and human oversight

Regulators increasingly emphasize explainability, particularly in credit decisioning and risk models. Explainability is not merely a technical property, it is a cultural one. Organizations must value understanding over automation and encourage questioning rather than deference to algorithms.

Data culture, therefore, mediates the effectiveness of AI, ensuring that automated insights are integrated responsibly into decision processes.

5. Barriers to developing data culture in financial services

Financial services firms that aim to develop data culture within their organizations face the following challenges:

- **Organizational inertia and legacy norms**: Many financial institutions possess deeply entrenched decision-making norms shaped by experience, hierarchy, and professional judgement. While expertise is valuable, over-reliance on precedent can limit openness to data-driven insights.
- **Siloed structures and fragmented ownership**: functional silos - between business units, risk, finance, and technology - impede the development of shared understanding. Data ownership fragmentation often results in inconsistent definitions and duplicated effort.
- **Data literacy gaps**: despite widespread availability of dashboards and reports, many employees lack the skills to interpret analytical outputs critically. Research indicates that superficial desk-top training programs are insufficient; sustained practice and contextual learning are required, and people need to be given the time to invest to develop their skills. All the tools in the world and learning about them do not solve data literacy gaps; organizations need to invest in ensuring that the underlying corporate data is well understood, traceable, and accessible. Without these basic underlying needs, it is unrealistic to expect staff to use data effectively and drive real value.
- **Leadership inconsistency**: leadership misalignment is a major impediment. When some leaders champion data-driven approaches while others rely on intuition, cultural signals become inconsistent, slowing adoption and creating divisions between those staff who work for leaders who embrace data and those who do not.
- **Short-term rationing mentality**: restricting licenses, tools, and environments for everyone to experiment prevents adoption and the ability for all in an organization to develop better data skills. Cost barriers (especially in cloud-based solutions) can create “have” and “have not” mentalities from data elitists and

budget holders who feel costs for less data literate staff do not warrant returns. Ignorance about the longer-term potential benefit for the organization for everyone to be more data literate can be a secondary consideration to hitting arbitrary annual budget numbers. Many elitists also forget they too had to start out building their skills from the ground up!

To overcome these challenges, organizations must focus on several foundational elements:

- **Leadership as cultural architects**: leadership sets the tone for organizational culture. When leaders prioritize data-driven decision-making, employees are more likely to follow suit. Leaders must not only advocate for the use of data but also model it in their own decision-making and demonstrate that they are using it.
- **Technology and infrastructure**: robust data culture requires appropriate infrastructure - from data warehouses and analytics platforms to modern tools that facilitate self-service analytics. However, technology alone is insufficient, it must be aligned with organizational goals and accessible to employees.
- **Developing data literacy as a core capability**: effective data literacy programs are role-specific, iterative, and embedded in day-to-day work. Executives require interpretive skills, not technical depth, while frontline staff need operationally relevant insights.
- **Collaboration and sharing**: breaking down silos is key to cultivating a strong data culture. Encouraging cross-functional collaboration ensures that insights are shared across departments, leading to more holistic decision-making. The use of AI and data clinics in financial services as an example to help "patients" in organizations resolve challenges in the same way as doctors. Prescribing both immediate solutions as well as longer term "well-being" plans help everyone to improve "data health" and reduce costs and duplication.
- **Incentives and reinforcement**: celebrating successes and rewarding employees who embrace data-driven practices reinforces the desired behaviors. Recognition helps to normalize data usage and motivates employees to integrate data into their workflows.

6. Measuring data culture maturity

Data maturity has long been measurable across a number of different dimensions with many proven and tested frameworks available. Enhancing these traditional assessments to include behavioral, capability, and outcome dimensions helps create a richer assessment in organizations.

Indicators include analytics adoption, cross-functional collaboration, literacy assessments, velocity to generate insights, and improvements in risk and performance outcomes. When combined, they collectively provide a good indicator of the "data culture health" of an organization, which can be measured by Division or Function to understand if adoption is consistent.

7. The future of data culture in financial services

Looking ahead, the importance of data culture will only continue to grow. Several trends highlight this trajectory:

- **AI and machine learning**: as AI becomes more widespread, a strong data culture will be essential to ensure that organizations use these technologies responsibly and effectively.
- **Data democratization**: empowering non-technical employees to access and analyze data will become increasingly important.
- **Regulation and compliance**: with growing concerns about privacy and data ethics, organizations will need cultures that prioritize responsible data usage.
- **Integration with sustainability**: data will play a crucial role in helping organizations meet environmental, social, and governance (ESG) goals.

In the future, organizations that neglect data culture will find it increasingly difficult to remain competitive.

8. Conclusion

Data culture represents a foundational organizational capability that cannot be automated or outsourced. In financial services, where trust, risk, and accountability are paramount, it underpins effective use of data and AI alike. Organizations need to consider their data and AI strategies together in the future, not as siloed strategies but aligned to business strategy and business outcome. The sins of the past, where vast data lakes and robotic automation would resolve all our challenges, failed because they were technology led and did not fully consider desired business outcomes or underlying cultures and adoption.

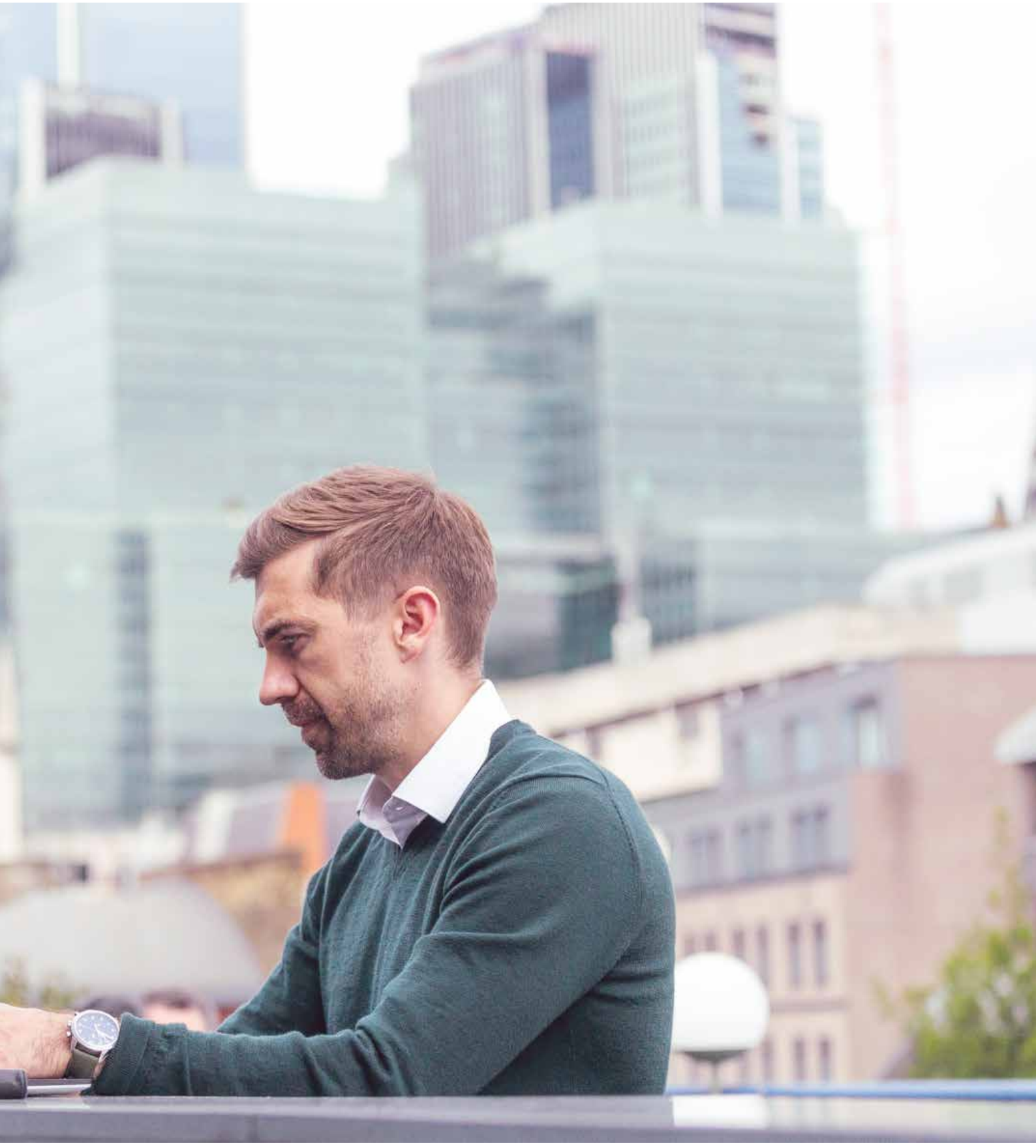
While technologies evolve rapidly, culture changes slowly. Institutions that invest in leadership engagement, literacy, governance, and collaboration will be best positioned to navigate uncertainty, harness innovation responsibly, and sustain long-term performance. AI may enhance analytical power, but it is data culture that determines whether insight becomes impact.

References

BCBS, 2013, "Principles for effective risk data aggregation and risk reporting," Basel Committee on Banking Supervision, January, https://tinyurl.com/5feh4bzr

Schein, E. H., 2017, Organizational culture and leadership, 5th edition, John Wiley & Sons

Schwaller, F., 2025, "Will AI improve your life? Here's what 4,000 researchers think," Nature, April 9, https://tinyurl.com/mry43cwa



How data and AI shape financial services

Data Governance: the tide that lifts all data initiatives

Shahina Khan,
EMEA Chief Data and Analytics Officer,
SMBC Group

ABSTRACT

This article examines the evolving role of data governance in financial services amid rapid AI advancement and intensifying regulation. It highlights the dual dynamic of AI for data governance, where AI automates discovery, quality monitoring, lineage, and metadata enrichment, and data governance for AI, where trusted, well-governed data underpins compliant and explainable AI systems. Driven by frameworks such as BCBS 239, GDPR, DORA, NIS2, and the E.U. AI Act, institutions must strengthen metadata, ownership, and lineage across complex, multi-vendor environments. The paper proposes practical, phased strategies, promotes governed data products, and envisions AI-enabled governance ecosystems that deliver regulatory compliance, operational resilience, and scalable, trustworthy AI innovation.

1. Introduction

The financial services industry is witnessing a convergence of two powerful forces: an **AI revolution** and a **wave of new regulations**. On one hand, rapid advances in artificial intelligence (AI), especially large language models (LLMs) and automation tools, are offering novel ways to manage and leverage data. On the other, regulators (particularly in the E.U.) are raising the bar for how financial institutions govern and control their data, from risk reporting to privacy and operational resilience. At the nexus of these trends lies **metadata management**, the discipline of managing data about data (definitions, lineage, quality metrics, ownership, etc.). Once a niche concern of data architects, data governance has now become a strategic priority for financial institutions, essential for both extracting value from AI and meeting stringent compliance demands.

This article explores how data governance is evolving amid increasing AI capabilities and regulatory complexity. We examine how data Governance needs to be practical with two critical and intertwined dynamics: (1) AI for data governance, how new AI tools can accelerate and enhance data governance practices; and (2) data governance for AI, how robust governance, in turn, enables trustworthy and compliant AI deployment. We also discuss the primary E.U. regulatory drivers (from ECB supervisory requirements to GDPR and the E.U. AI Act) that are pressuring firms to get their data in order, the challenges of implementing data governance in complex organizations, and practical strategies to embed governance into business processes. Finally, we look ahead to the future, including the rise of "governed data products", the role of AI agents and why a strong metadata foundation is critical for next-generation AI.

2. Challenges

2.1 Regulatory pressure: compliance as catalyst for data governance

For financial institutions, effective data governance is no longer optional, it is mandated by an expanding web of regulations. Over the past decade, supervisors have raised the bar, requiring banks to know and manage their data with precision. A prime example is the ECB's Single Supervisory Mechanism (SSM) [ECB (2024)]. When a bank becomes a "significant institution" under direct ECB oversight, it must pass a rigorous "comprehensive assessment" and maintain ongoing compliance. Compounding this is the fact that the ECB effectively mandates compliance with BCBS 239 [BCBS (2013)] as part of SSM. The ECB expects institutions to strengthen data architecture so that risk, finance, and regulatory reports are consistent, reliable, and audit-ready, supported by enterprise-wide data dictionaries, clear ownership, and robust metadata management.

BCBS 239, introduced after the financial crisis, remains the cornerstone of risk data aggregation and reporting. Its 14 principles demand accurate, timely, and comprehensive risk data, driving cultural change towards strong governance and architecture. These expectations permeate European supervision: FINREP [E.U. (2015)] and COREP [EBA (2011)] reporting require unified definitions, while national regulators enforce similar standards.

Beyond risk, new regulations broaden the scope. The Digital Operational Resilience Act (DORA), effective January 2025, focuses on cybersecurity and continuity but also mandates data accuracy, completeness, and governance [EP/EC (2022a)]. Firms must implement quality controls, audits, and detailed records of processing activities, including third-party data risk management. Similarly, the NIS2 Directive obliges institutions to inventory information assets and secure data flows as part of critical infrastructure protection [EP/EC (2022b)].

Privacy laws add another layer. GDPR requires organizations to map personal data, track metadata, and maintain a centralized catalogue of processing activities [EP/EC (2016)]. Responding to a data subject access request (DSAR) demands rapid retrieval across systems - possible only with well-governed metadata and lineage. Global banks also juggle other regimes like California Consumer Privacy Act of 2018 [CCPA (2018)], making automated classification essential.

Capital markets face their own pressures. MiFID II enforces strict retention and reporting for trade and communications data, requiring consistent lineage from front office to regulator [EP/EC (2014)]. Outside the E.U., frameworks like the Sarbanes Oxley ACT (SOX) demand complete reconciliation of financial reporting data [U.S. Congress (2002)]. Meanwhile, anti-money laundering (AML), know your customer (KYC), and counter-terrorism financing (CTF) rules push banks to integrate and monitor customer and transaction data in real

time. Regulators increasingly expect firms to "know your data" in the AML context , understanding definitions, sources, and transformations [Protiviti (2015)]. As Protiviti (2015) notes, effective AML compliance hinges on mastering metadata and lineage to justify suspicious activity reports.

The newest frontier is the E.U. AI Act, the world's first comprehensive AI regulation [EP/EC (2024)]. In 2026, high-risk AI systems, such as credit scoring or AML monitoring must comply with strict data governance requirements. Article 10 mandates that training, validation, and testing datasets be "relevant, representative, free of errors and complete," supported by documented governance practices. Firms must show how data was collected, curated, and checked for bias, ensuring full traceability from raw data to model output. AI cannot be a black box fed by unknown data; regulators will demand its pedigree. Meeting these obligations requires robust metadata repositories, lineage, and quality controls, making governance a prerequisite for AI innovation.

In short, overlapping regulations - from SSM and BCBS 239 to DORA, NIS2, GDPR, MiFID II, SOX, and the E.U. AI Act - are converging on one principle: strong data governance. Institutions that invest now in comprehensive frameworks will not only stay compliant but gain a strategic edge, enabling trusted analytics and AI in a highly regulated world.

2.2 Internal challenges in a complex data landscape (people, process, and technology)

While regulations provide external impetus, the internal reality of many financial institutions makes enterprise data governance challenging. Banks and insurance companies today operate in extremely complex data environments. Typical large institutions have data spread across old on-premises legacy systems and modern cloud platforms, across multiple business lines and geographic regions, and often use a patchwork of vendor tools for data management. This heterogeneity makes it difficult to get a unified view of metadata. For instance, a bank might use one vendor's solution for its data catalogue, a different tool for data lineage, and various databases and data lakes each with their own metadata stores. Without integration, metadata remains siloed, mirroring the underlying silos of the organization.

Multi-cloud and hybrid infrastructure add another layer of complexity: critical data may reside in an on-premises core banking system, a private cloud data warehouse, and a SaaS CRM, all at once. Ensuring consistent data definitions and quality across such diverse platforms is nontrivial. Yet, consistency is exactly what regulators and business leaders expect; for example, that terms like "customer" or "probability of default" (PD) are defined uniformly in risk reports and customer databases. Managing a central business glossary that syncs with all these systems is a major undertaking.

2.3 Third party data providers

Another pain point is the involvement of third-party data providers and outsourcing. Modern institutions do not only create data internally; they ingest significant data from external vendors (market data, credit ratings, KYC data feeds, etc.) and may outsource data processing or reporting functions. This raises the question: do vendors supply metadata (data definitions, lineage) along with the data itself and are they governing it to the standards the financial institution expects? Increasingly, regulators say they should. Under DORA's third-party risk management, for example, a bank must ensure that an outsourcer or supplier meets the bank's own data standards. In practice, if a bank relies on a fintech for, say, AML transaction monitoring, the bank is expected to know the fintech's data processing and controls in detail, effectively extending data governance across organizational boundaries. Similarly, if a data vendor provides a feed of ESG scores or market prices, banks are now pressed to understand how those figures were computed (the "data lineage" on the vendor side) and how reliable they are. This is leading to contractual requirements for data transparency. But aggregating external metadata into internal catalogues remains an open challenge.

2.4 Volume and velocity of data

Traditional manual governance approaches strain to keep up with today's data scale. An international bank may have tens of thousands of data elements and pipelines, updated constantly. Maintaining accurate metadata manually is error-prone and tedious, a problem that, as discussed below, AI can help ameliorate. Without automation, metadata repositories can become stale as systems change. Likewise, ensuring quality metadata is tough when underlying data quality is imperfect. Many institutions discover issues like inconsistent codes, missing values, or mismatched definitions during

governance projects, issues that require going back and fixing source systems, which can be expensive. Despite these challenges, the direction is clear: complexity or not, regulators expect firms to have command of their data environment.

The rest of this article discusses how to meet these expectations pragmatically and even turn them into an advantage.

3. Opportunities and strategies

3.1 Embedding data governance into the organization and processes

To truly reap the benefits (and avoid the pitfalls) of enterprise data governance, banks must weave governance into the fabric of how the organization operates. It is not enough to publish a data governance policy or establish a governance team, ingraining practices into daily workflows, IT projects, and business decision-making is critical. Here are some strategies and best practices for doing so.

3.1.1 Design for practicality and culture

A governance framework should align with the organization's culture and ways of working. For agile, product-driven firms, embed lightweight checks in sprints rather than heavy quarterly reviews. If decentralization is valued, a federated model with local data owners may work better than a top-down structure. The adage "If you build it, they might not come" applies, rolling out strict processes will not ensure adoption. Involve business users in designing governance so it solves real pain points and adds value. For example, link documentation to outcomes: "documenting lineage expedites regulatory approvals" makes the benefit clear. Keep processes simple at first and automate wherever possible, such as capturing technical metadata directly from databases.

3.1.2 Phased implementation and tiered maturity

Recognize that domains differ in governance maturity. Start with a pilot or critical area, such as regulatory reporting or a dataset that caused issues. Demonstrating success (e.g., fewer data errors after governance measures) builds momentum and lessons learned. Then expand in phases. Set realistic interim targets based on each domain's baseline: one division might define and approve all critical data elements in three months, while a less critical area starts with its top 10. Different plans for different parts of the business are fine, as long as there is a unified vision and eventual convergence. This incremental approach avoids overload and lets the central team better support each wave.

3.1.3 "People change" enables the roles

Assigning governance roles (owners, stewards, custodians) is common but must be backed by enablement. Simply naming someone a steward will not deliver well-governed data. They need training, time allocated for the role, and management support. Invest in skills through training and communities of practice, such as open "data surgeries" where practitioners solve challenges together. Embed data responsibilities into performance objectives: if a manager is accountable for data quality, make it part of their goals. Some firms add incentives or recognition, like awards for improved data quality metrics. Management buy-in is critical: when direct managers reinforce these tasks and allow time, cultural adoption improves. Studies show Executive data ownership and accountability lead to increased success [Ninio et al. (2025).

3.1.4 Leverage tools, but fully

Selecting and deploying a governance tool is often an early step. Many enterprise catalogues exist, but a common mistake is under-using their capabilities, banks sometimes treat them as glorified spreadsheets, listing data elements without leveraging lineage mapping or issue tracking. To avoid this, explore the tool's full features and align processes to use them. Even if it is not perfect, a formal tool meeting 80% of needs beats ad-hoc spreadsheets. Understand its capabilities, get vendor support or expert help early, and configure it for workflows. For example, use APIs to sync metadata with development pipelines or store data quality rules for central monitoring. For more upcoming functionality, see Section 3.2.

3.1.5 Fail fast, and iterate

Governance is not a one-and-done project, treat it as iterative. Not every process will work perfectly at first, and that is fine. Encourage a "fail fast" mindset: if a metadata review workflow is too cumbersome, acknowledge it and refine. Better to adjust than enforce something that drives poor compliance. Hold periodic retrospectives, perhaps quarterly

forums, to discuss what works and what does not. This adaptability improves effectiveness and signals governance is about continuous improvement, not rigid bureaucracy. Balance flexibility with consistency: core principles (e.g., "document definitions for all critical data") should remain firm even as tactics evolve.

3.1.6 Embed governance in the data lifecycle

Integrate governance into every stage of projects and the data lifecycle, not as a separate task. For all data development, new pipelines, data products, or onboarding applications, governance should be "baked in". Use project checklists: is the data model in the catalogue? Are owners identified and definitions approved? Is lineage logging in place? Do this during the project, not after deployment. Apply the same principle to changes and related processes (e.g., replacing a data owner in the movers and leavers process). Many organizations now adopt "data governance by design," ensuring new or updated systems and pipelines meet governance requirements from the start and throughout their life.

3.1.7 Governance where it matters

For existing data, focus governance where it matters most: prioritize key datasets, those feeding many critical reports or analytics, for metadata enrichment and quality improvement. This two-pronged approach (govern new things, gradually fix important old ones) avoids "boiling the ocean". For example, procurement can embed governance: when contracting a data vendor or SaaS tool, include a review to map data and ensure metadata and security provisions. This way, vendor management becomes part of governance, with contracts linked to data domains in a central register. Over time, such practices make considering data and metadata standard in every initiative. Done right, governance moves from abstract policy to operational reality, delivering regulatory compliance and internal benefits like efficiency, better decisions, and fewer errors. But manual effort will not scale, which is why AI plays a key role.

3.2 AI for data governance

Traditionally, maintaining metadata has been a labor-intensive endeavor: armies of analysts manually defining terms, documenting data flows, and reviewing data quality. Now, however, we are at an inflection point where AI technologies can turbocharge data governance, automating many tasks, and providing intelligent insights. In fact, a virtuous cycle is emerging: AI can accelerate data governance, which in turn creates better data foundations for more advanced AI (a topic we discuss in the next section). Here we focus on how AI is being leveraged to make metadata management faster, more comprehensive, and more proactive than ever before.

3.2.1 Automated data discovery and classification

One of the most immediate uses of AI in this domain is automated data discovery and classification. Financial institutions deal with huge volumes of data spread across numerous repositories, and manual cataloguing of all this data is nearly impossible. AI techniques like natural language processing (NLP) and machine learning (ML) can comb through data sources to identify what data is there and apply initial tags or classifications. For example, AI algorithms can scan database schemas, file contents, or even free text documents to detect sensitive information (like personal data indicators for GDPR) or categorize data by type (customer data, transaction data, logs, etc.).

3.2.2 Data quality management

AI also enhances **data quality management**, a key aspect of data governance. Traditional data quality checks (for completeness, consistency, accuracy, etc.) rely on predefined rules and periodic reports. AI techniques, particularly anomaly detection and pattern recognition, can monitor data continuously and flag issues that static rules might miss. For instance, machine learning models can learn what "normal" data patterns look like (say, typical ranges or distributions for certain fields) and raise alerts when incoming data deviates significantly.

3.2.3 AI for data lineage

One particularly promising area is applying AI to metadata enrichment and data lineage. In many organizations, metadata exists but is incomplete; for example, a data element might have a technical name and source system, but no clear business definition, owner, or relationship details. Generative models can help fill these gaps by drafting business-friendly descriptions, inferring relationships from code-to-text analysis, and even suggesting classifications based on upstream sources. Modern data catalogues now embed AI that mines

query logs, schemas, and data patterns to surface insights such as "analysts who use Dataset X frequently join it with Dataset Y," making implicit joins and dependencies explicit. Beyond enrichment, AI-driven lineage engines can automatically infer column-level flows from SQL and ETL patterns, reconciling lineage across tools and clouds.

3.2.4 Active metadata

The concept of active metadata is rising, powered by AI. Rather than static documentation in a repository, active metadata means the metadata system is continuously updated and provides real-time recommendations. For instance, an AI enabled data catalogue might observe that no one has queried a certain data table in a year and recommend flagging it as a deprecation candidate. Or it might see many users repeatedly adding the same filter when using a dataset and suggest capturing that filter as a formal data quality rule or a new derived field in the metadata. Vendors now incorporate such features, their catalogue can suggest improvements to metadata based on how users interact with data. This makes governance more dynamic and usage-driven, as opposed to a one-time documentation effort.

3.2.5 Data governance AI agents

One of the most exciting developments is AI "copilots" for governance roles. Imagine a governance AI agent available to every steward, analyst, or business user. Instead of digging through manuals or catalogues, users could ask: "where does this field come from?", "what does "exposure at default" mean and who owns it?", or "show me all reports using this dataset?" Connected to the organization's metadata knowledge base, the agent could instantly answer with links. This lowers barriers and democratizes data knowledge. Agents can also proactively alert users: "this dataset has a known quality issue" or "a newer version is available." Early versions exist, chatbots built on data catalogues and conversational interfaces for governance. Giving every stakeholder an AI assistant empowers informed decisions without relying solely on a central team.

3.2.6 AI governance workflows

AI is transforming governance by automating routine tasks and reducing manual effort. Access control can use AI to suggest roles based on usage patterns, flagging anomalies for review. Compliance checks can run continuously, detecting misplaced personal data or verifying encryption, enabling real-time audits. Gartner predicts that by 2027, 60% of organizations will be unable to realize the full value of their AI initiatives due to incohesive data governance [Sinha (2025)]. Conversely, firms using AI-driven governance report up to 30% less time on compliance reporting, proving efficiency gains offset investment.

Introducing AI requires care: algorithms need training on sparse metadata, and oversight is critical. A human-in-the-loop approach works well: AI proposes classifications, stewards review high-impact decisions until trust builds. Privacy matters too: tools should process metadata, not raw content, to avoid exposing sensitive identifiers.

Ultimately, AI accelerates governance by automating discovery, quality checks, and metadata enrichment, while making it more engaging, think chat interfaces instead of static wikis. As AI handles the grunt work, data teams can focus on strategy, creating the reverse dynamic: governance enabling better AI.

3.3 Data products as an enabler for users and AI accessibility

The relationship between AI and data governance is symbiotic. Yes, AI can improve data governance, but equally, strong data governance supercharges AI initiatives [Arnold and Everaert (2025)]. Financial institutions are discovering that their ambitious plans for AI and advanced analytics run into a basic roadblock: messy, poorly understood data. The motto "garbage in, garbage out" is as true of AI models as it was of traditional reporting. Thus, investing in data governance is not just a compliance or data management move, it is becoming a strategic prerequisite for AI success. Banks that have curated, high quality, well understood data can deploy AI faster, with more confidence and transparency, than those still wrangling unknown data sources.

3.3.1 Data discoverability and readiness

One immediate way that good governance accelerates AI is through data discoverability and readiness. Imagine a data scientist trying to build an ML model to predict customer churn or credit risk. In a poorly governed data environment, that data scientist might spend 80% of the time just finding and cleaning data: searching for relevant datasets, guessing meanings of fields, writing custom code to reconcile inconsistencies, etc. In contrast, in a

well-governed environment, the data scientist can quickly, in a single interface, search the enterprise data catalogue for datasets and marketplace for data products, looking for, "customer transactions last 2 years" and get a curated dataset with clear documentation.

A robust metadata-driven architecture means that structured and unstructured data are harmonized for both human and AI consumption. Leading organizations are building centralized data marketplaces where all "AI ready" datasets are published with their metadata. Modern governance tools supporting metadata-driven architecture allow enterprises to surface trusted data in seconds and integrate it seamlessly, reducing the time to find and prepare data drastically. This instant discovery of relevant, quality data sets, accelerates AI development cycles. Instead of spending months on data cleanup, teams can rapidly prototype models, enabling more experiments and faster iteration toward business value.

3.3.2 Trusted data

Trusted data yields trusted AI. Well-governed data makes AI outputs more reliable and explainable. For example, a risk model for capital reserves becomes auditable if every input is documented with lineage and quality metrics, addressing AI's "black box" issue. Governance adds context: an AI-driven lending decision can include a report of inputs, sources, update dates, and caveats, enabling regulators and auditors to trust the process.

Mature governance also mitigates bias by flagging issues like class imbalance or missing demographics before training, allowing data scientists to compensate and document fixes. The E.U. AI Act reinforces this: Article 10 mandates full lineage and proof of data quality for training datasets. Firms with BCBS 239-style governance will find compliance easier, reusing existing frameworks to ensure AI training data is "error and bias free".

3.3.3 Ontologies and knowledge graphs

Ontologies and knowledge graphs, advanced metadata capturing relationships and semantics, are increasingly important with AI's rise. LLMs excel at free-text understanding but lack domain specific reasoning. Ontologies, formal representations of domain knowledge, provide context and logic that pure ML misses. In essence, they are structured metadata. By building ontologies (e.g., for financial instruments or risk concepts like exposure, collateral, default), organizations create a knowledge layer that strengthens AI.

We are moving toward neuro-symbolic AI systems [Vsevolodovna and Monti (2025)], combining neural networks with symbolic reasoning. Research shows that pairing LLMs with ontological reasoning improves consistency and factual accuracy: LLMs generate fluent outputs, while ontologies enforce rules. For example, an ontology can encode "all loans must have an interest rate and maturity date," guiding AI assistants and reducing errors. Curating ontologies and knowledge graphs now, as part of governance, prepares financial institutions for this next wave of AI [Balestriero and LeCun (2025)].

3.4 Third-party vendor data integration: from data feeds to data products

Adopting a data product approach offers a powerful framework for integrating third-party data within financial institutions while maintaining intellectual property boundaries. Instead of treating external data feeds as raw inputs, vendors deliver them as fully managed data products. These data products come equipped with functional metadata layers that explicitly capture input/output relationships, transformation types, and business context, enabling banks to seamlessly incorporate third-party data into their governance frameworks.

This approach decouples the what (functional lineage) from the how (proprietary implementation logic), allowing vendors to share coarse-grained lineage maps that satisfy regulatory requirements such as BCBS 239 and DORA without exposing trade secrets. By packaging data with clear metadata definitions and standardized APIs, data products facilitate transparent data access, auditability, and quality assurance downstream, while vendors retain control over their "secret sauce."

Further trust can be built through independent third-party certifications or code escrow arrangements, which assure banks about the validity and compliance of underlying algorithms without direct exposure. Ultimately, embracing a data product model transforms third-party data from a compliance risk into a governed, audited, and reusable asset, streamlining regulatory adherence and fostering innovation.

4. A Practical application using data products

Banks increasingly rely on external market data vendors to enrich analytics and decision-making. Imagine an internal AI powered data marketplace where consumers use natural language to discover governed data products, complete with rich metadata: standardized descriptions, PII classification, ontology-mapped definitions, lineage, quality scores, remediation history, and role-based access controls. This metadata is vital for compliance and efficiency, yet vendor data often arrives in incompatible formats with incomplete metadata.

Vendors should be adopting the same data product paradigm, packaging datasets with metadata aligned to open standards like OpenLineage for lineage, DCAT or Schema.org for descriptions, and industry ontologies for finance. Metadata could include machine-readable files detailing schema, definitions, lineage, quality metrics, compliance tags (GDPR, DORA, BCBS 239), and links to ontologies. Proprietary logic is referenced at a business-rule level, protecting IP while enabling governance.

Delivery is streamlined: vendors expose data via secure APIs or encrypted transfer, bundling metadata in standard formats. The bank ingests both, auto-mapping definitions to its ontology, validating lineage, and integrating quality scores into dashboards. AI agents can then answer queries about vendor data, definitions, lineage, and quality, without manual effort. Contracts mandate metadata standards, periodic quality updates, and lineage documentation, balancing transparency with IP protection.

This approach transforms external data management: trust and transparency rise, manual effort drops, and regulators see end-to-end governance, even for third-party data. Both banks and vendors gain efficiency, reduce risk, and build scalable, governed ecosystems. Ultimately, this is the future of financial data: interoperable data products, intelligent marketplaces, and contracts that enable compliance and competitive advantage in a data-driven world.

5. Conclusion: the future is governed data products and AI synergy

In financial services, data has always been an asset but in the years ahead, governed data will be the true differentiator. Regulatory compliance, operational excellence, and AI-driven innovation all rest on the same foundation: **robust data governance**. Institutions that succeed will transform what looks like a compliance burden into a strategic advantage, turning governance into trusted data products and knowledge assets that fuel new digital services.

European regulations, ECB guidelines, BCBS 239, GDPR, DORA, and the E.U. AI Act are pushing firms to elevate governance practices. Leading organizations see this not as overhead but as an opportunity to build enterprise-wide frameworks that deliver business value. By breaking silos, embedding governance into daily processes, and tackling complexity, they create environments where data is understood, controlled, and ready for use.

AI amplifies this imperative. Without well-governed data, AI initiatives risk failure or flawed outcomes. Conversely, governance and AI reinforce each other: better governance enables better AI, which in turn improves governance. Expect a maturity curve where basic metadata management evolves into intelligent ecosystems, self-updating catalogues, AI assistants guiding data access, and policies enforced automatically. Financial institutions will develop “data governance brains”: systems that continuously optimize data usage, like traffic control for information.

The rise of “governance AI agents” will embed AI into governance workflows, and governance into AI workflows. AI will not only consume data but actively participate in governance: flagging issues, suggesting improvements, and negotiating access under preset rules. As AI’s role grows, regulatory scrutiny will extend from data governance to AI governance. Regulators will demand lineage not just for data but for AI decisions. Firms with strong metadata practices will be ready, connecting the dots from data to model to decision.

The goal is clear: a data environment where most used data is a governed data product, quali-

ty-checked, secure, discoverable and every AI is a governed consumer, monitored and accountable. Achieving this unlocks "digital trust at scale", enabling rapid innovation with confidence and demonstrable compliance.

Governed data will become a standard unit of value, much like financial products today. These data products, enriched with metadata and accessible via secure APIs, will power internal analytics and new client services, such as delivering insights as a service, backed by trusted data.

Ultimately, data governance may not make headlines like AI, but it is the critical infrastructure for an AI-driven, highly regulated future. By investing now, through AI tools, improved processes, and closing regulatory gaps, financial institutions can position themselves as both compliant and competitive. The destination is a future of governed, intelligent, and trusted data ecosystems, where governance is not an afterthought but an intrinsic quality of data and a pillar of strategic advantage. Organizations that embrace this will turn information into a decisive edge.

References

Arnold, A., and E. Everaert, 2025, "Springboard your AI journey through our data product approach," Projective Group, October 9, https://tinyurl.com/3edthf87

Balestriero, R., and Y. LeCun, 2025, "LeJEPA: provable and scalable self-supervised learning without the heuristics," working paper, November 11, https://tinyurl.com/4t5a46nu

BCBS, 2013, "Principles for effective risk data aggregation and risk reporting," Basel Committee on Banking Supervision, January, https://tinyurl.com/37e85vjb

CCPA, 2018, "California Consumer Privacy Act of 2018," https://tinyurl.com/2wr2wzbt

EBA, 2011, "Guidelines on Common Reporting – Revision 2 (2011)," European Banking Authority, https://tinyurl.com/nkpec6yx

ECB, 2024, "Overview of our supervisory framework," European Central Bank, https://tinyurl.com/dkux7s7h

EP/EC, 2014, "Directive 2014/65/EU of the European Parliament and of the Council on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU," Official Journal of the European Union, May 15, https://tinyurl.com/2hzsrs79

EP/EC, 2016, "Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)," Official Journal of the European Union, April 27, https://tinyurl.com/y5xaw3hf

EP/EC, 2022a, "Regulation (EU) 2022/2554 of the European Parliament and of the Council on digital operational resilience for the financial sector and amending Regulations (EC) No 1060/2009, (EU) No 648/2012, (EU) No 600/2014, (EU) No 909/2014 and (EU) 2016/1011," Official Journal of the European Union, December 27, https://tinyurl.com/39xjupa7

EP/EC, 2022b, "Directive (EU) 2022/2555 of the European Parliament and of the Council on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive)," Official Journal of the European Union, December 14, https://tinyurl.com/bdh88h39

EP/EC, 2024, "Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)," June 13, https://tinyurl.com/musfevps

E.U., 2015, "Regulation (EU) 2015/534 of the European Central Bank of 17 March 2015 on reporting of supervisory financial information (ECB/2015/13)," Official Journal of the European Union, March 31, https://tinyurl.com/44r4efjr

Ninio, B., J. Holdowsky, C. Iannacone, D. Kearns-Manolatos, and P. Mackie, 2025, "Five leadership and teaming choices that can help drive increased value," Deloitte, August 26, https://tinyurl.com/3ffpnpm2

Protiviti, 2015, "AML and data governance: how well do you KYD?" https://tinyurl.com/27hun38e

Sinha, M., 2025, "How AI is transforming data governance: automation, metadata, and beyond," April 8, https://tinyurl.com/yf55659r

U.S. Congress, 2002, "H.R.3763 - Sarbanes-Oxley Act of 2002," July 24, https://tinyurl.com/zhpvca4j

Vsevolodovna, R. I. M., and M. Monti, 2025, "Enhancing large language models through neuro-symbolic integration and ontological reasoning," working paper, April 10, https://tinyurl.com/3a566uus

How data and AI shape financial services

The leadership of **cyber resilience**: from controls to choices[1]

Martijn Dekker,
Professor of Business and Cybersecurity,
University of Amsterdam, and CISO,
ABN AMRO Bank N.V.

ABSTRACT

We live in an increasingly digital, complex, and volatile world. Organizations need to be able to prevent, absorb and respond to shocks caused by disruptions of digital services. These are, for example, due to ransomware attacks on suppliers or unavailability of cloud services. In this article, we will explain that the required digital resiliency is not only a technical problem, but a leadership and decision-making topic and hence requires new or improved governances. Cyber resilience is not only about controls, but about choices and how we make them. We conclude with providing practical guidance to improve the leadership of cyber resilience.

1. This article reflects the author's personal views.

1. Introduction

Ever since the invention of the Internet in the last century, we have witnesses a rapid rise in complexity and scale of cyberspace. The number of connected devices has grown astronomically and nowadays practically all organizations and businesses rely for their daily operations on the availability and trustworthiness of the digital domain. This ever-growing technology estate needs to be protected properly, and a large industry of security vendors has emerged.

To protect digital assets like servers, workstations, and data, extensive cybersecurity frameworks have been developed. A well-known (and much used) cybersecurity framework is the Cyber Security Framework (CSF) of the American National Institute of Standards and Technology (NIST) [NIST (2025)]. This framework has grown over the years into a set of more than 1200 control and control objectives.

Despite significant investments, cybersecurity incidents continue to disrupt services and cause data to be stolen or lost. This is not primarily a technical failure, but a leadership and decision-making issue. In fact, the focus on technical implementations can cause new operational risks. The cyber system itself is becoming very complex and hence fragile; in fact, the security systems themselves can bring about massive global incidents sometimes [see, for example, the global security disruptions caused by CrowdStrike; FCA (2024)]. These and other trends lead to an emerging phenomenon called “cyber senescence”. The topics we discuss in this article can also help in managing cyber senescence [Dekker (2025)].

Next to the increase in control frameworks, we see a significant amount of cybersecurity regulation coming into force. In Europe, we have seen the Network and Information Security Directive (NIS2), the Digital Operational Resilience Act (DORA), the Cyber Security Act (CSA), and others. In the U.S., we have seen Health Insurance Portability and Accountability Act (HIPAA) in 1996, in the healthcare sector, Gramm-Leach-Biley Act (GLBA) in 1999 for the financial sector, Cybersecurity Information Sharing Act (CISA) in 2015, and many others.

Investments by organizations in cybersecurity have grown very fast. These security products use advanced technologies like detection supported by artificial intelligence (AI) and automated response scripts. They perform regular risk assessments, security tests, and execute self-assessments to ensure regulatory compliancy or to measure maturity against frameworks like NIST CSF. To implement any of these processes or mitigate any of the identified security issues, multiple vendors promise technical solutions and services. There is no shortage of solutions in the market, and security investments have grown to become a significant component of most organizations’ IT spend. In fact, Gartner predicts security spend to grow by 12.5% in 2026, as compared to 2025 [Scroxton (2025)].

Next to technical controls, the governance of security decisions has been recognized as a key element in cyber resilience. For example, in the last NIST CSF update, a new category of “governance” controls was added. Cybersecurity clearly requires implementation and operational effectiveness of many technical controls. But to be resilient against cyber incidents requires fast and coordinated decision making under uncertain conditions. That is why cyber resilience is not a purely technical topic, it is also about leadership capabilities of the organization.

Hence, cyber resilience is not just an extension of cybersecurity into a wider technology domain, it is also an extension into a decision-making and leadership discipline. Organizations can fail during a disruption not because their systems were not protected properly, but because their leaders were unprepared to make effective decisions in a timely manner. Subsequent to an organization managing to make its basic cybersecurity mature, becoming cyber resilient is not just a technological issue but predominantly a leadership challenge.

2. The boundaries of control-centric cybersecurity

Cybersecurity started many years ago as an attempt to prevent security incidents from occurring. If a computer system operates in a simple environment, one could try to enumerate all insecure states and then implement controls that prevent them from occurring. Although this may work in very simple cases, this approach has proven to be naïve. The digital environment is not simple but very complex and enumeration or *a priori* identification of unsecure states is not possible. This means leadership needs to cope with low predict-

ability and high uncertainty.[2] Security specialists have recognized that controls need to be strengthened, and although secure-by-design principles in software development methodologies make sense and are best-practice, security incidents or unintended behavior cannot be ruled out. For this reason, prevention, detection, and response controls have been added to information security management systems. Next to controls, risk management has become part of cybersecurity, to prioritize security investments. In Dekker (2025), we argue that the digital world has become so unpredictable that risk management is only a rough approximation, and we need to find better ways to deal with uncertainty in the cyber domain to increase resiliency.

The current cyber threat landscape is characterized by at least three realities:

1. **Inevitability**: incidents are inevitable, and one must accept breach,
2. **Uncertainty**: it is hard to know the likelihood or the impact of security incidents. Likelihoods are hard to determine not only because of lack of information, but since incidents can be emerging phenomena caused by complexity or by humans with unknown motives. The impact of incidents includes elements that are hard to measure, like loss of trust or reputation,
3. **Speed**: cyber incidents can occur very fast, and they can escalate more quickly than governance processes.

Many security-frameworks are not well-equipped for these conditions. Managing security metrics like patching rates, survival-rates of software vulnerabilities or operational effectiveness of key controls are necessary but not sufficient for resiliency. A good example of a set of such metrics is the BOOM framework [Seiersen (2022)]. These metrics do not provide insights into how the organization will function when systems are unavailable or compromised, and, more importantly, they do not provide leaders with the means to effectively make decisions, both under normal circumstances and under stress.

3. Cyber incidents: decision making under stress

It requires a mature security implementation to be able to timely detect a security incident. This includes a high number of implemented security controls, but also, to be resilient, it should ensure optionality to provide different response scenarios. After detection, the incident response mechanisms require fast and effective decision making between different options, so cyber incidents can be viewed as a decision crisis. Decision makers are often confronted with many difficult and impactful responses with little available information, high uncertainty, and conflicting objectives.[3] Examples of such choices are:

- Should the compromised system be shutdown to stop the incident from becoming bigger, or should it be kept running to limit customer impact?
- When and how should we inform stakeholders, like customers, the press, or supervisory authorities?
- In what order should we start restoring business services in case of, for example, a denial-of-service attack?
- What trade-offs should we make between restoring systems and forensics?

Clearly, these choices are not only of a technical nature. These decisions have business, legal, financial, reputational, and trust implications. Consequently, these decisions should be made under pre-agreed conditions like decision ownership, authority, and risk tolerance. This is why cyber resilience is also, next to a technological and organizational capability, the capability of an organization to make and execute these decisions effectively, both under stress as well as under normal conditions. A resilient organization is prepared to make these decisions timely and effectively. To be prepared requires low decision latency, strong leadership alignment, and good information without blind spots. It also requires both testing decision processes and measuring preparedness. We will discuss these items below.

2. van den Berg (2024) identifies five ways to deal with uncertainty in cyberspace: increase control, increase resilience through risk management and adaptability, manage uncertainty through regulation or through suspension, and finally by simply ignoring it. We live in a world where ignoring uncertainty is not an option anymore.
3. Even under non-stress situations, decision makers face amplified agency problems in the cyber domain [Cortez and Dekker (2022)].

3.1 Decision latency

The time between the moment a security alert is triggered, and a meaningful response has been decided upon, is called "decision latency". It is well understood that this latency is an important element of cyber incident management. All three threat landscape realities mentioned above impact decision latency.

Decision latency arises from the following conditions that can occur in any organization:

- **Unclear authority**: unclear ownership between IT, security, business, legal, and risk leadership.
- **Escalation overhead**: heavy reliance on approval processes and formal committees for operational decisions.
- **Information distraction**: too many detailed metrics and dashboards distracting leadership from what needs to be decided.
- **Fear**: concerns about regulatory or reputational consequences leading to avoidance of decision making.

For cybersecurity incident decision making, latency is urgent because the three threat landscape realities add further pressure. The first (inevitability) makes decision latency very relevant. The second one (uncertainty) complicates decision making and hence most likely increases latency, as decision makers will seek more information or wait for authorization from others. The third one (speed) plays a role, as cyber incidents often spread fast and impacts quickly grow, so time is of the essence and low latency is required.

Most organizations implement IT incident and crisis management via an escalation model. Incidents can trigger an incident response process that activates an incident response team. When that team needs help or the incident is growing, the team can escalate to an "amber" status and activate a Business Continuity Team (BCT) to involve business leadership and others. If the event escalates further, status "red" can be invoked and a crisis management team led by senior management activated. If an incident threatens the continuity of the organization, in some cases a status "black" is also part of the escalation ladder.

Even when crisis management governance is clear, well defined, and formalized, and the people involved are regularly trained, activating the properly authorized response team takes time. In practice, it can take hours or even days after detecting an event for a cyber incident to be escalated to the right level (for example code red) for authorized decision making. For some cybersecurity incidents, the number of options or decisions that can be effectively implemented after that time is very limited and damage is irreversible. Consequently, the speed of decision-making is as important as technical recovery capabilities. The longer it takes to make decisions, the smaller the solution space of possible choices. Given the high level of uncertainty in cybersecurity, a large, prepared, and ready-to-use solution space is required to have sufficient optionality to react to an unpredictable future. Before we look at how to deal with this, we will look at the type of decisions that are relevant in incident response.

3.2 Decisions and leadership alignment

As argued above, cyber resilience requires a shift of focus from technical control implementations to decision making frameworks. One needs to shift from thinking in terms of controls, to thinking in terms of choices and how to make them. For example, to move beyond DORA compliancy, and really become cyber resilient, one should not only ask "how secure are we" but also "how prepared are we to decide when incidents occur"?

In other words, next to having effective preventative security controls (to reduce the amount of security incidents), effective detection and response controls (to timely discover and activate responses), fast and effective decision making is required to determine how to recover from an incident. Recovery can involve restoring data, restarting servers or applications, or any other technical implementations. It is important to realize that the classical response strategy to return to the last good state may no longer suffice. We now live in a world where a shock can change the environment permanently and the "last-good-state" is no longer acceptable or even possible.[4] In that case, a response to create a "new-good-state" is the only solution. To define a decision framework for cyber resilience, one should answer questions like:

4. An example of such a shock could be the legally forced ejection of foreign high-risk suppliers from critical sectors [Moreau and Hartmann (2026)].

- What are the business critical or important functions?
- In what order should services be restored when disrupted?
- Who is authorized to make these decisions and under what conditions?
- How quickly can decisions be executed or implemented?
- Can we fall back to last-good-states, or do we have the ability to create new good-states under stress?
- What roles do third parties play in these decisions?

Regulations like DORA have recognized some of these challenges and have made it mandatory for European financial institutions to proactively define their critical and important functions (CIF). For CIFs, special regulatory security requirements are defined. Implicitly, DORA expects resilience-by-design principles to be applied, embedding resilience characteristics into technology, processes, and leadership structures. However, there is very little information about how that can be done. In particular, it does not describe how decision making should be organized in case a CIF is part of a security incident or disruption. Most regulations simply make executive leadership liable and accountable for cyber resilience and leave it up to the organization to implement adequate governance to support that accountability.

Cyber resilience cannot be completely delegated to technology teams. Clearly, cybersecurity specialists play an important role, but they require guidance regarding decisions made by senior leadership. Consequently, the alignment between the security leadership and the business leadership is becoming more urgent. Unfortunately, there is a gap between these two leaderships as both are using very different languages [Cortez and Dekker (2022)] and, in most organizations, they rarely meet. Cyber resiliency requires this gap to be closed and to provide agency to the leadership. Hence it requires:

- Shared understanding of the cyber risks and business implications.
- A priori agreement of priorities during incident responses.
- Clear ownership of decisions, authority under described condition, and escalation processes.
- Defined tolerances and risk appetites.

Resilience requires optionality, hence having a large solution space ready is a good thing. However, a big solution space needs maintenance and can increase costs and reduce agility, which can even reduce resiliency. Resilience decision making is about making tradeoffs between enlarging the prepared solution space and agility and cost. In case of an incident, resilience decision making is about choosing responses. When autonomy for this is placed low in the organization, decision latency will be lower. The possible impact of the chosen response can, however, be so significant that one would like higher management to be involved. Resilience governance must, therefore, include clarity about who is authorized to make these tradeoffs, both under normal conditions and under stress. Governance under normal conditions can be different from periods of stress. For example, a security officer might be allowed to switch off a critical client facing system, but only under clearly defined conditions, such as in an identified imminent threat.

We argue that security leaders should step up and lead the discussion about articulating those tradeoffs and the dilemmas involved and by doing so establishing this leadership alignment. By doing that, they will grow into the digital resiliency leadership of the organization.

3.3 Third party and blind spots

To make decisions, one should minimize information blind spots. Blind spots can be the existence of shadow IT (assets and systems that are not well managed or owned), unknown dependencies, or unknown processes. We will look at blind spots in the supply chain to illustrate what blind spots to consider and how they relate to resiliency.

Until now, we have focused on the organizations themselves. But we live in a world in which organizations have become entangled with large numbers of vendors, suppliers, and third parties. These dependencies have economic benefits as they can provide efficiency, skill, or scale. These parties have become significant parts of the attack surface of the organization; they, therefore, add complexity to security decision making.

When new vulnerabilities are published, it is difficult to assess whether the organization is impacted. So how can one even hope to know which third parties are impacted by this vulnerability? How does one know that a third party is compromised? This requires deep insights into the state of the supply

chains. To establish these insights, most organizations include requirements in their contracts to ensure rights to perform risk and compliance assessments or the obligation to provide security assurance reports. However, when a third party is compromised, questions like the following need to be answered too:

- Can we continue without a critical provider?
- What contractual rights exist to ask for information or support from the third party?
- Can we quickly exit from this provider and switch to another?
- Who is in the lead for customer communication or regulatory reporting?
- What is the ability of the provider to respond to the incident? What is their decision preparedness?

Even if the third party is fully compliant with their contractual obligations, and the security professionals have a good situational awareness of the security posture of the third party, knowing whether the supplier can make timely and effective security decisions is usually a blind spot.

Given the size of the supply chains, cyber resiliency of the organization requires mapping third-party dependencies to critical and important functions and inclusion of those relations in critical decision making. Consequently, security leaders should expand their views of the supply chains beyond understanding security control effectiveness and include understanding the decision preparedness of the supplier. These factors should play an explicit role in sourcing decisions and in the complete contract life cycle. We see that to become resilient, new types of data and factors should be taken into account in existing decision-making processes like sourcing and procurement processes.

3.4 Testing decision preparedness

Many organizations have implemented security testing programs or crisis simulations to train crisis management teams. Some have formalized ethical hacker programs or responsible disclosure policies to help identify vulnerabilities in the security setup. All these approaches require deep technical expertise and tend to be quite complex. While these technical testing programs are necessary, they are not sufficient, as they do not, or only partly, test decision preparedness. Fortunately, there are relatively simple and non-technical approaches for testing the readiness of leadership. We will mention four examples of such approaches.

Firstly, decision walkthroughs are structured discussions in which leaders are led through realistic scenarios of incidents that could happen. The goal of these walkthroughs is to identify decision points in the management of the incident. Next to that, these walkthroughs also help senior management to understand/accept the fact that these are not hypothetical, but that these scenarios can really happen.

Secondly, authority mapping is an exercise for identifying or clarifying which leader is authorized to decide what and under what conditions. For example, what are the thresholds of acceptable operational losses or what information needs to be available to whom when making decisions?

Thirdly, time-based scenarios allow for testing decision making under stress. How do decisions change when time pressure is introduced? In some cases, for example, a certain response is only acceptable if there is no time for alternative responses.

Lastly, dependency reviews can be conducted to review how third-party issues would influence options for decision making.

All these approaches are non-technical and can be executed with combined teams of security and business leadership. They can help identify gaps in authority, alignment, and decision preparedness, which would be missed in security testing or crisis simulation exercises. The key point of these approaches is to focus on how decisions are made instead of what decisions are made or what controls are in place. They help shift attention from controls to choices.

3.5 The metrics of preparedness and board involvement

Now that we have described what decision preparedness is and how it can be trained, it is also clear that new metrics should be added to the security posture of an organization. These metrics should reflect the level of preparedness of decision making under stress and, hence, are indicators of cyber resilience.

Although no common wisdom or best practice exists as yet in this domain, it seems reasonable to explore decision-centric metrics like time-to-decision during incidents, clarity of ownership of decisions, traceability of agreed priorities and actual

responses, ability to function with damaged or unavailable systems ,and balance between speed and risk.

As executive boards and leadership teams increasingly recognize cyber resiliency as part of their accountability, assessing and reporting on resilience becomes imperative. Existing security reporting tends to focus on compliance, control effectiveness, and the threat landscape. Security leadership need to step up and expand their oversight of technical detail with reporting on decision preparedness. Business leadership should, in turn, improve their communications with security leaders to include questions like:

- Are leadership decision roles clear?
- How quickly can the organization decide which business services to shut down or in what order to restore them?
- What third-party failures or external events would constrain our options?
- Have we practiced making these decisions under stress?
- Do we have the ability to create new-good-states in case of permanent changes in the environment?

By having a high-quality conversation like this, between security leaders and business leaders, executive boards can help build a more resilient organization.

4. Conclusion

Cyber resilience is often defined as the ability to withstand, absorb, or respond to cyber incidents. Given the nature of the cyber domain, this is often considered a technology problem. We have argued that building true cyber resilient organizations requires both a strong technical security posture and effective leadership decision making. Consequently, security leaders should extend their view to include decision preparedness, and business leaders should recognize their role in defining decision clarity, authority, alignment, and practice preparedness.

Cyber resilience can be reached by aligning the optionality provided for by security control implementations, and the decisions that fit the tolerances of the organization and its stakeholders.

In the highly volatile digital environment where cyber incidents are inevitable, but unpredictable and develop quickly, the most resilient organizations will be those that combine strong defenses with leaders who are prepared to decide.

References

Cortez, E., and M. Dekker, 2022, "A corporate governance approach to cybersecurity risk disclosure," European Journal of Risk Regulation 13:3, 443-463

Dekker, M., 2025, "Uncertainty in security: managing cyber senescence," https://tinyurl.com/2sn6fw8v

Moreau, C., and T. Hartmann, 2026, "Commission sets out plan to eject foreign high-risk suppliers from critical sectors," Euractive, January, https://tinyurl.com/25mb4u8u

FCA, 2024, "CrowdStrike outage: lessons for operational resilience," Financial Conduct Authority, October 31, https://tinyurl.com/yhue8nu3

NIST, 2025, "Cybersecurity framework NIST," National Institute of Standards and Technology, https://tinyurl.com/544dbf5p

Scroxton, A., 2025, "Global cyber spend will top $200bn this year, says Gartner," ComputerWeekly, July 29, https://tinyurl.com/54cbh93p

Seiersen, R., 2022, The metrics manifesto: confronting security with data, Wiley

van den Berg, B., 2024, "Dealing with uncertainty in cyberspace," Computers & Security 144: 103939

How data and AI shape financial services

Building data resilience: adapting to modern threats and evolving technologies in a financial markets infrastructure

Laure Molinier,
Director, Group Operational Resilience,
Euroclear

ABSTRACT

Modern organizations face a rapid shift from traditional, infrastructure focused resilience toward data centric resilience, driven by evolving cyber threats, increasingly complex hybrid technology stacks, and heightened regulatory expectations. Cyberattacks, including ransomware, large scale data corruption, and hybrid warfare now target data directly, posing greater systemic risk than historical physical disruptions. This article explores this evolution based on the experience of Euroclear, a globally systemic Financial Market Infrastructure that safeguards €40.7 trillion in assets. Given the magnitude of assets protected, any disruption could have global repercussions, making perfect data and robust recovery strategies essential.

1. Introduction

In today's rapidly evolving threat landscape, organizations must rethink their approach to resilience. The rise of sophisticated cyber threats, regulatory demands, and the growing strategic value of data have shifted the focus from traditional infrastructure and application recovery to data resilience. This article explores the drivers behind this transformation, including changes in technology, regulations, and the nature of modern threats.

Drawing on the experience of Euroclear, one of the world's largest Financial Market Infrastructures, this article highlights the importance of operational resilience and the need to focus on protecting and recovering critical data. Euroclear plays a vital role in global capital markets, handling trillions in assets and ensuring the integrity of securities ownership across multiple markets. The company's operations demonstrate the systemic importance of data resilience: any disruption could have global repercussions, making perfect data and robust recovery strategies essential.

Through practical examples, the article examines how Euroclear is adapting its resilience strategies to address evolving threats, regulatory requirements, and the complexities of hybrid technology environments.

2. Drivers for the shift: evolving threats, technology, and regulatory landscape

In the aftermaths of the September 11th, 2001, terrorist attacks, companies began investing heavily in systems and processes to strengthen the resilience of their infrastructures to support the recovery of their services against a broad range of man-made or natural physical disasters. They have invested in multiple data centers that are synchronously linked and permanently staffed operational centers spread across regions with different risk profiles. These resilience strategies, which have been quite effective in limiting the impact of subsequent terrorist attacks and the Covid-19 pandemic, are being challenged by the evolution of a new threat environment, predominantly involving "physical" to "logical" scenarios (e.g., ransomware and supply chain attacks). These new threat types, which impact data rather than infrastructures, are exacerbated by increasingly interconnected eco-systems and require different resilience strategies.

Hybrid warfare tactics observed in the current geopolitical environment combine conventional physical military attacks, such as attacks on undersea cables or use of drones, with non-conventional cyber and disinformation operations to undermine institutions. The scope of resilience strategies should, therefore, cover a broader spectrum of threats and scenarios.

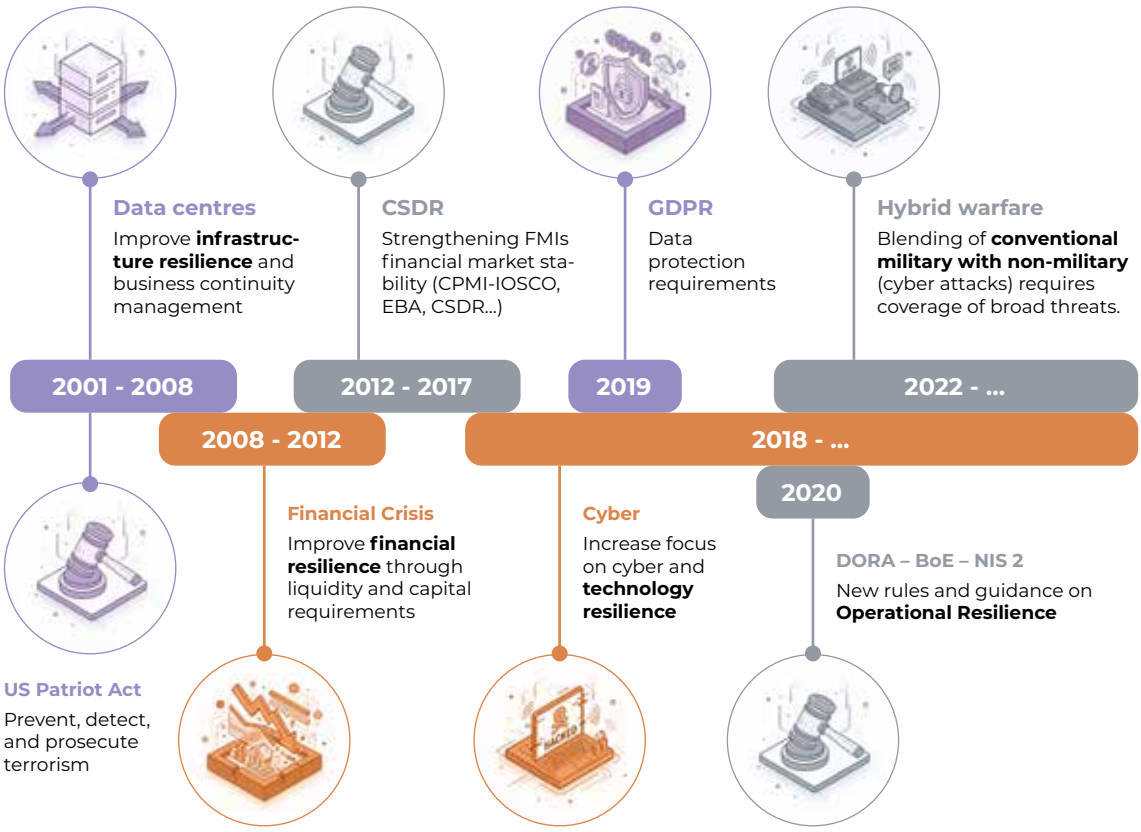
The evolution of technology includes a shift from monolithic applications and isolated systems running onsite, to hybrid systems involving multiple applications and platforms - such as mainframe, cloud, Windows, and Linux - and utilizing diverse communication channels. Key data may also be dispersed. Whilst single application services could rely on backup restore, recovery is becoming much more complex, requiring managing latency and reconciliation issues at different levels, between internal and external systems, and with the firm's ecosystem of clients, providers, and other third parties.

Finally, recent regulations across the globe, including the E.U.'s Digital Operational Resilience Act (DORA), the Bank of England's Operational Resilience Framework, the Network and Information Security (NIS 2) Directive, the Cyber Resilience Act (CRA), and the U.S. Federal Financial Institutions Examination Council (FFIEC) guidelines, place a strong emphasis on operational and cyber resilience across critical sectors.

DORA raises the bar for digital operational resilience in the financial services sector, requiring robust ICT risk management, incident reporting, resilience testing, and third-party oversight. The FFIEC guidance in the U.S., stresses enterprise-wide business continuity and operational resilience, urging financial institutions to prepare for, withstand, and recover from disruptions, with a focus on governance, risk assessment, and third-party risk management. Collectively, these frameworks reflect a regulatory shift from reactive recovery to proactive resilience, aiming to ensure the stability and reliability of critical services in an increasingly complex and interconnected environment.

Figure 1 depicts the different drivers of a shift from infrastructure to data driven resilience strategies with examples based on the experience of Euroclear.

Figure 1: From business continuity...to operational resilience



Euroclear is one of the world's largest Financial Market Infrastructures (FMI), playing a critical role in the smooth functioning and stability of global capital markets. As a trusted post-trade partner, Euroclear connects issuers and investors across more than 50 countries, facilitating the settlement, safekeeping, and servicing of securities, including bonds, equities, and investment funds. Euroclear's operations are vital for market integrity and financial stability. Every day, major financial institutions rely on Euroclear to process their transactions. Any disruption could have global repercussions, underlining its systemic significance.

The group is handling €40.7 trillion assets under custody and processing over €1 quadrillion annually. Euroclear maintains the authoritative record of securities' ownership and determines property rights for trillions of assets across multiple markets. And as settlement is legally final and irrevocable, and the integrity of positions has a ripple effect on global lending and liquidity across multiple financial systems, data must be perfect and cannot be undone.

3. Operational resilience is providing a framework to support data recovery

Operational resilience is a holistic discipline focused on ensuring that an organization can continue to deliver critical operations through disruption, whether caused by physical events (like natural disasters or terrorist attacks) or logical threats (such as cyberattacks or data corruption). It involves preparing for, responding to, and recovering from incidents that could impact business continuity. Operational resilience covers both infrastructure resilience, protecting physical assets and IT systems, and data resilience, ensuring the availability, integrity, and recoverability of critical business data.

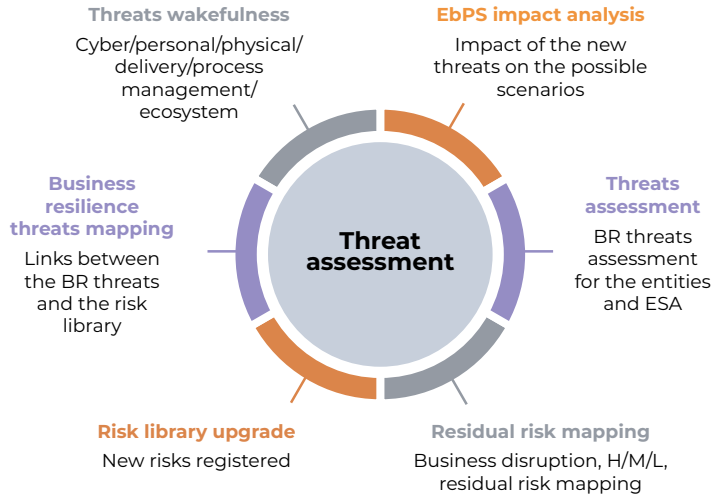
Whilst Euroclear continuously invests on protection from a very broad spectrum of threats, keeping its defense up to date to the complexification of cyber-attacks, the company also assumes that dis-

ruption will happen and that disruption is not an "if" but a "when" question. Preparing to respond to and recover from disruptions has always been Euroclear's focus and over the years the company has demonstrated its capacity to withstand severe events.

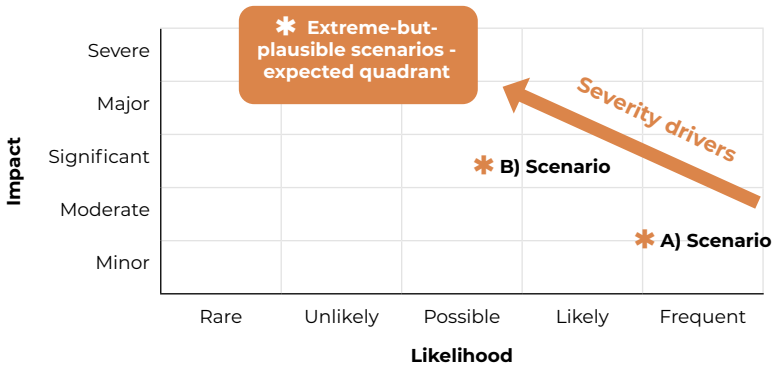
To measure, monitor, and improve its capacity to respond and recover from a wide range of disruptions, Euroclear has developed an "operational resilience framework", including regular assessments and continuous monitoring of its threats environment and identified a set of impact-based extreme-but-plausible scenarios. On a regular basis, Euroclear monitors the evolution of threats that may trigger a disruption to its critical business services and ensures that protection and detection measures are constantly adapted to reduce the likelihood of an impact. We also perform regular assessments of capabilities supporting the response and recovery from extreme-but-plausible scenarios. Euroclear defines extreme-but-plausible scenarios as impact-based scenarios covering all outage scenarios, extreme or not, and uses the following elements to model and define these scenarios:

- Scenarios are primarily based on impact rather than specific trigger, as there may be many different and/or combined triggers (cyber and non-cyber) for any given scenario.
- Scenarios relate to one or more underlying critical activity/business service.
- Scenarios have attributes that will escalate the impact of that scenario, also increasing the complexity and, generally, reducing the likelihood.

Figure 2: Extreme-but-plausible scenarios assessment framework



External and internal sources are monitored to ensure adequate threat coverage. A **vulnerability assessment** is performed to assess the inherent and residual risk of each of the threats identified as potential triggers for disruptive scenarios by looking at the existing controls in the relevant domains.



Map and understand **the operational disruptions** caused by the threats. Euroclear baselined a number of impact-based extreme-but-plausible scenarios, representing the most extreme but plausible impact scenarios, triggered by internal and external threats.

Euroclear regularly monitors and performs an annual review of the threat landscape with the purpose of identify the most relevant threats and to assess and understand its capabilities to respond to the threats that may impact the continuity of its critical business services.

Euroclear also assesses its response and recovery capabilities from impact-based extreme-but-plausible scenarios.

These scenarios build upon the "classical" business continuity loss of assets' scenarios, such as loss of staff, offices, IT infrastructure, and applications, adding a specific focus on scenarios impacting data. Euroclear monitors the recoverability of its key services against the following three generic types of impact to data:

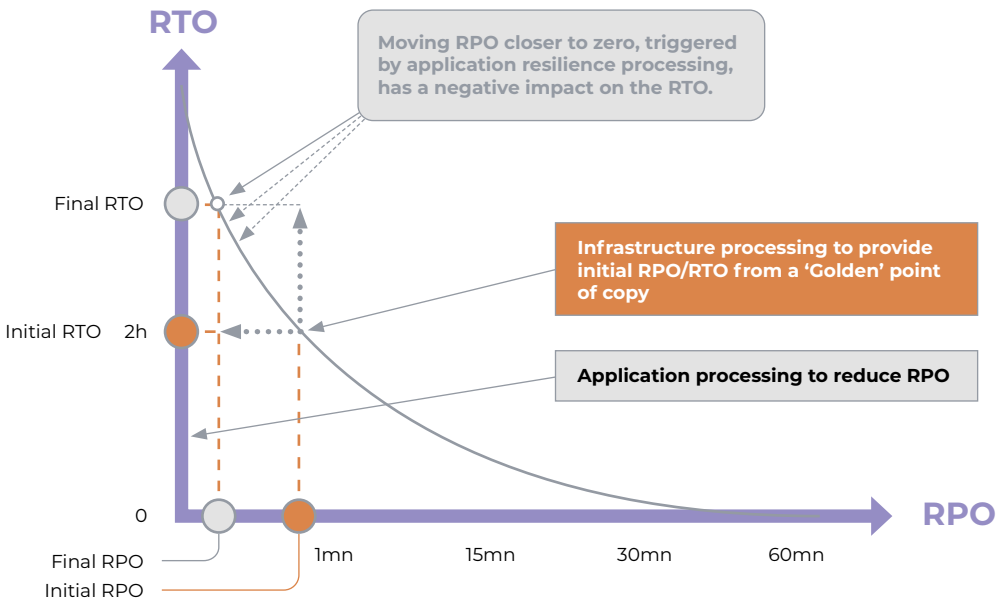
- **Confidentiality**: that can materialize through theft or leakage of data.
- **Availability**: that can result from an infrastructure or applicative incident. In the worst-case scenario, it may also be the result of a successful ransomware attack and result in loss of data.
- **Integrity or data corruption**: that can be either unintentional, due to a system or applicative issue, or intentional, resulting from a cyber-attack.

Confidentiality breaches may have major reputational implications, as well as result in fines. Euroclear is constantly adapting its security defenses to reduce the likelihood of such breaches, and to detect them. However, data leakages or theft are less likely to create a major disruption to Euroclear services, consequently the company focuses on improving its recovery strategies to address the "black swans", worst cases of major data corruption or data encryption/ransomware. In the next section, we will explain the challenges and potential solutions for achieving data resilience. Figure 2 depicts the threats and extreme-but-plausible scenarios assessment framework.

Depending on the nature of their business activities and criticality in the market they operate, organizations need to define recovery objectives. We will introduce a few key concepts that are relevant to this document:

- **RTO (recovery time objective)**: is defined as the period of time following an incident that a system, application, or process can be down after a disruption before operations must be restored and supports the organization's recovery planning after an incident to minimize business impact. The timer starts at the moment the negative impact of the incident materializes and includes detection and decision times, up until the resumption of the service up to the predefined level.
- **RPO (recovery point objective)**: stands for the acceptable amount of data loss measured in time. It defines the point in time to which data must be restored after a disruption, and guides backup frequency and data protection strategies.
- **Impact tolerance**: new regulations such as Bank of England and EU-DORA introduce the concept of impact tolerance, which is defined as the maximum level of disruption to critical services that an organization can withstand in an extreme-but-plausible

Figure 3: Speed versus correctness



scenario, without causing intolerable harm to the market, market participants, wider financial stability, or the organization's own safety and soundness. It is typically expressed as a duration (e.g., number of hours or days) or another measurable threshold and assumes the disruptive event has already occurred.

- Financial market infrastructures such as Euroclear need to make trade-offs between the speed of recovery (RTO) and the acceptable amount of data loss (RPO), as efforts to lower RTOs and availability may potentially negatively impact RPO as illustrated below.

4. Strategies for achieving data resilience

4.1 Framework and governance

To ensure proportionate investments into resilience measures, companies should understand what matters, what they need to protect and recover. To do so, they should identify what critical services (a.k.a. critical business services/important business services) they deliver to their clients and what services and assets are required to deliver them. Business impact analyses are a foundational component of strong resilience frameworks. They aim to identify critical services and recovery objectives and should cover interdependency mapping, enabling the construction of an end-to-end lineage from critical services provided to clients down to applications, IT assets, data, people, and workplaces. Interdependency mapping supports the assessment of resilience capabilities and the scoping of resilience investments. Returning to data, in today's environment, companies need to understand where their critical data are stored, which may be combined onsite with off-site locations (cloud), and implement strong data governance processes.

4.2 Technology evolutions

The evolution of technology is introducing a higher level of complexity, with many companies moving from a single on-premises technology stack to a hybrid architecture built on distributed, multi-technology environments, leading to the need to manage different data recovery strategies and requiring additional reconciliation efforts.

Technology evolution also brings great opportunities with resilient by design applications, intentionally architected to withstand, adapt to, and recover from disruptions, whether caused by technical failures, cyberattacks, or broader operational crises. Newly designed applications integrate redundancy with extra copies of critical components, automated failover, and error detection.

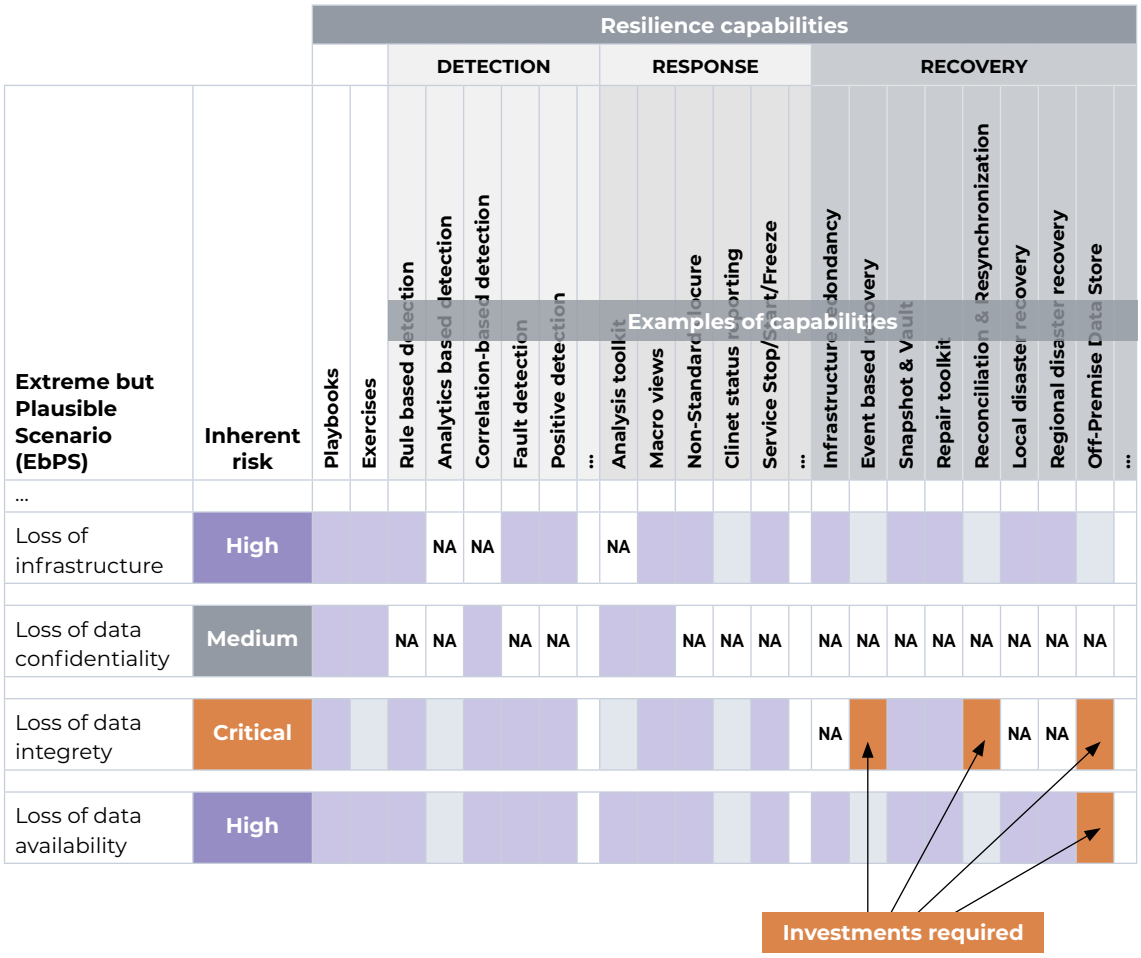
Intelligent infrastructure, automation and architecture principles like idempotency[1] may also contribute to faster data recovery as it leverages real-time data, sensors, and predictive analytics including AI and machine learning to anticipate risks and detect anomalies before they escalate. Earlier detection of data integrity issues significantly improves recovery time and reduces data loss and reconciliation needs. Automation enables systems to respond to incidents automatically, such as restarting failed services, rebalancing workloads, or rerouting traffic without waiting for human intervention. This leads to much faster recovery times and higher system availability. Intelligent infrastructure continuously monitors applications, networks, and resources, providing visibility across the entire environment. This data-driven approach supports real-time optimization, ensuring resources are allocated efficiently and systems remain resilient under changing conditions. Modern infrastructure can adapt to shifting workloads, evolving threats, and new business requirements. Intelligent automation allows systems to learn from outcomes, adjust policies, and self-heal, making resilience a built-in feature rather than an afterthought.

4.3 Data recovery strategies

Recovering services after key data have been corrupted or encrypted requires a strategy different from the traditional IT disaster recovery strategies, focused on addressing the two biggest sources of systems' failure, software bugs and failures of the underlying hardware, infrastructure and middleware components. Indeed, these strategies were all targeting a short recovery time and with data synchronously replicated across several data centers at the speed of light, they could both rapidly switch to a back-up infrastructure and restart without losing

1. **Idempotency** refers to the property of an operation where performing the same action multiple times produces the same result as performing it once, if the inputs remain the same. In practical terms, this means that if a request is repeated due to network issues, retries, or duplicates the outcome remains consistent and does not cause unintended side effects.

Figure 4: Combining scenarios and detection, response, and recovery capabilities



any data. These strategies are, by design, not providing any effective protection against data corruption, as on the contrary, corruption will get replicated at the same speed as clean data. One of the challenges will be to identify what was the last clean record to start recovering from while reducing to the minimum the potential data loss.

4.4 Companies need to adapt their thinking patterns

Let us look at a few practical approaches that should support recovery from dooms day scenarios impacting data. There is no one-size fits all strategy and even within the same business, different recovery strategies will be available, and the complexity of recovery will be highly dependent on the nature of data (e.g., more static technical data, source code versus dynamic transactional data) and of the scale of the corruption.

Whilst Euroclear has a strong recovery track-record and achieves very high systems' availability objectives for its critical services, thanks to a multiple data centers strategy, in-built resilience, and strong Business Continuity Planning, we also recognize the need to strengthen our capabilities to support recovery from extreme data integrity or availability losses in an acceptable RTO.

Euroclear has defined different data-centric recovery strategies that would be activated depending on the nature of the data impacted, dynamic data such as financial transactions versus more static systems or applicative data, and on the scale of the corruption. A strategy aiming to recover data integrity by correcting impacted data, correcting or reversing transactions, would be preferred in case of limited scale scenario as it limits the need for reconciliation with the ecosystem. This strategy would be less effective in case of wides-

cale data corruption for which data recovery from a back-up would be the best option. Depending on the scale and type of data, number of application/business services impacted, different recovery methods and back-up types would be used, from applications or database back-ups for contained impacts, to immutable data snapshots in the most extreme cases. Because recovery from back-ups would restore data at a point back in time, lost data would need to be rebuilt, and reconciliation of early transactions with the Ecosystem would be required.

What organizations should retain is that there is no one-size fits all strategy but a combination of resilient architecture patterns and tools covering the very diverse nature of incidents and disruptions (Figure 4).

5. Challenges and key considerations

In strengthening their data resilience and recovery frameworks, organizations will have to overcome a number of challenges while recognizing the need for gradual maturity increase.

- Some organizations may lack a formal data strategy or governance plan, with executives sometimes believing they can succeed without one. This leads to resistance when implementing new data governance or recovery strategies, especially as technology stacks become more complex. Without a clear understanding of the value and necessity of adapting data recovery approaches, organizations risk relying on outdated methods that are ineffective for today's hybrid and cloud environments.
- As regulatory frameworks have matured over the period across the globe with mostly similar risk reduction objectives, international companies may be confronted with **conflicting regulatory requirements** stemming from different regions. As an example, while the E.U. Cloud Sovereignty framework focuses on ensuring that data generated and stored in the E.U. is protected from foreign access, with strict requirements for data localization, privacy, the U.S. Cloud allows U.S. authorities to compel U.S.-based cloud providers (including subsidiaries operating in the E.U). to provide access to data, regardless of where it is physically stored. This extraterritorial reach means U.S. law can override local E.U. protections.
- Cultural resistance often manifests as a belief that "it never happened to us, we have strong cyber security, and if it happens, we will find our way to deal with it." This mindset leads organizations to underestimate the complexity of modern technology stacks and the risks posed by new threats such as ransomware or large-scale data corruption. For instance, when legacy systems coexist with hybrid or cloud applications, management's inaccurate perception of risk exposure and the level of investment required to achieve the desired resilience maturity may lead to resistance in updating recovery strategies, relying on outdated methods that are ill-suited for today's interconnected environments. This resistance can result in delayed response to incidents, ineffective recovery, and prolonged business disruption. A lack of understanding about the need for tailored, data-centric recovery approaches can leave organizations vulnerable, as traditional backup and restore methods may replicate corruption across systems or fail to address the challenges of dynamic transactional data recovery.
- Overcoming these challenges requires defining extreme-but-plausible scenarios, having a very good understanding of the architecture of the critical business services, raising awareness through testing and crisis exercises, and fostering a resilience culture that embraces continuous adaptation and modernization of recovery strategies.
- Alongside the definition and implementation of new company-wide data recovery strategies, capabilities, and technical tools, companies need to evolve their testing strategies. Validating the effectiveness of each technical component or IT disaster recovery and business continuity plan through isolated tests remains a pre-requisite to establish strong foundations. For technical testing, companies may think about using chaos engineering, introducing controlled failures or disruptions into systems such as server shutdowns, network failure to test its resilience and uncover weaknesses before real incidents occur.
- "End-to-end" tests and exercises are as essential for validating the effectiveness of the chain of recovery actions.

- Indeed, end-to-end testing of resiliency strategies offers several critical benefits for organizations aiming to ensure operational continuity and compliance with regulations:
  - End-to-end testing goes beyond isolated technical checks and tabletop exercises by simulating realistic disruption scenarios across the entire recovery chain, including IT systems, business processes, crisis management, and third-party dependencies. This holistic approach confirms that all components work together to restore critical services within required timeframes (RTO, RPO, impact tolerance).
  - By testing the full sequence from detection to containment, eradication, restoration, and resumption of business services, organizations can uncover overlooked dependencies and gaps in their response plans. This reduces the risk of unexpected failures or outdated procedures during actual incidents.
  - End-to-end testing provides evidence of resilience, demonstrating to regulators (e.g., under DORA) that the organization can withstand and recover from disruptions. It supports compliance with requirements for annual, comprehensive, and ecosystem-inclusive testing.
  - End-to-end testing may be purely internal or it may involve external stakeholders such as ICT providers, business partners, and clients. Organizations engaging in testing with their ecosystem, sharing their recovery options and strategies will ensure alignment across the ecosystem and strengthens collective preparedness.
  - Finally, advanced simulated tests help teams practice their roles, refine crisis management procedures, and build confidence in their ability to respond effectively under pressure. This continuous improvement culture is essential for adapting to evolving threats.

6. Conclusions and recommendations

The threats landscape has shifted from primarily physical risks to complex, logical threats such as ransomware and data corruption. This evolution, combined with regulatory challenges and the increasing value of data, requires a move from traditional infrastructure/application resilience to a focus on data resilience. Indeed, traditional backup and disaster recovery methods are insufficient for today's hybrid and cloud environments, as they may replicate data corruption and fail to address the complexity of dynamic, interconnected systems. Cultural resistance and lack of formal data governance remain significant barriers to effective resilience, with many organizations underestimating the risks posed by modern threats.

Organizations should adapt to these fast-evolving threats and technology environments and consider the following measures:

- **Adopting data-centric resiliency strategies**, shifting focus from infrastructure to data, ensuring strategies address data integrity, availability, and confidentiality across all platforms and environments.
- Developing and maintaining **strong data governance**, map critical data, understanding dependencies, and implementing robust governance processes to support effective recovery and compliance.
- Ensure **comprehensive knowledge of the architecture** of all critical business services and their interconnections, supported by a robust Business Impact Analysis (BIA) framework, to enable accurate risk assessment and effective resilience planning.
- **Leveraging technology evolution**, investing in resilient-by-design applications, intelligent infrastructure, and automation to enhance adaptability and recovery capabilities.
- Defining **tailored recovery strategies** depending on the type and scale of data impacted and ensuring reconciliation processes are in place for dynamic transactional data.
- Fostering a **culture of resilience** by raising awareness, conducting regular crisis exercises, and promoting continuous adaptation of recovery strategies.
- Implementing **end-to-end testing**, including business processes and third-party dependencies to uncover gaps and demonstrate resilience.

How data and AI shape financial services

Is ESG reporting **data fit for purpose**?

Tensie Whelan,
Distinguished Professor of Practice Emerita, Business & Society, Stern School of Business, New York University

ABSTRACT

This paper highlights the inherent tension in environmental, social, and governance (ESG) reporting: ESG reporting metrics generally do not measure outcomes, they measure outputs. Outputs, such as a policy on a sustainability topic, do not necessarily drive performance, societal or financial. This article reviews the Sustainability Accounting Standards Board (SASB) and European Sustainability Reporting Standards (ESRS) standards and metrics through the lens of driving potential societal and financial performance and makes recommendations for the standards and corporate leaders.

1. Introduction

ESG data reporting is being required by regulators, investors, and other stakeholders, due to concerns that corporate performance on those topics, when material, can have a significant impact on both society and the company's "bottom-line".

However, arguments over the utility of ESG data and reporting abound, as do academic studies and consultant reports, with strong opinions on both sides of the ledger. Regulators and investors continue to ask for ESG data, though in some important cases those requests have been rolled back entirely or become more focused (e.g., the Securities and Exchange Commission (SEC) no longer requiring climate disclosure or the Corporate Sustainability Reporting Directive (CSRD) reducing its scope). Some companies continue to collect ESG data, regardless of regulatory or investor requests, because they find the data useful for their decision-making. Others no longer report on ESG performance.

Our research at NYU Stern Center for Sustainable Business using a framework we have named the Return on Sustainable Investment (ROSI)[1] has explored whether sustainability efforts linked to material ESG issues can drive better financial performance. We should note that the definition of sustainability includes impact on profit, while ESG does not, which is why we use the term sustainability when we are not specifically talking about ESG data and reporting.

As we worked with ROSI, we found many shortcomings in commonly used ESG metrics that made it difficult to assess societal or financial performance, hence we undertook a more formal analysis of the SASB and ESRS standards.

The positive and negative critiques of ESG reporting by stakeholders are summarized briefly (and not exhaustively) as follows:

Why we need ESG data:

- There are ESG issues that may in the future, or already, have a material impact on a company's bottom line, such as climate change and biodiversity loss [Giglio et al. (205)]. That data, therefore, needs to be collected and disclosed as per any other material financial risk factor.
- Customers and consumers are looking for sustainable solutions and companies that ignore that demand may put their business at risk. For example, consumer demand for healthy and sustainable consumer packaged goods (CPG) products is growing 2.3X faster in the U.S. than conventional products, and at a 27% premium (conventional is losing market share) [NYU CSB (2024)]. Robust ESG data is, therefore, needed to demonstrate to customers that claims are real.
- Data such as energy use and cost [Ibishova et al. (2024)], health and safety incidents and related comp and lost productivity costs, waste disposal costs, and so on are necessary to optimize performance, especially as costs related to energy, natural resources, waste disposal, and employees increase.
- We need a common standard for ESG data [Rouen et al. (2022)], otherwise data quality can be compromised and regulators and investors cannot make grounded assessments or comparisons related to performance.

Why ESG data is a scam/unnecessary:

- ESG data is not material to financial performance and its collection is forcing companies to make decisions based on non-financial metrics that will drive negative impacts to their bottom line, e.g. Florida anti-ESG investing law [Malone (2023)].
- ESG data reporting requirements themselves are onerous and, therefore, cost too much to implement, e.g. first version of CSR-D requiring more than 1100 data points.
- ESG data provider rankings differ widely in their assessments of corporate ESG performance, creating confusion and questions about the validity of the data [Billio et al. 2021)].
- A lot of ESG data is not audited and may be based on problematic processes or assumptions, e.g., greenhouse gas (GhG) accounting for scope 3 emissions may have different approaches and overlapping accounting.

Our own research finds an interesting contradiction, which partially explains the differing opinions described above:

1. Most ESG reporting data is output, not outcome, based and, therefore, not likely to drive better financial or societal performance. For example, the reporting metric asks if the

1. https://tinyurl.com/4nwr34sy

company has a policy on waste management, ignoring the outcome of that policy [Whelan (2025)].

2. ESG reporting data, even when outcome-oriented, do not link to financial returns. For example, reducing GhG can lead to lower energy use, which can lead to lower cost and energy pricing volatility, lower maintenance cost, lower regulatory fees, etc., but that is not tracked or reported on.
3. Despite these challenges, our research does find that well-designed and executive sustainability strategies and practices focused on material ESG issues for a given industry/company are likely to drive better financial performance [Whelan and Douglas (2021)].

In other words, sustainability can drive better financial and societal performance, but we postulate that ESG reporting metrics are not generally the right tool.

We have researched the financial value drivers associated with sustainability strategies and practices in different industries (automotive, apparel, food and beverage, health care, real estate, and energy) and reviewed those drivers across strategies (i.e., decarbonization, circularity, sustainable sourcing, health & safety, and water stewardship). We have found that business drivers such as operational efficiency, risk mitigation, increased sales, innovation, employee productivity, and retention are often associated with effective sustainability practices.

For example, a company may determine that its operations materially impact the environment through GhG emissions. By integrating energy management into its decarbonization strategy, the company can save on energy costs, reduce exposure to energy price volatility, and potentially command market premiums. The company can use a Return on Sustainability Investment (ROSI™) [NYU CSB (2024)] process to define and monetize the financial benefits for each identified opportunity (Table 1).

Circularity provides another good example: it can drive lower virgin input costs, lower waste disposal costs, reduced dependency on geopolitically risky supply chains, less explore to tariffs or extended producer responsibility taxes, increased sales, benefits and so on (Table 2). Notably, the ESG reporting standards do not highlight those opportunities.

Table 1: Potential financial benefits of improved energy management

Decarbonization strategy	Practices	Value drivers	Financial benefits
Energy management	**Implement on-site renewable energy**	Sales and marketing	Additional revenue due to green premiums
		Operational efficiency	Tax incentives and rebates
		Risk management	Reduced dependence on grid and protection against rising energy prices
		Operational efficiency	Reduces utility costs
	Use power purchase agreements	Operational efficiency	Tax incentives and rebates
		Operational efficiency	Lower costs of energy procurement through negotiated contract rates.
		Sales and Marketing	Green premiums
	Improve energy efficiency	Operational efficiency	Lower operational costs (energy, maintenance, increased employee productivity)

Table 2: Potential benefits of circularity

Circularity Strategy	Practices	Value drivers	Financial benefits
Circularity	Implement consumer circular take-back program	Customer loyalty	Reduced new customer acquisition costs
		Media coverage	Earned media generated from increased visibility associated w/the program
		Sales and marketing	New revenue stream and incremental profit generated for parent company from customers
	Use reprocessed component parts and/or equipment	Operational efficiency	Reduced procurement costs due to reduced need to purchase new equipment/parts
		Operational efficiency	Reduced waste management costs associated with disposal
		Risk management	Reduced GHG emissions
		Risk management	Reduced risk of supply chain disruption
	Increase percentage of recycled materials in product	Operational efficiency	Lower operational costs (energy, maintenance, increased employee productivity)
		Sales and marketing	Avoided sales loss/sustained sales from more sustainability-minded consumers

2. An Assessment of SASB and CSR-D with regards to sustainability performance

We researched both SASB (now part of International Sustainability Standards Board (ISSB)) and the original ESRS under CSR-D to assess the strengths and weakness of both standards with a view toward understanding how likely they would be to drive financially and societally material benefits. We find that both exhibit weaknesses that may partially justify some of the criticisms.

2.1 SASB/ISSB

There are four challenges associated with SASB/ISSB as currently constructed:

1. Most reporting metrics are activities/outputs rather than outcomes.
2. Most metrics are neutral or risk-focused rather than opportunity-focused.
3. ESG reporting metrics neither integrate financial performance nor provide guidance on how to understand and drive better financial performance.
4. Our own engagements with companies and investors find that some focus on these reporting metrics at the beginning and end of their sustainability initiatives, which given the above weaknesses, are unlikely to drive sustainability benefits or financial improvement.

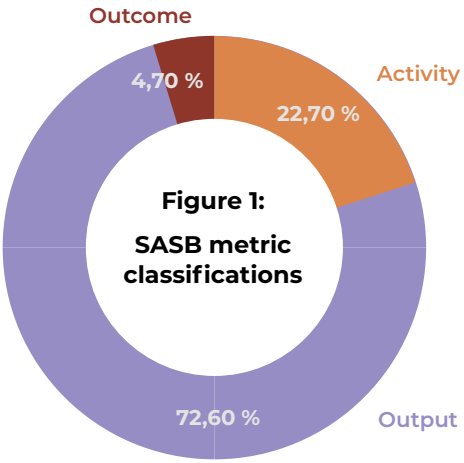


Figure 1:
SASB metric classifications

Our assessment of SASB metrics finds that 95% of SASB metrics are primarily process/output-based (including activity classification) versus performance/outcome-based (Figure 1) [Whelan and Marx (2025)]. Certain verticals are also missing material metrics on topics such as living wage and circularity. Social metrics seem to be lacking the most.

In addition, the majority of SASB's existing metrics are neutral or risk metrics. The "neutral" metrics provide a snapshot of performance but do not help us understand what the positive or negative financial impact might be. For example, tracking GhG emissions, even reducing them, does not tell us if we have reduced a given financial risk or created revenue growth. The "risk" metrics focus on mitigating environmental and social risks that could threaten the bottom line, but the standards overall do not adequately explore the upside, such as operational efficiency and innovation and growth.

- **Activity metrics**: these metrics encourage organizations to describe the environment of a given issue and what needs to be done, without actually implementing said activities. Often, they describe strategies or risks and opportunities. Activity metrics do not require organizations to monitor these strategies' results or develop recommendations.
- **Output metrics**: we classified output metrics as those that document an output such as a policy or training or a chemical/GhG metric, but do not drive or document specific improvements/changes. They generally have no benchmark, target, or context.
- **Outcome metrics**: such metrics require a methodology to assess whether progress has been made and whether they have been meaningful. Knowing that current water use is x, does not tell us if that is positive or negative because we have no context. Knowing that water use has been reduced by 10%, say, is still incomplete information because we do not know if that is meaningful or not. Setting context, targets and tracking progress over time, thus, becomes an important element to ensure outcomes.

Output metrics can be converted to outcome metrics by requirements for short and long-term targets, disclosure of the base year, and performance to those targets. Furthermore, disclosures should include a description of the target-setting process and the suitability of those targets to reach company goals and commitments. These metrics allow organizations to set targets to improve their efforts and document performance.

For example, the number of recalls is a material issue for the automotive sector. But the SASB metrics just ask for the total number of recalls, which gives us no sense of performance. We can adjust the metric as follows: (1) number of vehicles recalled in current year: (a) short and long-term targets with a defined period, (b) base year, and (c) performance to time-based targets.

In another example from the agricultural products sector, SASB asks for: "the percentage of agricultural products sourced that are certified to a third-party environmental and/or social standard, and percentages by standard." We can add the following: (1) percentage of agricultural products sourced by all suppliers that are certified to a third-party environmental and/or social standard, and (2) percentages by standard: (a) short- and long-term target with a defined period, (b) base year, and (c) performance to time-based targets.

In a final example from the insurance sector, SASB asks for: "a description of approach to incorporation of environmental, social, and governance (ESG) factors in investment management processes and strategies." We turned this activity into an outcome-oriented metric as follows: number and total value of investments incorporating the integration of (1) ESG factors: (a) short and long-term targets with a defined period, (b) base year, and (c) performance to time-based targets.

2.1.1 Lack of alignment with financial impacts

These revisions from outputs to outcomes still do not align the ESG reporting with financial returns.

The revised recall metric does not address the cost of these recalls, for example. In separate research, we have found that companies only track the direct costs of fixing the faulty component, but in fact, there are many other costs ranging from litigation, media outreach, crisis management time and effort, employee training, and new systems to avoid recurrence of the problem, etc., which could be tracked.

To follow-up with our agricultural certification example, the outcome-based metric does not by itself determine whether financial value will be created, but our research has found that sustainable sourcing with certification programs does lead to improved sales and marketing benefits, as well as reduced supply chain risk. We, therefore, would want to include metrics related to increased sales, premiums, or market share improvement.

With regards to the insurance company's integration of ESG into its investment portfolio - to understand if this shift had financially material results - we would need to add another metric asking for performance of the investment portfolio versus a conventional benchmark.

These examples underscore our concern that identifying something as financially material does not actually drive better tracking or assessment of the financial impacts.

There are two paths to potentially follow:

1. Change the metrics to integrate financial metrics with the ESG metrics (e.g., energy savings associated with energy management).
2. Provide guidance on which metrics can drive value (e.g., note that improved energy management can reduce energy pricing volatility, increase operational efficiency, and reduce potential or existing carbon fees).

2.2 CSR-D/ESRS

The ESRS/CSR-D provides more strategic insight and outcome-oriented guidance than SASB. For example, CSR-D requires companies to identify and disclose all material impacts within their value chain as per the following (impacts can be actual or potential, negative or positive over the short-, medium-, or long-term):

- Their governance processes, controls, and procedures used to monitor, manage, and oversee material impacts.
- How their strategy and business model interact with its material impacts.
- The processes by which they identify, assess, and manage material impacts.
- Their performance, including targets it has set and progress towards meeting them.

However, despite the guidance outlined in the Materiality Assessment Implementation Guidance (EFRAG IG 1), a standardized methodology for conducting a materiality assessment is not provided, nor does ESRS offer detailed advice on what constitutes material impacts, risks and opportunities (IROs). The absence of a clear, standardized methodology or list of IROs can create inconsistency in how companies approach these assessments. This can lead to companies identifying different or incomplete lists of material matters, making it difficult for stakeholders to compare data points against companies in the same sector.

While the regulation aims to enhance transparency and accountability for a company's material impacts and to encourage positive outcomes, it does not mandate that companies set outcome-oriented targets. Instead, it requires disclosure only for targets that companies have already established. For any targets set, the regulation imposes a set of minimum disclosure requirements, including the nature and scope of the target, baseline value, baseline year, milestones, interim targets, and more (ESRS 2). Consequently, despite robust disclosures for set targets, the standard may actually disincentivize the setting of targets as companies aim to avoid additional disclosure that could expose them to risks if there is not a requirement to do so.

The ESRS does require companies to report the anticipated financial effects of material sustainability risks and opportunities. However, the standards primarily focus on identifying and managing negative financial risks rather than highlighting potential financial upside opportunities and the metrics are few. The ESRS includes around 100[2] (out of ~1,000) data points related to financial disclosures (Table 3), with most found in E1. (The original version of the CSR-D required 1200 plus data disclosures; the newest version has reduced that by more than 60%.) However, less than five of these data points explicitly address the financial upside

2. In order to identify financial disclosures leveraged EFRAG IG 3 List of ESRS Data Points and filtered for "Monetary" data types, disclosures in the "Anticipated Financial Effects" sub reporting area, and disclosures related to monetary fines and penalties.

Table 3: ESRS/CSR-D number of financial metrics

ESRS Standard	Number of financial disclosures by standard
ESRS 2	17
E1	56
E2	11
E3	1
E4	7
E5	6
G1	5
S1	2
	105

(i.e., cost savings from climate mitigation and adaptation plans and revenues from low-carbon products). Otherwise, the disclosure requirements are risk focused (e.g., operating costs, capital expenses, or monetary losses) or remain vague, asking companies to disclose anticipated financial effects from material risks or opportunities but not explicitly indicating which financial measures to disclose. In the newest version of the standard, some financial metrics have been eliminated; E1 now has 5 financial metrics, for example.

For example, "Assets at material physical risk before considering climate change adaptation actions" (E1-9a, 66a) highlights exposure to climate-related vulnerabilities. Some disclosures are neutral or depend on interpretation, such as "Disclosure of quantitative information about anticipated financial effects of material risks and opportunities arising from pollution-related impacts" (E2-6, 39a). Few disclosures explicitly emphasize opportunities, such as "Expected cost savings from climate change adaptation actions" (E1-9, 69a), which recognizes the financial benefits of proactive sustainability measures.

2.2.1 Recommendations

For standard setters, specifically IISB and EFRS, we have published two reports that provide specific recommendations for improving the standards and metrics: how to improve IISB/SASB reporting to drive better societal and financial performance [Whelan and Marx (2025) and evaluating the effectiveness of E.U. CSRD in driving sustainability outcomes and improved financial performance [Cannon and Whelan (2025)]. However, we appreciate the variability in topics and approaches across industries and value chain players makes it difficult to provide the type of granularity and specificity that is required to drive real impact and financial benefit. Consequently, we turn our attention to how corporate leaders can best manage material ESG issues to improve performance.

3. Setting smart sustainability and ROSI KPIs and mapping to ESG reporting metrics

Most importantly, implementation of sustainability strategies and target setting should be strategic and embedded in business strategy, as is in every other area of business endeavor, rather than approached as a reporting and compliance exercise.

Myriad management studies and guidance point to mapping the strengths, weaknesses, opportunities, and threats (SWOT) to an organization and its ecosystem (e.g., customers, suppliers, and investors) and building that assessment into business strategies with innovative products and or services [Helms and Nixon (2010)]. SMART (specific, measurable, achievable, relevant, and time-bound) KPIs are required to drive progress and effect mid-course corrections [and can be enhanced by AI; Schrage et al. (2024)]. Sustainability should be no different.

In other words, companies can themselves set outcome-oriented KPIs that they can map to ESG reporting metrics. Hence, the reporting starts first with business strategy and ends with compliance, rather than the other way around.

Common practice today has companies undertaking an ESG materiality analysis with internal and external stakeholders. SASB provides a list of material topics by industry sector (though missing some topics) and CSR-D outlines a double-materiality assessment process. With a materiality matrix in hand, companies can use that analysis to undertake a SWOT and determine prioritiza-

tion and appropriate sustainability strategies and practices – and the right ESG metrics to track. For example, if a manufacturing company uses a lot of water and is located in an area under water stress, it could be exposed to cost increases, regulatory risk, and license to operate challenges. It needs to then identify potential solutions, assess the cost/benefits of different scenarios, and set up KPIs tracking water use per unit, costs per unit (including cost of water, cost of wastewater disposal, cost of energy needed to move, heat and cool the water). This sound obvious, but current ESG reporting metrics do not drive this kind of approach.

The logic flows as follows:

1. **Identify the sustainability KPI for the material ESG issue**: for example, reduce water use at all mills by 30 percent from 2023 baseline by 2026 (note this is an outcome metric, but it does not have a financial metric).
2. **Identify the potential financial benefits of that action and determine which benefits are likeliest to drive financial value** (a comprehensive analysis will look at the full range of benefits, ranging from productivity, to risk, to sales, to efficiency): for example, operational efficiency (reduced water and energy costs as well as reduced wastewater disposal costs), risk mitigation (continued access to water - no stranded asset), and stakeholder relations (improved relationships with NGOs, local community, and regulators).
3. **Develop financial KPI(s) and monetization methods to track those benefits**: for example, operational efficiency: energy, water, and wastewater disposal costs reduced by x% at each mill, totaling $x by 2026.

4. Conclusion

Critics and supporters of ESG reporting are both justified in their assessment of ESG reporting. While research demonstrates that sustainability is indeed financially material (based on topic and industry) and can create financial risks and opportunities, ESG reporting generally may not include metrics that would actually improve management, impact, and financial performance. Sustainability should be managed as would any other material topic for a company or an investor, with appropriate KPIs and monitoring that assess outcome-oriented societal and financial performance. ESG reporting metrics can be improved to focus more directly on outcomes and to incorporate more direct financial impacts, however, companies should define their own sustainability and ROSI KPIs as a first order of business.

References

Billio, M., M. Costola, I. Hristova, C. Latino, and L. Pelizzon, 2021, "Inside the ESG ratings: (dis)agreement and performance," Corporate Social Responsibility and Environmental Management 28:5, 1426-1445

Cannon, S., and T. Whelan, 2025, "Evaluating the effectiveness of the EU Corporate Sustainability Reporting

Directive (CSRD) in driving sustainability outcomes and improved financial performance," Center for Sustainable Business, Stern School of Business, New York University, April, https://tinyurl.com/566y7j68

Giglio, S., T. Kuchler, J. Stroebel, and O. Wang, 2025, "Nature loss and climate change: the twin-crises multiplier," NBER working paper no. 33361

Helms, M., and J. Nixon, 2010, "Exploring SWOT analysis – where are we now?: A review of academic research from the last decade," Journal of Strategy and Management 3:3, 215-251

Ibishova, B., B. Misund, and R. Tveterås, 2024, "Driving green: financial benefits of carbon emission reduction in companies," International Review of Financial Services 96:B, 103757

Malone, L., and E. B. Holland, S. Thacher, and Bartlett LLP, 2023, "Florida passes farthest-reaching anti-ESG law to date," Harvard Law School Forum on Corporate Governance blog, https://tinyurl.com/3k3p9aan

NYU CSB, 2024, "Sustainable Market Share Index™," Center for Sustainable Business, New York University, https://tinyurl.com/274m8ma8

Rouen, E., K. Sachdeva, and A. Yoon, 2022, "The evolution of ESG reports and the role of voluntary standards," Harvard Business School working paper no. 23-024

Schrage, M., D. Kiron, F. Candelon, S. Khodabandeh, and M. Chu, 2024, "The future of strategic measurement: enhancing KPIs with AI," Sloan Management Review, February 12, https://tinyurl.com/38y23vdf

Whelan, T., 2025, "ESG reporting needs a refresh. Here's how to fix it," Trellis, December 1, https://tinyurl.com/3sd2z38a

Whelan, T., and E. Douglas, 2021, "How to talk to your CFO about sustainability," Harvard Business Review January-February, https://tinyurl.com/mes58kfb

Whelan, T., and L. Marx, 2025, "Recommendations for ISSB/SASB, policy-makers, investors and corporate leaders: how to improve ISSB/SASB reporting to drive better societal and financial performance," Center for Sustainable Business, Stern School of Business, New York University, May, https://tinyurl.com/534enwr4

How data and AI shape financial services

Securing the future of **Payments**

Johan Gerber,
Executive Vice President and Head of Security Solutions, Mastercard

Kate Pohl,
Global Sales Lead, Projective Group

ABSTRACT

Challenging tomorrow's cybercriminals requires intelligent, adaptive, and collaborative tools. Revolutionary technologies such as artificial intelligence (AI), and its many derivatives - generative AI (GenAI) and Agentic AI, and quantum computing - will provide the foot soldiers for battles no longer fought with static rules and blacklists, but through dynamic, real-time analysis and intervention. Sadly, these same technologies are also aiding cybercriminals in their efforts to undertake fraud on a larger scale, and more globally. This is an arms race in which threat and response evolve in tandem and in tension. In this article, we explain how challenging tomorrow's cybercriminals requires not only the application of the latest and most advanced technologies but also collaboration among different parties involved in the Payments value chain, including among competitors. We further describe how Mastercard has been able to develop solutions that meet the challenges of managing, and mitigating, Payments fraud in today's highly complex technological environment while remaining compliant with the myriad of global regulations.

1. Introduction

When Phil Brandenberger used his Mastercard to purchase a Sting CD for $12.48 from NetMarket in August of 1994, in what was the first-ever encrypted purchase on the internet [Lewis (1994)], he could have never imagined how his trust in his Payments provider to protect him in case of failure-to-deliver by the merchant would revolutionize how commerce and finance would be conducted for decades to come. That single transaction set in motion a chain of events that made today's e-commerce viable and paved the way for how business is conducted, and paid for, online; resulting in a digital Payments marketplace that makes in excess of U.S.$20 trillion of e-commerce annually possible [Statista (2025)].

That single transaction also resulted in one of the most transformational eras in the history of modern finance. Banks and other established Payments providers were forced to familiarize themselves with the new technologies, and their pitfalls, so as to make these types of online transactions viable, fast, and safe. And the established players were not alone in this endeavor. The e-finance revolution resulted in the establishment of a completely new type of competitor, young technologists who were not only natives of the online world, but who also knew the benefits that came from crisp and elegant websites and apps. More importantly, these newcomers, who would later come to be known as Fintechs, were free from the legacy systems that the more established players were hampered with.

The first task was, of course, to make these transactions possible for clients, many of whom had become increasingly demanding as they were becoming accustomed to services that were instantaneous, frictionless, and free from errors. The new online world had also made clients significantly less loyal to brands than their predecessors [Schultz and Block (2012)]; they were quite happy to move from one provider to another to get a better price and/or service.

While many Payments providers, particularly banks, initially faced challenges adapting to the digital landscape [Porfírio et al. (2024)], they ultimately succeeded in embracing the necessary changes. The emergence of Fintechs played a significant role in accelerating this transformation [Breidbach et al. (2020), Diener and Spacek (2021)]. As a result, robust technological capabilities were established, enabling clients to make online Payments through various channels.

The problem, of course, was that since these transactions were now taking place outside of the banks' established networks, and predominantly on smartphones, computers and over the internet, securing transactions end-to-end became an even bigger challenge. Unscrupulous actors, later to be called cyber criminals, quickly realized the opportunities that such an environment provided and started looking for ways to interrupt these transactions and divert funds to their own accounts [Clark (2025)].

And the faster these transactions became, the more susceptible the system became to cyber-attacks, with the number of attacks having doubled since the covid pandemic [IMF (2024)].

In response, an entire industry in cybersecurity was developed to prevent cyber criminals from exploiting the vulnerabilities that inevitably exist within the financial and technological value chain. The issue, of course, was that while banks and other Payments providers were investing billions of dollars safeguarding these transactions, cyber criminals were also investing and developing advanced capabilities to stay one step ahead.

And the sums invested are not negligible. According to Global Market Insights (2024), the payment security market was valued at U.S.$23.4 billion in 2023 and is projected to reach U.S.$78.5 billion by 2032. Sadly, the sums that were stolen from the system were not small either, with U.S.$27.6 billion of losses reported by the FBI for the U.S. alone between 2018 and 2022 [Judd (2024)].

Just how successful Payment providers have been in preventing large scale cyber events is debatable [Tuominen (2024)], but as these transactions become faster and the markets grow, the risks become even larger. And the concern is not just that faster payments, as they are called, will result in fraud and data breaches for individual participants, it is that a successful large-scale cyber attack could quickly engulf the entire financial system [World Bank (2025)]. Consequently, banks and other Payment providers are also being forced to act as the last line of defense against massive systemic risks to the financial system.

There are a number of ways in which Payments providers are protecting themselves against cyber-attacks, a number of which will be discussed in the next section, but most would agree that none is foolproof. In this article, we suggest that as tokenization and passkeys become the backbone of

secure digital payments, combined with fact that the next decade will redefine how we protect identity, data, and trust at scale; cybersecurity must evolve from reactive threat mitigation to autonomous, proactive defense systems to improve how we protect clients as well as the financial system.

2. Contemporary cybersecurity solution

It goes without saying that no organization wishes to become a victim of a cyberattack, none more so than payment providers, given the vast sums involved. The ramifications could be quite severe, and they typically tend to be much larger than the sum of initial money lost.

From a business perspective, cyberattacks can severely impact an organization's reputation and brand image, which in turn could undermine the trust that customers have in that organization. The implications of the loss of business could be significantly greater than the initial amounts lost to the cybersecurity incident. And it could be long lasting, with some studies finding that over 80% of customers will move to a competitor if they do not trust the company to responsibly look after their data [Lieberman (2017)].

There are also legal ramifications that businesses need to be aware of. Data breaches and other cybersecurity incidents can have serious legal consequences for the organizations involved, with the E.U.'s Digital Operational Resilience Act (DORA) [E.U. (2022)] and the U.K.'s Financial Conduct Authority's (FCA) Supervisory Statement (SS) 6/24 on Critical Third Parties [FCA (2024)], and the broader Operational Resilience regulations in the U.K. [BoE (2025)], being cases in point. [1]

It is for this reason that Payments providers, including banks, spend millions of dollars a year on cybersecurity. Protecting themselves from the ramifications of cyberattacks is one, if not the most important issue for senior financial executives [EY/IIF (2025)]. Although some might view spending hundreds of million dollars a year just on cybersecurity excessive, when one considers that major banking institutions fight off billions of hacking attempts a day, that figure suddenly seems more reasonable [Naysmith (2024)]. In fact, when one considers that there are estimates that cybercrime is expected to increase from U.S.$9.22 trillion in 2024 to over U.S.$15 trillion by 2029, then the sums invested by major global financial services organizations might even seem inadequate.

It is undebatable that financial services organizations need to spend what is necessary to protect themselves and their clients from cyberattacks. What is debatable is whether these investments have been as effective as they were intended to be.

And there is no shortage of tools to manage cybersecurity. Payment providers have access to a large array of such tools, including:

- **Encryption**: transforms sensitive/private data into unreadable code during transmission. The encryption could be symmetric, where the same key is used to lock and unlock the data, and asymmetric, which uses a public key to lock and a private key to unlock the data.
- **Payment gateways**: act as a secure intermediary between customers, merchants, and payment providers. They transmit payments details to banks to verify them.
- **Factor authentication (MFA)**: using single-factor authentication (SFA), which uses one form of identification, such as a password or a PIN, two-factor authentication (2FA), which uses a combination of a password and a one-time code sent to a client's mobile or app, or a multifactor authentication (MFA), which uses a combination of biometric data, security questions, or physical tokens, users are verified.
- **Digital wallets**: securely store payment information in a digital format to enable clients to undertake transactions without the need for physical cards.
- **EMV (Europay, Mastercard, and Visa) chip cards:** use small microprocessor chips attached to credit and debit cards that generate unique transaction codes for each transaction.
- **Tokenization**: replaces sensitive payment data, such as credit or debit card numbers, with unique tokens generated by a secure system. Given that the tokens simply reference the original payment information, they cannot by themselves be used to obtain the original payment data.

1. It should be noted that while DORA is prescriptive, with a focus on technology, U.K.'s Operational Resilience regulation is not, and goes beyond IT.

The challenge is identifying if the solutions that are most effective against fraud can keep up with the latest technological developments and remain one step ahead of cybercriminals.

This is made more challenging by efforts to remain compliant with the myriad of global payment and money transfer regulations, not to mention data protection.

Payments providers need to navigate their way through the complex web of national and jurisdictional requirements to be anti-money laundering (AML) and know-your-customer (KYC) compliant. Should the transaction involve card payments, they also need to ensure compliance with Payment Card Industry Data Security Standards (PCI DSS). Numerous international bodies, such as the Basel Committee on Banking Supervision (BCBS), and the Committee on Payments and Market Infrastructures, have published cybersecurity guidelines for payment providers. The E.U.'s Revised Payment Services Directive (PSD2) has also mandated a number of requirements, including the need to ensure compliance with the strong customer authentication (SCA) for electronic and credit card payments.

In addition, payment providers need to ensure they are compliant with national and international data protection requirements, such as the E.U.'s General Data Protection Regulation (GDPR), which mandates how payment providers can collect, process, and store payment data.

In the following sections, we will describe how Mastercard has been able to develop solutions that meet the challenges of managing, and mitigating, payments fraud in today's highly complex technological environment while remaining compliant with global regulations.

3. Preparing for the future - proactive security

Effective cybersecurity will require both the adoption of advanced technologies and strong collaboration across all participants in the payments ecosystem, including secure information sharing, even among competitors.

Technologies such as AI are transforming cybersecurity from static rule-based approaches to dynamic, real-time detection and response, enabling faster and more adaptive threat mitigation.

When combined with large-scale intelligence sharing, functioning as a global early-warning system, future cybersecurity solutions will detect emerging threats early and identify fraudulent activity in real time across multiple points in the value chain.

The focus of businesses, institutions, and governments will be on the effective management of increasingly autonomous technologies, as well as forming alliances to exchange insights, best practices, and defensive techniques to strengthen collective resilience against cyberattacks. Trust networks using privacy-enhancing technologies and federated learning systems will ensure that intelligence sharing is configured for specific activities, such as fraud or identity theft, without sharing underlying data.

3.1 Threat intelligence: early warning systems

Intelligence and insights are increasingly recognized as the sword and shield of tomorrow's cybersecurity—intelligence to analyze vast datasets in real time, and insights to drive effective, strategic responses. Organizations require real-time data feeds that provide instant visibility into emerging threats, attack methodologies, and vulnerabilities.

Threat intelligence platforms play a critical role by correlating vast amounts of data from diverse sources to strengthen collective defense strategies. Predictive analytics further enable organizations to understand attacker motivations, preferred tools, and behavioral patterns, supporting a proactive security posture.

Threat intelligence encompasses a range of capabilities, including malware analysis, vulnerability assessment, and monitoring of malicious actors across both physical and digital domains. As the threat landscape grows more sophisticated, the demand for real-time, automated intelligence has surged.

Companies are increasingly leveraging artificial intelligence (AI) and machine learning (ML) to predict and mitigate risks before they materialize.

Industry collaboration, such as sharing threat data across platforms and regions, remains essential to enhancing global cybersecurity resilience. Mastercard Threat Intelligence- powered by Recorded

Future - exemplifies this approach by delivering actionable intelligence that combines advanced analytics with global data-sharing initiatives. By integrating predictive modeling and real-time monitoring, such platforms enable organizations not only to defend against attacks but to anticipate and adapt to evolving threats.

3.2 AI at scale: autonomous defense

Cybersecurity is evolving from reactive threat mitigation to autonomous, proactive defense systems. These systems leverage AI to predict, detect, and neutralize threats before they cause damage. Automated security tools continuously analyze network traffic, endpoint activities, and system behavior to identify anomalies indicative of cyberattacks. The integration of self-healing systems ensures that vulnerabilities are automatically patched, minimizing downtime and reducing human intervention.

Dynamic and intelligent fraud detection systems are a leap forward in the payments industry, a principal target of fraudsters and cyberattacks. Here is how they make a difference: [2]

- **Real-time analysis**: AI-powered systems can analyze vast amounts of transactional data in real time, identifying anomalies and suspicious patterns as they occur.
- **Adaptive learning**: unlike traditional rule-based systems, they continually learn from new data, attack vectors, and criminal strategies.
- **Predictive analytics**: leveraging historical data makes possible anticipating potential fraud scenarios and implementing preventative measures.
- **Scalability and efficiency**: the growing volume and complexity of digital transactions are a perfect fit for AI analysis, reducing the false positives produced when traditional systems erroneously flag legitimate transactions as fraudulent.

3.3 Federated learning: collaborative fraud modelling

Federated learning is a way of training AI cybersecurity models that turns its usual logic on its head. In traditional methods, data flows away from individual nodes, or "clients", towards the model in the cloud - potentially exposing that data and violating data-governance regulations. In federated learning, the model itself is downloaded to the clients and trained there, locally, using local data. At the end of the process the various local versions are reassembled in the cloud into a single integral model, without risking exposure or sharing of the underlying data.

The key benefits are:

- **No data movement or centralization**: data remains within each organization's secure environment, which reduces the risk of data breaches, compliance violations, or leaks of proprietary information.
- **Improved fraud detection accuracy**: by training on insights from multiple institutions, the model gains a more comprehensive understanding of fraud behaviors, making it better at identifying suspicious activity.
- **Privacy protection**: federated learning uses privacy-enhancing technology (PET). Since raw data is never shared, organizations maintain control over their sensitive information, aligning with privacy regulations and other industry-specific standards.
- **Preserving competitive intelligence**: organizations are often reluctant to share data because it can reveal business strategies or customer behaviors. Federated learning enables collaboration without compromising competitive advantages. [3]

3.4 First party fraud

First-party fraud is committed by people directly engaging in transactions. A consumer might falsely claim never to have authorized a transaction, for example. In a type of first-party fraud known alternatively as friendly or chargeback fraud, the buyer of an expensive smartphone online might falsely claim never to have received it. A quarter of consumers admit to having committed friendly fraud in 2024 [Sift (2024)]. Businesses both big and small need support with burdensome time and resource-intensive issues, such as researching and addressing these claims. Mastercard's First Party Trust program provides enhanced data-sharing,

2. For example, Mastercard's Safety Net global fraud monitoring tool prevented U.S.$50 billion in fraud between 2022 and 2024. It detects organized card testing by fraudsters, indicative of a BIN (bank identification number) attack. Another AI tool, Decision Intelligence, assesses fraud risk across Mastercard's 159 billion processed transactions each year, issuing a score for each one within 50 milliseconds.
3. Improvement in fraud detection accuracy by financial institutions as a result of a federated learning platform is expected to be around 20%. Federated learning market is expected to grow from U.S.$128.3 million in 2023 to U.S.$260.5 million in 2030 [Ohaba (2024)].

either at the time of transaction or at the time a dispute is raised to help issuers better identify third-party fraud.

3.5 Know your agent: onboarding AI

The rush to deploy AI to organizational systems creates a whole new set of often unpredictable risks. For example, despite the immense power of AI tools, around two-thirds of organizations have no processes in place to vet their security risks before deploying them [WEF (2025)]. Effective AI governance will be crucial to prevent privacy violations, data breaches, and nefarious manipulations. A focus on explainable AI (XAI) and the advent of "know your agent" (KYA) processes will enhance transparency, ensuring that cybersecurity decisions made by AI-driven systems are interpretable and accountable. Additionally, continuous audits and adversarial testing will be necessary to prevent cybercriminals from exploiting AI-driven security measures. Among the steps inherent in the KYA process are verification of an AI solution's or other digital agent's data sources and algorithms, bias testing, and the drafting of permission and access management protocols.

Organizations could also consider:

- **Conducting regular adversarial testing**: simulate attacks to identify weaknesses in AI systems. This includes testing for adversarial inputs, data poisoning, and model vulnerabilities.
- **Implementing robust data security measures**: encrypt sensitive datasets used for training AI models and restrict access to authorized personnel. Regularly audit data sources to ensure they are free from malicious manipulation.
- **Adopting explainable AI (XAI) solutions**: invest in AI models that are interpretable and transparent, allowing organizations to understand the reasoning behind decisions. This builds trust and more easily detects anomalies.
- **Human-in-the-loop mechanisms**: incorporate human oversight into critical AI processes, especially for high-stakes decision-making. Human operators can review and verify AI outputs to minimize errors or exploitation.
- **Monitoring and updating AI models**: continuously monitor AI systems for unusual behavior and deploy updates to improve resilience against emerging threats. This includes patching vulnerabilities and improving algorithm robustness.
- **Compliance and governance**: develop comprehensive policies to ensure AI systems align with data protection regulations (e.g., GDPR, CCPA). Clear governance frameworks can guide ethical and secure AI implementation.

Governments worldwide are introducing new regulations around AI. The European Union AI Act categorizes AI applications according to risk and has strict requirements on AI systems dealing with biometric identification, security, and critical infrastructure, for example. This introduces compliance challenges for multinational organizations with complex global operations.

4. Fighting the scammers

The commoditization of personal data, fueled by dark web marketplaces and AI-driven fraud tactics, is being met by innovation in identity verification and authentication. The next five years will see a decisive shift toward continuous, real-time, risk-based authentication, building on behavioral analytics, biometrics, and identity networks.

4.1 Continuous authentication: always-on security

As tokens and passkeys substitute for data and passwords, continuous authentication will evolve to leverage a wider range of biometrics, behavioral analytics, device intelligence, and contextual solutions that ensure persistent and dynamic verification of individuals and entities. This approach significantly reduces identity fraud.

Continuous, or always-on, authentication offers dynamic security in a protean threat environment, assessing on a rolling basis whether devices and users should be trusted. It uses context-aware methods like geolocation assessment and transaction data and picks up on anomalies like atypical typing rhythms or mouse movements. AI analysis of multiple datapoints enables a smarter, more reasoned assessment of risk than passwords and PINs. Global markets for password-less authentication, risk-based authentication, and passive (silent) authentication are forecasted to grow rapidly over the next several years, driven by tech advances,

demands for enhanced security, and rising adoption across industries like finance, health care, retail, and workplace security.

Continuous authentication is related to zero-trust architecture, an approach to cybersecurity that eschews single logins and simple perimeter protection in favor of more holistic security. Every access request, whether inside or outside the network, is treated as untrusted until proven otherwise. Zero-trust architecture incorporates multiple strategies, such as least privilege access (limiting access to only most needed), micro-segmentation, and real-time monitoring.

4.2 Tokenization and passkeys: the twin turbo of fraud prevention

Tokenization and passkeys are both critical tools in modern security and fraud prevention, though they serve distinct purposes.

Tokenization protects sensitive data, such as credit card numbers or personal information, by replacing it with a unique token bound by a cryptogram, making it meaningless outside of the secure system, minimizing exposure and reducing the risk of data breaches.[4]

Passkeys, on the other hand, are a password-less authentication method that uses cryptographic key pairs - one private and one public - ensuring secure account access while eliminating vulnerabilities associated with traditional passwords, like reuse or phishing.

While tokenization focuses on securing data in storage and transactions, passkeys emphasize secure authentication.

According to the FIDO Alliance, 20% of the world's top 100 services and websites have already adopted passkeys. By 2030, Mastercard will eliminate manual card entry and static passwords by integrating tokenization and biometric authentication, ensuring every online transaction is both secure and seamless across its network.

4.3 AI scam protection: patterns and anomalies

Tomorrow's scam protection tools will also use context-aware methods to assess risk. The defining feature of "authorized" scams is that the victims themselves play an active role in their misfortune, making it far harder to police using traditional authentication tools. Authorized Push Payment fraud (APP) happens when the victim is deceived into transferring funds directly to accounts under the control of the fraudster.

Fraudulent push payment transactions are expected to rise from 190.1 million in 2026 to 334.3 million globally by 2029 [Juniper (2025)].

4.4 Consumer fraud risk

New technologies are increasingly able to determine whether manipulation and deception are clouding victims' judgment. Mastercard's network spans banks and fintechs, enabling analysis of over 95% of real-time payments in the UK. At present, 15 banks employ its AI-powered Consumer Fraud Risk tool to safeguard their customers. This solution assesses numerous transaction data points and generates a risk score in real time for the sender's bank. Further AI enhancements deliver this score to the receiving banks, allowing them to identify potential mule accounts associated with fraudulent activity. The introduction of Mastercard A2A Protect—an advanced global solution—will strengthen consumer security, bolster banks' fraud prevention efforts, and implement a comprehensive reporting system for enhanced insight into fraud trends. The U.K. is bucking the global trend for APP fraud, seeing a 12% reduction across 2023 [PSR (2024)].

4.5 Real-time deepfake detection: a cat and mouse game

Advances in deepfake detection are focused on developing more robust and efficient methods to identify AI-generated content across various modalities, including video, audio, images, and documentation. Techniques include sophisticated machine learning, computer vision, and audio analysis techniques to differentiate between real and manipulated media. AI-powered deepfake detection tools can analyze pixel-level differences between authentic and manipulated images with high precision. However, very few solutions can detect deepfakes in real time, which limits their ability to detect a scam in progress. In time, we can expect real-time capability to become a standard feature of all conferencing apps, and not just identity verification platforms.

4. Mastercard aims to have 100% of e-commerce transactions tokenized by 2030.

U.S. deepfake detection platform Reality Defender recently announced its platform was capable of real-time detection, enabling the verification of individuals on web conferencing platforms [Reality Defender (2024)]. Meanwhile, Asia-Pacific's largest cybersecurity firm, Ensign InfoSecurity, says its Aletheia solution can identify AI-driven manipulations in audio and video content inside a second and with 90% accuracy [Knowles (2024)].

5. Innovation in anti-money laundering

As illicit financial networks become faster, smarter, and more decentralized, financial institutions and regulators are adopting next-gen anti-money laundering (AML) solutions powered by AI, blockchain intelligence, advanced analytics, and cross-industry collaboration. These solutions aim to detect suspicious activity at machine speed, streamline compliance, and future-proof financial ecosystems against evolving criminal tactics.

5.1 Advanced transaction monitoring: spotting patterns, mapping relationships

AI-driven platforms detect anomalies and suspicious activity that might indicate money laundering across vast payment networks in real time. These systems leverage ML algorithms to analyze millions of financial transactions per second, benchmark them against expected behaviors, and instantly flag deviations - such as rapid crypto conversion, cross-border layering activity, and signs of large transactions being divided into small, less conspicuous amounts (structuring). Such real-time detection enables compliance teams or automated systems to respond immediately, halt transactions, request further verification, or launch deeper investigations. This can disrupt money laundering schemes as they unfold, rather than relying solely upon post-transaction audits or static rule checks.

Advanced fraud detection platforms leverage ML and behavioral analytics to monitor customer activity across devices and communication channels. A mule account might pass KYC checks at onboarding but later demonstrate unusual login locations, rapid crypto transactions, or interaction patterns indicative of bot control, all of which can trigger AML alerts under behavioral profiling.

Burnett (2025) finds that AI-enhanced behavioral analytics modules have been integrated into approximately 40% of new AML software solutions.

5.2 Trace financial crime

Mastercard's Trace Financial Crime, another solution designed to make A2A payments more seamless and secure, uses ML and advanced graph analytics to detect complex, hidden financial activities across multiple money mule accounts [5], often revealing indirect relationships across disparate financial entities to identify criminal networks. This goes beyond traditional rule-based detection methods to identify sophisticated laundering techniques. ML algorithms enable Trace to analyze vast volumes of transactional data and learn to identify subtle anomalies that may suggest involvement in a criminal scheme, such as irregular fund flows, unusual transaction frequencies, or atypical recipient behavior.

In addition, AI tracing tools leverage graph analytics, which map out the complex web of connections between accounts, individuals, and financial institutions. This allows the tool to uncover indirect relationships. Imagine, for example, an account that was not previously involved in fraud but shares links with known mule accounts through common intermediaries or shared transaction paths. These insights help identify hidden networks of criminal activity that might otherwise remain undetected when analyzing accounts in isolation.

5.3 Identity as the foundation of trust

According to Mastercard 80% of global consumers were the targets of a scam attempt in the last year. As the digital ecosystem grows, it is increasingly imperative for businesses and consumers to establish trust.

Identity is the foundation of trust in digital interactions, ensuring that every connection, between consumers, devices, businesses, and AI agents, is secure and reliable. Enhanced identity verification solutions that streamline and simplify the process of proving one's identity, as well as confirming the identity of others, will be essential to the growth of the digital ecosystem. Mastercard is powering

5. According to the European Banking Authority [EBA (2023)] "2822 Banks and financial institutions join forces with law enforcement agencies in global effort against money laundering."

seamless, secure, and intelligent digital experiences by enabling robust authentication and verification across its network.

The adoption of digital identity wallets is expected to facilitate access to financial, governmental, and other services, including age verification, while accelerating the ability to create verified aliases for crypto transactions, which can help reduce the risk of fraud associated with complex addresses. Extending digital ID services to developing markets may also accelerate inclusion into the digital economy. In summary, the goal is to enable digital identity solutions that are as intuitive and dependable as making a payment.

5.4 RegTech innovation: Automated compliance

Modern regulatory technology (RegTech) tools leverage advanced AI to automate AML workflows. These tools are designed to handle vast amounts of financial data, detect anomalies, and flag suspicious activities with a high degree of accuracy. By automating repetitive and time-consuming tasks such as transaction monitoring, customer due diligence, and suspicious activity reporting, RegTech platforms reduce operational costs and human error while improving the speed and consistency of compliance processes. One of the key advantages of these tools is their ability to adapt to rapidly changing regulatory environments. Using ML, natural language processing (NLP), and real-time data analysis, RegTech solutions can update compliance rules dynamically, ensuring institutions remain compliant with evolving local and international regulations without constant manual intervention.

Adoption use-cases include:

- Automated KYC/AML checks.
- Continuous transaction monitoring.
- Real-time suspicious activity reporting.
- Risk scoring and customer segmentation.

6. Conclusion: towards a safer future

In a few short decades, the online world has become the foundation of global communication, financial flows, and information exchange, and the catalyst for cultural, political, and economic transformation. It is where innovation flourishes, ideas are shared, and communities come together. What happens online increasingly determines what happens offline.

Now, we are at a critical juncture. Cybercrime is evolving into a highly autonomous, scalable, and sophisticated industry that destroys livelihoods, erodes confidence, creates instability, and undermines growth, all while extracting huge amounts of value from the global economy.

The features and risks of this new landscape are in a constant state of flux. There is a proliferation of autonomous bots, polymorphic malware, deepfake scams, geopolitical antagonists - and a burgeoning cybercrime industry with evolving tactics and increasingly sophisticated tools.

Yet the technology that bad actors deploy to threaten, steal, and subvert is being met by unprecedented levels of innovation in cybersecurity and fraud prevention. This is an arms race in which threat and response evolve in tandem and in tension. Even as AI creates new opportunities for cyber criminals, it offers more effective ways of thwarting them. Looking further ahead, quantum computing will present a challenge to the cryptography that underpins the global digital infrastructure, while also promising to be a powerful tool for securing it.

Agentic AI, real-time data, behavioral intelligence, and advanced identity technologies will collectively form a foundation for a future cybersecurity that is smarter, faster, and more resilient. It will counter autonomous threats with more vigilant autonomous defenses, identity fraud with smarter verification, and money laundering with more discerning intelligence. Cybersecurity stakeholders also have an advantage that criminals lack: the will to work together. Global collaborations, trust networks, and insight-sharing are key to this.

The Payments sector has always been a target for criminals, and consequently a trailblazer in the development of security technologies. Mastercard has long pioneered the best and most secure ways to transact and increasingly interact and will continue to drive this innovation, to ensure the resilience of a financial ecosystem being rapidly redefined through technological change.

References

BoE, 2025, "Operational resilience of the financial sector," Bank of England, https://tinyurl.com/36edmmwy

Borges, E., 2024, "5 threat intelligence solution use cases," Recorder Future, June 25, https://tinyurl.com/2pd9y52u

Breidbach, C., B. Keating, and C. Lim, 2020, "Fintech: research directions to explore the digital transformation of financial service systems," Journal of Service Theory and Practice 30, 79-102

Burnett, S., 2025, "Anti-money laundering software statistics 2025: innovations and market growth," CoinLaw, June 16, https://tinyurl.com/3hyxpjrf

Clark, A., 2025, "Cybersecurity in the UK," research briefing, House of Commons Library, May 1, https://tinyurl.com/yc49u3vw

Diener, F., and M. Spacek, 2021, "Digital transformation in banking: a managerial perspective on barriers to change," Sustainability 13, 20-32

Ohaba, E., 2024, "Federated Learning for cybersecurity: collaborative intelligence for threat detection," March 18, https://tinyurl.com/2k4pyhaa

EBA, 2023, "Paper trail ends in jail time for 1013 money mules," European Banking Authority, https://tinyurl.com/387a47va

E.U., 2022, "Regulation (EU) 2022/2554 of the European Parliament and of the Council of 14 December 2022 on digital operational resilience for the financial sector and amending Regulations (EC) No 1060/2009, (EU) No 648/2012, (EU) No 600/2014, (EU) No 909/2014 and (EU) 2016/1011," European Union, https://tinyurl.com/39d8pef4

EY/IIF, 2025, "Bank risk management survey," EY and Institute of International Finance, https://tinyurl.com/533mzuts

FCA, 2024, "PS24/16: Operational resilience: Critical third parties to the UK financial sector," Financial Conduct Authority, https://tinyurl.com/3e632w5m

FIDO, 2024, "New survey: half of people use passkeys as frustrations with passwords continue," FIDO Alliance, May 2, https://tinyurl.com/vv756z9a

Fleck, A., 2024, "Cybercrime expected to skyrocket in coming years," Statista, Feb 22, https://tinyurl.com/3j2umpuv

Global Market Insights, 2024, "Payment security market size, industry analysis report 2032," July, https://tinyurl.com/4zxwwc5s

IMF, 2024, "The last mile: financial vulnerabilities and risks," International Monetary Fund, Global Financial Stability Report, April, https://tinyurl.com/5n7tj97h

Judd, E., 2024, "Seven cybersecurity threats for banks in 2024—and some smart precautions," ABA Banking Journal, February 8, https://tinyurl.com/yurez28w

Juniper, 2025, "Fraud detection & prevention in banking market: 2025-2030," Juniper Research, https://tinyurl.com/txdywjxt

Knowles, C., 2024, "Ensign InfoSecurity launches real-time deepfake detection tool," SecurityBrief Asia, September 20, https://tinyurl.com/msmyt53w

Lewis, P. H., 1994, "Attention shoppers: internet is open," New York Times, August 12, https://tinyurl.com/fwjnhhp3

Lieberman, M., 2017, "Mind the trust gap: how companies can retain customers after a security breach," Forbes, December 8, https://tinyurl.com/3c4abhj3

Naysmith, C., 2024, "JPMorgan faces 45 billion hacking attempts per day as Jamie Dimon calls cyber attacks 'biggest threat to the US financial system'," Yahoo Finance, January 24, https://tinyurl.com/5e5fmn92

Porfírio, J. A., J. A. Felício, and T. Carrilho, 2024, "Factors affecting digital transformation in banking," Journal of Business Research 171

PSR, 2024, "Authorised push payment (APP) scams performance report," Payment Systems Regulator, July, https://tinyurl.com/8s8z3d77

Reality Defender, 2024, "Reality Defender announces real-time video deepfake detection for web conferencing platforms," PR Newswire, October 15, https://tinyurl.com/2jmz8zyz

Schultz, D., and M. Block, 2012, "Rethinking brand loyalty in an age of interactivity," The IUP Journal of Brand Management 9:3, 21-39

Sift, 2024, "Q4 2024 Digital Trust Index: decoding dispute trends," https://tinyurl.com/48dsb7xe

Statista, 2025, "Payments – worldwide," https://tinyurl.com/ytbzfbu3

Tuominen, A., 2024, "Enhancing banks' resilience against cyber threats – a key priority for the ECB," The supervision blog, European Central Bank, July 26, https://tinyurl.com/5e4zxpmv

World Bank, 2025, "Cyber risks in fast payment systems," World Bank Focus Note, February, https://tinyurl.com/2s3p5rax

WEF, 2025, "Global cybersecurity outlook 2025 – insight report," World Economic Forum, January, https://tinyurl.com/vavuyvey

How data and AI shape financial services

Big tech in finance: how to ensure a level playing field in a data-driven economy

Judith Arnal,
Senior Research Fellow, Centre for European Policy St udies (CEPS), Elcano Royal Institute, and the European Credit Research Institute (ECRI)

ABSTRACT

Big Tech firms have expanded rapidly into financial services, integrating payments, wallets, and data-driven offerings into broader digital ecosystems. This article examines the economic logic behind that expansion and the resulting policy challenges for ensuring a level playing field. It first explains why Big Tech emerged in highly concentrated digital markets, driven by network effects, economies of scale and scope, and data feedback loops. It then analyses why financial services are a natural extension of platform models and maps Big Tech participation across the financial value chain in the U.S., Europe, and China. The paper highlights a key trade-off: Big Tech can enhance competition, innovation, and inclusion in the short term, yet may reinforce concentration and distort incentives over time. It proposes policy design principles and applies them to the E.U.'s FiDA debate on DMA-designated gatekeepers.

1. Introduction

Over the past two decades, a small number of large technology firms have come to dominate core segments of the digital economy. These firms, commonly referred to as Big Tech, are characterized by their massive market capitalizations, control over essential digital platforms, extensive user bases, and unparalleled access to data. While the term is most often associated with U.S.-based companies such as Google (Alphabet), Amazon, Apple, Meta, and Microsoft (the so-called GAFAM), it also encompasses major Chinese technology groups, notably Baidu, Alibaba, and Tencent (BAT), which occupy similarly dominant positions in their domestic markets and have expanded regionally and internationally.

Big Tech dominance is not the outcome of firm-specific strategies or regulatory arbitrage alone. Rather, it reflects the interaction of structural economic forces intrinsic to digital markets, including strong network effects, extreme economies of scale, data-driven feedback loops, and pronounced economies of scope. Once these forces take hold, digital markets tend naturally towards highly concentrated "winner-takes-most" outcomes, in which a small number of platforms capture a disproportionate share of value and user attention.

In recent years, Big Tech firms have increasingly expanded beyond their original digital activities into financial services. Payments, digital wallets, consumer and merchant credit, wealth-related services, and insurance distribution have become integrated components of broader platform ecosystems. This expansion raises fundamental questions for competition, financial stability and regulatory design. On the one hand, Big Tech participation in finance may enhance competition, foster innovation, and improve financial inclusion. On the other hand, it may reinforce concentration, distort incentives along the financial intermediation chain, and challenge the notion of a level playing field between platform-based firms and traditional financial institutions.

This article examines Big Tech participation in financial services from an economic and policy design perspective. It argues that the key challenge is not whether Big Tech should be allowed to operate in finance, but how regulatory frameworks should be adapted to reflect the ecosystem-based nature of platform business models while preserving competition, innovation, and financial stability.

2. Why Big Tech emerged: the economic forces behind digital concentration

The rise of Big Tech firms is not the result of idiosyncratic business strategies or regulatory arbitrage alone, but rather the outcome of structural economic forces inherent to digital markets. Large technology platforms such as Google, Amazon, Apple, Meta, and Microsoft in the U.S., or Baidu, Alibaba, and Tencent in China have emerged and consolidated their dominant positions in market environments characterized by strong network effects, extreme economies of scale, and pronounced economies of scope. Once these forces interact, they naturally give rise to highly concentrated market structures of the "winner-takes-most" type.

Network effects constitute a central driver of concentration in digital markets [Katz and Shapiro (1985)]. **Direct network effects** arise when the value of a service increases with the number of users adopting the same service. This mechanism is particularly evident in communication services and social networks, where platforms become more valuable as more users join. Once a critical mass is reached, user growth becomes self-reinforcing, making it increasingly difficult for competing platforms with smaller user bases to attract participants, even if they offer technically superior features.

Indirect network effects further amplify these dynamics. Many digital platforms operate as multi-sided markets, connecting distinct but interdependent groups of users. Search engines and social networks link users and advertisers, marketplaces connect buyers and sellers, and cloud platforms bring together firms and developers. In such settings, growth on one side of the platform increases the platform's value on the other side, generating positive feedback loops [Rochet and Tirole (2003)]. Platforms often subsidize one side of the market, typically end-users, while monetizing the other, which allows dominant firms to offer services at zero or very low prices, reinforcing barriers to entry for smaller competitors that lack comparable scale.

Economies of scale represent a second fundamental force shaping digital market outcomes

[Chandler (1990)]. Digital businesses typically exhibit cost structures characterized by very high fixed costs related to infrastructure, research and development, data centers, and algorithm design and near-zero marginal costs for serving additional users. Once the core infrastructure is in place, expanding the user base involves minimal additional expense. As a result, average costs decline sharply with scale, conferring decisive advantages on firms able to operate at very large volumes.

Unlike traditional industries, where organizational complexity may eventually generate diseconomies of scale, digital markets often exhibit persistent and extreme scale advantages. Large platforms can amortize fixed investments over vast user bases, enabling them either to operate with higher margins or to reinvest aggressively in quality improvements and innovation. This dynamic is visible in search engines, social networks, e-commerce platforms, and cloud computing services, where smaller rivals face structural cost disadvantages that are difficult to overcome.

Beyond conventional cost-based scale effects, digital markets also display **data-driven economies of scale**, which are particularly relevant in the context of artificial intelligence (AI) [Shapiro and Varian (1998)]. User activity generates large volumes of data that can be used to train algorithms, improve recommendations, personalize services, and optimize operations. More users generate more data, more data improve algorithmic performance, and better services attract more users. These data feedback loops create self-reinforcing advantages that strengthen over time and are difficult for new entrants to replicate, even with comparable technical capabilities.

Economies of scope further reinforce concentration by allowing firms to reuse key digital assets across multiple products and markets. Infrastructure, data, software capabilities, and specialized human capital can be deployed simultaneously in different services at relatively low incremental cost [Jacobides et al. (2019). This enables platforms to expand efficiently into adjacent activities, leveraging existing user bases and technological capabilities as launchpads for new services.

The reuse of data across services is particularly powerful. Information collected in one context - search queries, purchase histories, social interactions, or location data - can be exploited to enhance performance in other business lines, from advertising and content recommendation to logistics and cloud services [Crémer et al. (2019)]. These cross-service synergies increase the overall value of the platform ecosystem and make specialized, single-product competitors structurally less competitive.

The interaction of network effects, economies of scale, and economies of scope produce cumulative competitive advantages. Platforms that succeed early are able to grow rapidly, generate more data, improve quality, expand into adjacent markets, and further entrench their positions. Barriers to entry are, therefore, multidimensional: new competitors must overcome not only technological challenges, but also disadvantages in scale, data availability, user networks, and ecosystem breadth [Cabral et al. (2021)]. The probability of simultaneously surmounting all these obstacles is low, which explains the persistence of dominant positions once established.

Taken together, these forces explain why digital markets tend naturally towards high levels of concentration and why a small number of Big Tech firms have come to dominate core segments of the digital economy. This outcome is not primarily the result of firm-specific conduct, but of structural characteristics intrinsic to digital platform markets.

3. Why Big Tech enter financial services: data, networks, and the DNA loop

The expansion of Big Tech firms into financial services is not a peripheral diversification strategy, nor merely a response to regulatory opportunities. It follows directly from the economic characteristics that underpin digital platform business models. Financial services, most notably payments, but increasingly also credit, wealth-related services, and insurance distribution, are a natural extension of platform ecosystems built around data accumulation, network effects, and economies of scope.

At the core of this expansion lies what has been described as the **DNA loop**, a self-reinforcing interaction between data, network effects, and activities [BIS (2019)]. As platforms broaden the range of services they offer, they generate additional user data. These richer datasets improve service quality, personalization, and targeting, which in turn increase user engagement and attract new users. Higher

engagement strengthens network effects and creates opportunities for further expansion into adjacent activities. Financial services fit particularly well within this feedback loop.

From an economic perspective, financial services are highly complementary to existing Big Tech activities. Payments, in particular, are embedded in a wide array of digital interactions, including e-commerce transactions, app ecosystems, subscriptions, peer-to-peer transfers, and in-app purchases. By integrating payment functionalities, platforms reduce transaction frictions, enhance user experience, and increase the time users spend within their ecosystems. This deepens direct network effects and raises switching costs, reinforcing user lock-in [Farrell and Klemperer (2007)].

Payments also represent a strategic source of data. Transaction data provide granular and high-frequency information on consumption patterns, income flows, merchant activity, and behavioral traits. When combined with data already collected through search activity, social interactions, geolocation, or online purchases, payment data significantly enhance platforms' ability to profile users and tailor services [Crémer et al. (2019)]. This informational advantage is particularly valuable for the development of downstream financial services, such as consumer credit, merchant lending, or personalized financial offers.

Regulatory considerations further help explain why payment services often constitute the initial point of entry into finance. Compared to deposit-taking or lending, payment services generally face lower prudential requirements, lighter capital constraints, and more flexible licensing regimes [Frost et al. (2019)]. This makes them an attractive starting point for non-bank platforms seeking to establish a presence in financial services without immediately assuming the full regulatory burden associated with traditional financial intermediation. Once payment services are established and widely adopted, platforms can progressively expand into more regulated segments, either directly or through partnerships with licensed financial institutions.

The platform-based business models of Big Tech firms also allow for extensive cross-subsidization. Revenues generated in core activities, such as advertising, cloud services, or e-commerce, can be used to subsidize financial services, enabling platforms to offer low-cost or zero-fee payment solutions. This accelerates user adoption, reinforces network effects, and places competitive pressure on incumbent financial institutions that rely primarily on fee-based revenue models [Cornelli et al. (2023)].

Importantly, Big Tech expansion into financial services does not require replicating the full value chain of traditional finance. Platforms typically enter selectively, focusing on activities where their comparative advantages are strongest: customer acquisition, user interface design, data analytics, and ecosystem integration. Balance-sheet-intensive activities are often avoided or conducted through partnerships, allowing platforms to capture value while limiting regulatory exposure and capital requirements.

These dynamics help explain why financial services have become a natural extension of digital platform ecosystems. The extent to which Big Tech firms have moved along the financial services value chain, however, varies significantly across firms and jurisdictions, reflecting institutional and regulatory differences. Understanding how these patterns manifest in practice is essential for assessing the competitive and regulatory implications of Big Tech participation in finance.

4. Mapping Big Tech participation in financial services

The expansion of Big Tech firms into financial services has taken different forms across firms and jurisdictions, reflecting variations in institutional settings, financial system structures, and platform strategies. While the economic logic underpinning Big Tech entry into finance is broadly similar, the scope and depth of participation along the financial services value chain remain heterogeneous. Mapping these patterns is essential for understanding how Big Tech interacts with incumbent financial institutions and where competitive and regulatory tensions are most likely to arise.

Table 1 provides an overview of how major U.S. and Chinese Big Tech firms participate in key segments of the financial services ecosystem, including payments, credit and lending, asset management, and insurance. The table highlights both common entry strategies and important cross-jurisdictional differences.

A first and central observation is that payments constitute the primary entry point into financial

Table 1: Financial activities of GAFAM and BAT

Big Tech	Payments	Credit / lending	Asset management	Insurance
Google	Google Pay and embedded payment solutions	Limited (mainly partnerships)	No material activity	No
Apple	Apple Pay and wallet services	Consumer credit (Apple Card, BNPL, partnerships)	Limited (cash management via partners)	No
Meta (Facebook)	Peer-to-peer and in-app payments (limited scale)	No	No	No
Amazon	Integrated payments in e-commerce ecosystem	Merchant and consumer credit	Limited (cash management, partnerships)	Limited (embedded insurance offers)
Microsoft	Enterprise payment solutions (B2B focus)	No	No	No
Baidu	Digital payments (Baidu Wallet)	Consumer and SME lending	Wealth management products	Insurance distribution
Alibaba	Alipay and embedded payments	Consumer and SME credit	Wealth management and investment products	Insurance distribution and products
Tencent	WeChat Pay and embedded payments	Consumer and SME lending	Wealth management products	Insurance distribution

Source: Author's own analysis

services for virtually all Big Tech firms. Digital wallets, embedded payment solutions, and in-app payment functionalities are ubiquitous across both U.S. and Chinese platforms. This reflects the relatively low regulatory barriers associated with payment services, as well as their strategic value within digital ecosystems. Payments reduce transaction frictions, increase user engagement, and generate high-frequency transactional data, making them particularly attractive for platform-based firms.

Beyond payments, Big Tech participation becomes more selective. In the U.S. and Europe, involvement in credit and lending activities is generally limited and often mediated through partnerships with regulated financial institutions. Consumer credit, merchant lending, and buy-now-pay-later solutions are present, but typically without platforms assuming full balance-sheet intermediation. Activities in asset management and insurance remain marginal, fragmented, or entirely absent, and frequently rely on third-party providers rather than direct provision by the platforms themselves.

Chinese Big Tech firms, by contrast, have expanded further along the financial services value chain. In addition to payments, they play a more active role in consumer and small-business lending, wealth management products, and insurance distribution. This broader integration reflects not only differences in regulatory evolution, but also structural features of the Chinese financial system. Historically, traditional banking in China has been more heavily oriented towards lending to state-owned enterprises and large corporates, while household finance and SME lending have been comparatively underserved [Huang et al. (2020). Lower levels of retail banking penetration and limited access to formal financial services created significant unmet demand, particularly among households and small firms.

In this context, digital platforms were well positioned to fill existing gaps in financial intermediation. By leveraging extensive user networks, alternative data sources, and digital distribution channels, Big Tech firms were able to scale financial services rapidly and reach segments of the population that had limited access to traditional banking. This suggests that the depth of Big Tech penetration in financial services is likely to be greater in jurisdictions where traditional financial intermediation is less developed, and particularly in emerging economies with large unbanked or underbanked populations.

Overall, the evidence indicates that Big Tech participation in financial services is strategic and modular, rather than comprehensive. Platforms enter selectively, focusing on activities that complement their core businesses and reinforce their ecosystems. Financial services are not pursued as standalone profit centers, but as components of integrated digital environments in which data generation, network effects, and user engagement play a central role.

This mapping exercise underscores the importance of analyzing Big Tech involvement in finance from an ecosystem perspective. The competitive implications of platform-based firms cannot be assessed solely by examining individual financial services in isolation, but must instead take into account how financial functionalities interact with non-financial activities within broader digital ecosystems.

5. Benefits and risks of Big Tech participation in financial services

The growing participation of Big Tech firms in financial services has the potential to reshape retail finance in fundamental ways. As in previous episodes of structural change in the financial services sector, this development entails both important potential benefits and significant risks, which are likely to materialize at different horizons. A balanced assessment of these effects is, therefore, a necessary step before turning to questions of regulatory design and policy responses.

5.1 Benefits: competition, innovation, and inclusion

In the short term, Big Tech participation in financial services is likely to intensify competition, particularly in customer-facing segments such as payments, distribution, and retail financial services. In many jurisdictions, these segments have long been characterized by limited contestability, persistent inefficiencies, and relatively high costs. The entry of large digital platforms, equipped with extensive user bases and advanced technological capabilities, may force incumbents to improve service quality, reduce prices, and accelerate digital transformation, at least in the initial phase of entry.

Big Tech firms may also act as important drivers of innovation. Their strengths in data analytics, automation, and user-centered design facilitate the development of new financial products and delivery models, such as embedded finance, real-time payments, and personalized financial services. By integrating financial functionalities into existing digital ecosystems, platforms can reduce transaction frictions and improve convenience for consumers and firms.

In addition, Big Tech participation may contribute to financial inclusion, particularly in jurisdictions where traditional banking systems have historically undersupplied households and small businesses. By leveraging alternative data sources, digital distribution channels, and existing user relationships, platforms can extend access to basic financial services - especially payments and small-scale credit - to previously underserved segments. Evidence from emerging economies suggests that this effect can be particularly relevant where banking penetration is low and unmet demand is significant.

Taken together, these potential benefits help explain why outright exclusion of Big Tech firms from financial services is neither realistic nor desirable. Increased competition, innovation and inclusion represent tangible welfare gains that policymakers should seek to preserve.

5.2 Risks: market structure, incentives, and structural asymmetries

Alongside these benefits, Big Tech participation in financial services raises a number of structural concerns that become more salient over the medium- to long-term. Digital platform markets tend to exhibit strong network effects and economies of

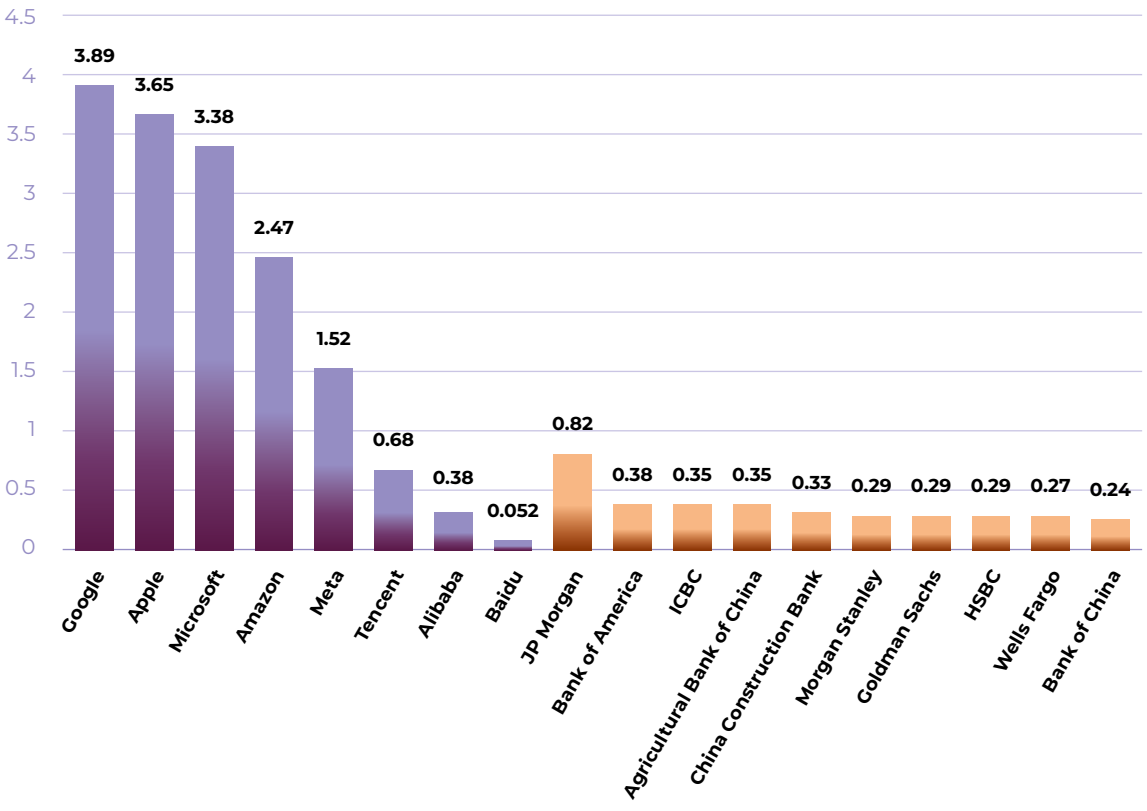
scale, which favor winner-takes-most dynamics. While entry may initially enhance competition, the integration of financial services into dominant digital ecosystems risks reinforcing user lock-in and reducing contestability over time.

A central concern is that Big Tech firms may come to dominate the origination and distribution of financial services, particularly credit to households and small and medium-sized enterprises, while traditional banks are progressively confined to funding and balance-sheet provision. Given their control over customer interfaces, superior access to data and ability to embed financial services into broader digital ecosystems, platforms are well positioned to capture the most profitable segments of retail finance. In such a configuration, value creation and customer relationships increasingly shift towards platforms, while banks risk evolving towards "narrow" or utility-like institutions with declining margins and reduced strategic autonomy (Padilla, 2020).

This progressive unbundling of financial intermediation, whereby origination and distribution are separated from funding, may generate incentive distortions. Platforms that do not retain meaningful exposure to credit risk may prioritize scale, engagement, or cross-selling over long-term loan quality. Over time, weakened screening incentives may exacerbate moral hazard and adverse selection problems, particularly if financial services primarily serve to reinforce other profitable platform activities rather than standing as independent business lines [De la Mano and Padilla (2018)].

These risks are further amplified by significant asymmetries in economic scale between Big Tech firms and traditional financial institutions. Figure 1 illustrates the stark differences in market capitalization between major Big Tech companies and the world's ten largest banks by market capitalization. Market capitalization proxies firms' capacity to absorb losses, sustain prolonged periods of low profitability, and cross-subsidize activities across

Figure 1. Market capitalization of GAFAM, BAT and the world's ten largest banks by market capitalization [in trillion dollars (January 21, 2026)]



Source: Author's own analysis based on companies' market capitalizations

business lines. When firms with such divergent economic profiles compete in overlapping segments of financial services, competitive dynamics may be affected even in the absence of direct balance-sheet risk.

The extensive use of data and algorithms further reinforces these structural asymmetries. While data-driven credit assessment can improve efficiency and reduce information asymmetries, the concentration of vast datasets across multiple non-financial activities confers advantages that are difficult for traditional financial institutions to replicate. Differences in data governance regimes may, therefore, strengthen market power and exclusionary dynamics beyond what would be justified by efficiency considerations alone [Padilla (2020)].

Finally, even when Big Tech firms seek to avoid direct balance-sheet exposure, their growing role in financial intermediation may have system-wide implications. A financial system in which origination, distribution, and customer relationships are increasingly controlled by platforms, while funding is provided by regulated institutions, may become more vulnerable to correlated risk-taking and shifts in incentives. Historical experience suggests that such configurations can increase fragility over time, even if short-term effects initially appear pro-competitive.

5.3 From benefits and risks to policy questions

The coexistence of short-term benefits and longer-term risks gives rise to a clear policy trade-off. Big Tech firms can enhance competition, innovation, and inclusion, yet may also contribute to concentration, incentive distortions, and structural asymmetries over time. This tension has been widely recognized in the literature analyzing Big Tech entry into financial services, which emphasizes that long-term outcomes depend critically on how platform-based models interact with existing financial intermediation structures [De la Mano and Padilla (2018), Padilla and Trento (2019)].

These considerations shift the focus from the question of whether Big Tech firms should participate in financial services to how such participation should be addressed from a policy design perspective. In particular, they raise questions about the existence of a level playing field between platform-based firms and traditional financial institutions, and about the adequacy of current regulatory frameworks to deal with the interaction between competition, data, and financial stability. The next section turns to these issues, discussing how policy design can respond to the challenges posed by Big Tech participation in financial services.

6. Policy design challenges: ensuring a level playing field in financial services

The trade-offs identified in the previous section highlight that the policy challenge posed by Big Tech participation in financial services is not whether such firms should be allowed to operate in financial markets, but how regulatory frameworks should be designed to preserve competition and innovation while limiting structural distortions and longer-term risks. Addressing this challenge requires a policy approach that goes beyond narrow institutional boundaries and reflects the economic reality of platform-based firms operating across financial and non-financial activities.

6.1 Activity-based regulation as a necessary, but insufficient, baseline

A natural starting point for policy design is the principle of activity-based regulation, according to which the same regulatory requirements should apply to any entity performing the same regulated financial activity. This principle supports competitive neutrality, limits regulatory arbitrage, and avoids discriminating between providers based solely on institutional form. Applying comparable prudential, conduct, and consumer protection rules to all entities engaging in payments, lending, or other regulated activities is, therefore, a necessary baseline for ensuring a level playing field.

However, recent policy analysis has emphasized that activity-based regulation, while necessary, is often not sufficient when applied to Big Tech groups [Collot et al. (2024)]. Platform-based firms combine financial and non-financial activities within highly integrated ecosystems, leverage data generated across multiple markets, and benefit from strong network effects and economies of scope. Focusing exclusively on individual financial activities could entail overlooking risks and com-

petitive distortions that emerge at the level of the group rather than within isolated entities.

6.2 Mixed-activity groups and consolidation gaps

A central challenge for policy design arises from the presence of mixed-activity groups, in which regulated financial activities coexist with large-scale non-financial operations. In such structures, risks may be shifted across entities, and economic power may be exercised through group-level strategies that are not fully captured by entity-based supervision. Fragmented regulatory frameworks, in which different financial activities are supervised separately and often at national level, may struggle to provide a comprehensive view of these group-wide dynamics.

This challenge is particularly visible in segments where Big Tech firms tend to concentrate, such as payments, e-money issuance, and certain forms of non-bank lending. Differences in national regimes and supervisory practices can facilitate structural arbitrage, allowing groups to organize their activities in ways that minimize regulatory burdens without reducing underlying risks. Policy work has, therefore, increasingly pointed to the need for more harmonized and consolidated approaches to supervision, including mechanisms that enable authorities to assess risks at group level rather than exclusively at the level of individual entities [Collot et al. (2024)].

6.3 Data leverage, reciprocity, and competitive neutrality

Data constitutes a key source of competitive advantage for Big Tech firms operating in financial services. Platforms typically combine financial and non-financial data across multiple activities, enhancing their ability to screen customers, personalize offers, and integrate financial services into broader digital ecosystems. While such data-driven models can improve efficiency and user experience, they also raise concerns about competitive neutrality when data access and sharing obligations are asymmetric.

In several jurisdictions, traditional financial institutions are subject to extensive data-sharing requirements, while large platforms effectively retain control over vast datasets generated outside the financial services sector. This asymmetry may reinforce structural advantages and affect entry dynamics in financial services. From a policy perspective, ensuring a level playing field, therefore, extends beyond prudential requirements to include data governance rules, encompassing issues such as reciprocity, limits on data combination, and safeguards for user control and transparency. As highlighted in recent policy debates, the level playing field in finance is not only a prudential issue, but also a data-driven one [(Collot et al. (2024)].

6.4 Operational resilience, conduct, and institutional coordination

The growing reliance of financial services on digital infrastructures further underscores the importance of operational resilience and conduct regulation in the context of Big Tech participation. Large platforms often act as critical providers of ICT services, including cloud computing, to financial institutions. Ensuring the resilience of these infrastructures and managing concentration risks requires regulatory tools that extend beyond traditional financial supervision.

More broadly, Big Tech participation in financial services cuts across multiple regulatory domains, including financial regulation, competition policy, data protection, and digital market regulation. No single authority can adequately address all relevant risks in isolation. Effective policy design, therefore, requires structured institutional coordination among regulators with different mandates, as well as mechanisms for information sharing and joint oversight. Without such coordination, regulatory gaps may persist, undermining efforts to ensure competitive neutrality and systemic resilience.

Taken together, these policy design considerations suggest that ensuring a level playing field in data-driven financial services cannot be achieved through a single instrument. It requires a coherent combination of activity-based requirements, effective oversight of mixed-activity groups, robust data governance, and institutional coordination. These principles are not merely abstract: they are already shaping concrete legislative choices in the E.U. The next section, therefore, “lands” this framework in a live legal and policy debate currently unfolding in the E.U., namely the proposed Financial Data Access Regulation (FiDA) and the question of whether firms designated as gatekeepers under the Digital Markets Act should be restricted from accessing financial data through the new Financial Information Service Providers (FISP) regime.

7. Gatekeepers and Open Finance: the FiDA debate as a case study

The policy challenges associated with Big Tech participation in financial services are particularly visible in the E.U.'s ongoing legislative debate on the proposed FiDA regulation. FiDA represents the E.U.'s attempt to extend the data-sharing logic introduced under Open Banking to a broader set of financial services, thereby enabling the development of Open Finance.

7.1 From Open Banking to Open Finance: the scope of FiDA

Open Banking, as introduced under the revised Payment Services Directive (PSD2), requires banks to grant licensed third-party providers access to payment account data, subject to user consent. While this framework has facilitated innovation in payments and account aggregation, its scope remains limited to payment accounts.

FiDA seeks to go significantly further by establishing a horizontal framework for financial data sharing across a wide range of financial services, including savings, investments, insurance, and pensions. Its objective is to promote competition, innovation, and consumer choice by enabling authorized third parties to access customer financial data in a secure and standardized manner, always subject to explicit user consent.

To this end, FiDA introduces a new category of regulated entity: FISPs. FISPs are subject to authorization and ongoing supervision and must comply with requirements related to governance, operational resilience, data security, and consumer protection. In this respect, FiDA does not create an unregulated access regime, but rather embeds Open Finance within the E.U.'s existing financial supervisory architecture.

7.2 Gatekeepers under the DMA and their relevance for FiDA

The E.U. has recently adopted the Digital Markets Act (DMA), which introduces an ex-ante competition framework targeting large digital platforms designated as gatekeepers. Gatekeeper status under the DMA is determined on the basis of quantitative and qualitative criteria, including scale, control over core platform services, and an entrenched position in the market.

Importantly, gatekeeper designation is service-specific. A company may be designated as a gatekeeper for certain core platform services, such as online search engines, app stores, or social networks, without being a gatekeeper across all its activities. As a result, several large technology companies are gatekeepers in some digital markets but not in others, and DMA designation does not directly relate to the provision of financial services.

Nevertheless, because many DMA-designated gatekeepers are large technology firms with established digital ecosystems, the FiDA debate has increasingly focused on whether gatekeepers should be excluded from accessing financial data through the FISP regime. In practice, such an exclusion would primarily affect Big Tech firms seeking to participate in Open Finance.

Table 2 summarizes the companies currently designated as gatekeepers under the DMA, the core platform services for which they are designated, and their existing or potential involvement in financial services.

7.3 Proposals to exclude gatekeepers from FiDA and why they raise concerns

During the legislative process, proposals have emerged to categorically exclude companies designated as gatekeepers under the DMA, and, in some versions, entities owned or controlled by them, from obtaining FISP licenses. The concern underpinning these proposals is that gatekeepers could leverage their ecosystem advantages to dominate emerging Open Finance markets, combine financial data with extensive non-financial datasets, and exacerbate competitive asymmetries vis-à-vis regulated incumbents.

While these concerns are legitimate, translating them into blanket restrictions raises a set of legal, economic, and regulatory coherence issues [Arnal and Andersson (2025)].

First, categorical exclusions risk constituting a restriction on the freedom to provide services within the internal market under Article 56 of the Treaty on the Functioning of the European Union (TFEU). By prohibiting a defined category of firms from providing financial information services

Table 2: Companies currently designated as gatekeepers under the DMA

Gatekeeper	Core platform service
Alphabet	Google Play
Alphabet	Google Maps
Alphabet	Google Shopping
Alphabet	YouTube
Alphabet	Android Mobile
Alphabet	Alphabet's online advertising service
Alphabet	Google Chrome
Amazon	Marketplace
Amazon	Amazon Advertising
Apple	App Store
Apple	iOS
Apple	Safari
Apple	iPadOS
Booking	Booking.com
ByteDance	TikTok
Meta	Facebook
Meta	Instagram
Meta	WhatsApp
Meta	Messenger
Meta	Meta Ads
Microsoft	LinkedIn
Microsoft	Windows PC OS

Source: Author's own analysis based on Arnal and Andersson (2025)

across the E.U., such exclusions may "impede or render less attractive" the exercise of that Treaty freedom and, therefore, require robust justification.

Second, the restrictions may also raise fundamental rights concerns under the Charter of Fundamental Rights, notably the freedom to conduct a business (Article 16) and the principle of equality and non-discrimination (Article 20). Gatekeepers and non-gatekeepers seeking FISP authorization are in a comparable situation with respect to FiDA's authorization logic: their DMA designation relates to specific core platform services and is not, in itself, indicative of their capacity to comply with FiDA's operational and consumer protection requirements. Categorically denying access on the basis of DMA status, therefore, risks creating a two-tier market access regime without objective justification.

Third, and closely related, blanket exclusions must satisfy a strict proportionality assessment [Article 5(4) TEU and settled case law]. Such measures may fail the appropriateness test if they do not effectively address the underlying concerns and may even reduce competition by excluding innovative entrants. They may also fail the necessity test because FiDA already contains a dense set of safeguards - licensing and supervision of data users, standardized technical interfaces, explicit and revocable consumer consent, and data protection requirements - which allow risks to be mitigated without wholesale exclusions. Finally, categorical bans may impose disproportionate burdens relative to the policy objective by excluding firms regardless of their actual risk profile or compliance capacity.

Fourth, the proposed exclusions raise regulatory coherence concerns. The DMA is explicitly designed as a targeted instrument that applies to specific designated core platform services, not to companies as a whole. Extending DMA-based restrictions to a different policy domain - Open Finance - stretches the logic of gatekeeper designation beyond its intended context and risks undermining legal certainty.

Finally, blanket restrictions entail substantial implementation challenges. Determining which entities fall within the scope of restrictions, monitoring complex group structures over time, and addressing future changes in DMA designations may generate administrative complexity and legal uncertainty. Such rules may also invite regulatory arbitrage, as groups restructure to circumvent exclusions, potentially making oversight harder rather than easier.

These considerations do not imply that the concerns motivating the proposals should be dismissed. Rather, they point to the need for a more proportionate response grounded in FiDA's own risk-based logic, namely participation under reinforced safeguards instead of categorical exclusion.

7.4 A proportionate alternative: participation under reinforced safeguards

An alternative approach, more consistent with the policy principles discussed in previous sections, is to allow gatekeeper-designated firms to participate in Open Finance under reinforced supervisory conditions, rather than excluding them outright. Under this approach, all firms would be subject to the same baseline FISP requirements, while supervisors would retain discretion to impose additional obligations where justified by scale, structure, or data-related risks.

Such safeguards could include enhanced governance and risk-management requirements, stricter data ring-fencing obligations, and closer scrutiny of how financial and non-financial data are combined. Where necessary, supervisors could also require increased transparency regarding business models and data usage, ensuring that participation in Open Finance does not translate into unfair leveraging of existing market power.

Crucially, this approach does not rely on automatic or blanket data reciprocity obligations for gatekeepers. Instead, it preserves regulatory flexibility by addressing data-related concerns on a case-by-case basis, in line with the firm's actual activities and risk profile. This is more closely aligned with FiDA's objectives and with the E.U.'s broader commitment to proportionate, activity-based regulation.

7.5 Implications for the development of Open Finance in the E.U.

The FiDA debate highlights a broader question about the future direction of data-driven competition in the E.U. If the policy objective is to liberalize access to data, empower data holders, and enable users to determine how their data are reused, focusing exclusively on the financial sector may prove conceptually and economically limiting.

Many of the competitive advantages associated with Big Tech participation in financial services do not stem from financial data alone, but from the ability to combine financial and non-financial data across sectors. Addressing these advantages solely within the perimeter of Open Finance risks treating the symptoms rather than the underlying source of data-driven market power.

From this perspective, a more coherent long-term approach may lie in the gradual development of a cross-sectoral data access framework, anchored in user consent, interoperability, and robust governance safeguards. Such an approach would allow the principle of data empowerment to be applied more consistently across the economy, rather than confining it to individual sectors in isolation.

Importantly, a cross-sectoral perspective could also provide a more neutral and proportionate way to address concerns about reciprocity. Rather than imposing ad-hoc or sector-specific obligations on particular categories of firms, reciprocity could be embedded more broadly in horizontal data governance rules, ensuring that access to valuable datasets is balanced by corresponding obligations, regardless of the sector in which data are generated.

This does not imply that FiDA should itself become a vehicle for cross-sectoral data liberalization. Rather, it suggests that FiDA should be understood as a building block within a wider data governance architecture, one that recognizes the increasingly cross-sectoral nature of data-driven competition. Anchoring Open Finance within such a broader framework may ultimately prove more effective than attempting to resolve complex ecosystem-level issues through sector-specific exclusions.

8. Conclusions

The expansion of Big Tech into financial services is not an anomaly, nor a transient phenomenon. It is the predictable outcome of economic forces that shape digital markets and platform-based business models. Financial services, particularly payments and data-intensive activities, fit naturally within digital ecosystems built around scale, networks, and data reuse. Attempts to address Big Tech participation in finance without acknowledging this underlying logic risk being ineffective or internally inconsistent.

This paper has shown that Big Tech entry into finance entails a genuine trade-off. In the short term, platform-based firms can enhance compe-

tition, accelerate innovation, and expand access to financial services. Over time, however, their ecosystem advantages may reinforce concentration, alter incentives along the financial intermediation chain, and create structural asymmetries vis-à-vis traditional financial institutions. These dynamics do not lend themselves to binary policy responses.

From a policy design perspective, ensuring a level playing field requires moving beyond narrow, entity-based approaches. Activity-based regulation remains a necessary foundation, but it is insufficient when applied in isolation to firms operating across tightly integrated financial and non-financial activities. Data governance, group-level oversight, and institutional coordination are equally central to addressing the challenges posed by platform-based models.

The E.U. debate on Open Finance and FiDA illustrates these tensions particularly clearly. Proposals to exclude DMA-designated gatekeepers from accessing financial data reflect legitimate concerns about data leverage and market power. Yet categorical exclusions risk undermining proportionality, regulatory coherence, and the very objectives of Open Finance. A more effective approach lies in allowing participation under reinforced safeguards, anchored in supervision rather than exclusion.

More broadly, the FiDA debate points to a deeper policy question. If the ultimate objective is to empower users, promote data portability, and foster competition in a data-driven economy, a purely sectoral approach may prove limiting. Competitive advantages increasingly arise from the ability to combine data across activities and sectors. Addressing these dynamics may ultimately require a more horizontal and cross-sectoral perspective on data access and governance.

Big Tech in finance is, therefore, not only a question of financial regulation. It is a test case for how modern regulatory frameworks can adapt to an economy organized around platforms, data, and ecosystems. Striking the right balance between openness and safeguards will be central to shaping the future of both financial services and the digital economy more broadly.

References

Arnal, J., and F. Andersson, 2025, "Should Open Finance exclude gatekeepers in the EU? A proportionate path through reinforced supervision," SUERF policy note no. 385, SUERF – The European Money and Finance Forum, https://tinyurl.com/3axu8e5d

BIS, 2019, "Big tech in finance: opportunities and risks," in BIS Annual Economic Report 2019 (Chapter III), Bank for International Settlements, https://tinyurl.com/33bbf9wu

Cabral, L., J. Haucap, G. Parker, G. Petropoulos, T. Valletti, and M. Van Alstyne, 2021, "The EU Digital Markets Act: a report from a panel of economic experts," Publications Office of the European Union, https://tinyurl.com/26r78uxr

Chandler, A. D., 1990, Scale and scope: the dynamics of industrial capitalism, Harvard University Press

Collot, S., A. Machover, and E. Rocher, 2024, "The growth of big techs in the financial sector: which risks, which regulatory responses?" Banque de France, International Affairs Directorate

Cornelli, G., J. Frost, L. Gambacorta, R. Rau, R. Wardrop, and T. Ziegler, 2023, "Fintech and big tech credit: drivers of the growth of digital lending," Journal of Banking & Finance 148, 106742

Crémer, J., Y.-A. de Montjoye, and H. Schweitzer, 2019, "Competition policy for the digital era," European Commission

De la Mano, M., and J. Padilla, 2018, "Big Tech banking," Journal of Competition Law & Economics 14:4, 494-526

Farrell, J., and P. Klemperer, 2007, "Coordination and lock-in: competition with switching costs and network effects," in Armstrong, M., and R. Porter (eds.), Handbook of industrial organization, volume 3, Elsevier

Frost, J., Gambacorta, L., Huang, Y., Shin, H. S., & Zbinden, P. (2019). BigTech and the changing structure of financial intermediation. Economic Policy, 34(100), 761-799. https://doi.org/10.1093/epolic/eiaa003

Huang, Y., C. Lin, P. Wei, and Z. Xu, 2020, "Financial inclusion and urban-rural income inequality: long-run and short-run relationships," Emerging Markets Finance and Trade 56:2, 457-471

Jacobides, M. G., C. Cennamo, A. Gawer, 2019, "Towards a theory of ecosystems," Strategic Management Journal 39:8, 2255-2276

Katz, M. L., and C. Shapiro, 1985, "Network externalities, competition, and compatibility," American Economic Review 75:3, 424-440

Padilla, J., 2020, "BigTech "banks", financial stability and regulation," Financial Stability Review, 38, 11-25

Padilla, J., and S. Trento, 2019, "No barbarians at the gate? The relatively slow progress of Big Techs in EU and US retail banking," Concurrences 4, 42-47

Rochet, J.-C., and J. Tirole, 2003, "Platform competition in two-sided markets," Journal of the European Economic Association 1:4, 990-1029

Shapiro, C., and H. R. Varian, 1998, Information rules: A strategic guide to the network economy, Harvard Business School Press

How data and AI shape financial services

Data risk: a strategic enterprise challenge that requires a holistic board-level offensive

Jim Halcomb,
Chief Research and Development Officer, EDM Association

Scott Beange,
Head of Data Management and Cyber Strategy, Projective Group

ABSTRACT

With the number of organizations attempting to integrate Artificial Intelligence (AI) within their operations increasing, data risk - which encompasses data quality, availability, governance, and security – has been transformed from a localized IT concern into a strategic enterprise challenge requiring a board-level offensive. In this article, we will analyze the business implications of data risk, including financial losses, regulatory consequences, operational disruptions, reputational damage, loss of resilience and trust, and personal accountability for executives. We will also discuss the key steps that should be considered, from a strategic holistic perspective, in order to address the rapidly changing data risk challenges faced by organizations.

1. Introduction

As organizations integrate AI at scale, the definition of data risk has expanded to include not just the loss of data, but the integrity and governance of the information fueling decisions across the business.

Data risk typically refers to a type of risk where the likelihood of a data-related event, which could adversely impact an organization's ability to achieve its goals, is measured by the probability of the event occurring and the severity of its consequences (which can be negative or positive). It is the potential for deviations from expected outcomes, encompassing threats (like losses) and opportunities (potential gains), that requires identification, analysis, and mitigation.

Adverse data-related events include data defects. Defective data is often the symptom of a design flaw or human error; however, irrespective of the cause, if data is defective, it can have serious implications for the organization – the old adage of garbage in, garbage out can certainly apply.

Put differently, data risk is the flipside of data quality.

Since design flaws are typically caught in testing before being released into production, human error tends to be the main culprit at the application level. Given that applications are typically not designed to talk to each other, in situations where we get a large number of defective data organizations can end up having to create operational workarounds to manage the flow of data across the value chain. Consequently, data risk is part and parcel of operational risk.

When detected, data defects create bottlenecks in business processing. The further in the value chain that these defects pass through the quality checks, the higher the potential cost to the organization. At a minimum, bottlenecks result in inefficiency and delays. However, when defects go unresolved, or worse undetected, they undermine decision-making, regulatory compliance, financial integrity, and trust.

Data risk can also appear when data is unavailable because systems are down. When critical information cannot be accessed when it is needed, it prevents the organization from operating, deciding, or complying effectively. Because modern enterprises are now "hyper-connected", a failure in data availability does not just stop in one department, it can trigger a cascading collapse of processes across the entire enterprise. When core systems go down, even though the data exists, it cannot be used. In a digital-first economy, this is equivalent to a physical store being padlocked from the outside.

Date loss, or breached data - which can be due to cyber-attacks, insider threats, poor access controls, third-party failures, or weak data handling practices - can also create data risk. When information is destroyed, stolen, exposed, or accessed without authorization it can undermine trust, compliance, and control, with potentially severe and long-lasting consequences. Ultimately, lost or breached data represents a failure of trust. Once data is exposed or destroyed, the organization must not only recover systems, but also rebuild confidence with regulators, customers, and the market.

Poor data governance is no longer viewed as a "messy office" problem; it is a control failure that acts as a heavy drag on business performance. When ownership is unclear and controls are weak, data becomes a liability that grows exponentially. In many organizations, data is orphaned: it has no clear named owner responsible for its quality, security, or lifecycle. Poor data governance is the silent killer of AI innovation. You cannot automate what you do not own, and you cannot secure what you do not govern.

"Data ownership" is one of the most critical responsibilities that most companies have as yet not written about.

In summary, data risk has transformed into a high-stakes driver of enterprise valuation and board-level accountability, moving beyond its technical origins to encompass the integrity, availability, and governance of information critical to decision-making.

2. Business implications of data risk

Data risk is no longer considered just an IT headache; it is a fundamental business threat that impacts everything from the balance sheet to brand equity. When data is lost, corrupted, or stolen, its implications span across the entire enterprise.

It is common for businesses to blame IT for data quality errors; however, the vast majority of data defects are caused by applications that were not designed to talk to each other. This interopera-

bility challenge poses significant operational risk to organizations, and requires focus from, and collaboration with, the business to address.

Moreover, executives should frame data risk in terms of business outcomes, not IT systems. The business impact of data failure is both the direct and the indirect damages caused by data being inaccurate, unavailable, insecure, or misused.

The questions that most businesses need to ask themselves are:

- What happens if critical data is wrong, unavailable, leaked, or manipulated?
- Which business decisions rely on this data (pricing, credit, liquidity, capital, AML, and reporting)?

Data risk is a strategic risk, with data failure having the potential to turn information from an asset into a liability: driving poor decisions, regulatory exposure, financial loss, operational disruption, and reputational damage.

2.1 Core business implications

2.1.1 Lack of trust

Lack of trust in data is no longer just a hurdle for the IT department, it is a systemic business risk that acts as a "tax" on every decision and operation. When data trust collapses, the organization flies blind, reverting to slower, more expensive, and less effective ways of working.

The average employee loses 4 working hours per week (roughly 10% of their time) simply resolving data discrepancies or manually validating information before use [Snaplogic (2023)].

The rise of unverified AI-generated data has led 50% of organizations to adopt a "zero trust" posture for data management. If the data is not verified at source, leaders refuse to act on AI recommendations, rendering expenditure in AI R&D useless [Gartner (2026)].

When a data failure occurs, the damage to the infrastructure is often repaired long before the executive trust is restored. For the C-suite, data is the "eyes and ears" of the organization. If these senses fail once, leadership often reverts to manual, instinct-based decision-making that slows the entire company down.

KPIs and metrics are challenged rather than acted upon. Making it difficult for senior leaders to hold their teams to account.

2.1.2 Financial loss

When data fails, it is rarely just a "glitch", it is a line item on a P&L statement. Whether it is through bad quality, poor governance, or a total system outage, the fallout is usually measured in cold, hard cash.

In financial services, downtime is measured in seconds because of the liquidity ripple effects. High-impact outages in the financial services sector are estimated to cost around U.S.$1.8 mln per hour [New Relic (2026)], with the average annual losses due to downtime (including lost revenue, regulatory fines, and legal settlements) being around U.S.$152 mln [Resolve (2025)].

The average cost of a data breach reached U.S.$4.44 mln in 2025, with the U.S. average being more than twice that figure, at U.S.$10 mln, predominantly due to higher detection and escalation costs [Bonnie (2025)]. The daily cost of downtime for financial services firms is estimated to be U.S.$5.08 mln, with recovery from ransomware taking days while the tail-end costs of cleaning up the mess persisting for months [IBM (2025)].

2.1.3 Regulatory and legal consequences

In 2026, the regulatory and legal landscape for data failure has shifted from reputational risk to existential financial risk. Regulatory bodies like the Financial Conduct Authority (FCA), Securities and Exchange Commission (SEC), and the E.U. regulators now view data quality and integrity as fundamental to financial stability.

Regulators are no longer just fining organizations for data breaches; they are also penalizing firms for data deficiencies that lead to operational failure.

Research shows that data deficiencies contributed to over 68% of FCA enforcement cases for Anti-Money Laundering (AML) failures. These data deficiency fines alone have totaled over £430 million in the U.K. recently [Kyckr (2026)].

If regulators receive inaccurate data, they view it as evidence that there are weak controls within the organization. This can lead to fines and other penalties. Regulators rarely just stop at fines, they can force organizations into costly multi-year remediation programs based on regulator mandated timelines and milestones. Regulators may also require

organizations to hold increased capital, meaning that it cannot be invested in business operations.

Future regulatory submissions will be scrutinized more closely, leading to ongoing operational and compliance burdens.

When critical information is lost or becomes inaccessible, regulators take this seriously because it threatens reporting accuracy, operational integrity, and consumer protection.

2.1.4 Operational disruption

Data failure is no longer viewed as a temporary pause, but as an event that has the potential to cause operational paralysis. With modern businesses becoming increasingly AI-interdependent, a data failure in one department can trigger a cascading effect across the organization.

Poor data is not just a reporting problem, it directly disrupts operations, slows execution, and increases costs across the organization.

The "manual workaround" tax forces staff to spend time verifying, reconciling, or manually correcting information, hence focusing on firefighting instead of creating value for the business. There is also the potential for disruptions to directly impact customers, with inaccurate account information, inconsistent service, and reputational damage.

Poor data governance often means no clear data ownership, inconsistent data standards, and weak controls, creating systemic risk in reporting and decision-making.

2.1.5 Reputational damage

For a data-driven enterprise, a data failure is essentially a public admission of loss of control. When your data infrastructure breaks, the market assumes your leadership is either negligent or incompetent.

When customers encounter incorrect or inconsistent data, trust is eroded, causing complaints to become amplified through public forums. As a result of the so-called "screenshot" effect, a single screenshot of a data failure can reach millions of people on social platforms, creating a permanent digital footprint that buries your positive marketing efforts for years.

In the age of instant-portability (open banking, easy SaaS migration), customer loyalty is paper-thin. Research indicates that 81% of customers would stop engaging with a brand online following a data breach or significant data-driven error [Nasdaq (2019)].

Publicized data incidents attract media attention and social scrutiny. This can lead to reputational damage affecting sales and customer loyalty. There can be a negative impact on board and executive credibility as well, which in turn impacts investor confidence. Investors rely on accurate, timely data for reporting and decision-making, data failures signal weak controls within the organization.

Partners and suppliers may also lose confidence with reduced collaboration, delayed deals, or renegotiation of contracts under less favorable terms.

2.1.6 Loss of resilience and agility

The cost of data failure has evolved from a simple IT expense into an anchor that drags down a company's ability to move fast and survive shocks.

Gartner predicts that 60% of AI projects will be abandoned by the end of 2026 because the underlying data is too inconsistent [Edjlali (2025)]. Without "AI-ready" data, the speed promised by automation becomes a liability, making bad decisions faster.

Data is central to decision-making, operations, and strategic adaptability; hence failures can ripple across the entire organization. Data failures can halt business critical processes, with a single point of failure cascading across multiple areas.

Executives rely on real-time or near-real-time data. If the data is inaccurate, outdated, or inaccessible, strategic and tactical decisions are delayed or misinformed. Poor data quality impedes collaboration across departments, slowing response to opportunities or threats. Agility in adapting products, services, or pricing depends on trustworthy customer, market, and operational data. Data failure undermines this responsiveness.

Data-driven innovation (AI, analytics, product development) is compromised if the underlying data is flawed.

2.1.7 Personal accountability for executives

Personal accountability for data failure has shifted from a corporate slap-on-the-wrist to a direct legal and career risk for the individual. Regulatory frameworks such as the General Data Protection Regulation (GDPR), the Digital Operational Resilience

Act (DORA), the NIS2 Directive, the Basel Committee on Banking Supervision standard number 239 (BCBS239), and the U.K.'s Senior Managers and Certification Regime (SM&CR) now treat data negligence as a personal failure of oversight.

Under the NIS2 Directive (fully applicable across the E.U. by late 2024/2025), national authorities can temporarily ban executives from exercising managerial functions at the C-suite level if the organization fails to implement mandatory risk management measures.

In the U.K., the FCA uses the "fitness and propriety" test, where a significant data failure can lead to a "prohibition order", effectively ending an individual's career in regulated financial services.

Under DORA and the SEC's 2024/25 rules, boards are required to approve and monitor digital resilience strategies personally.

Senior Manager Regimes (SMRs) legally assign personal accountability to senior executives for failures in their areas of responsibility, including data failures. Senior managers risk regulatory sanctions and personal consequences.

The personal accountability of executives is directly affected by data failure because executives are ultimately responsible for the reliability, governance, and use of data in the organization. Executives can be held personally liable if data failures lead to violations of regulations, especially where there is negligence in data governance.

Executives are accountable for ensuring that decisions are made on accurate and reliable data. Poor data that leads to financial loss, misreporting, or strategic missteps can reflect a failure in leadership. Boards expect executives to identify, mitigate, and report risks, with data failures often indicating gaps in risk management practices fiduciary duty.

3. Key steps to reduce your data risk

3.1 Critical data identification

Not all data is equal. Some data elements are more critical, and yet many organizations struggle to agree on a common definition of criticality. Let us recall that data quality underpins data risk, and that the most common data quality issue is the breakdown of interoperability between systems that were not designed to "talk" to each other. The interoperability challenge results in workarounds developed and implemented through ad-hoc solutions by business users.

Workarounds are informal solutions built by business users to address the challenge of keeping data flowing from system to system. As such, workarounds represent significant risks to operations, and IT departments generally fail to resolve them. Again, this is due to an underestimation of the data quality challenge, because it is emerging between, and across, the business application layer. In other words, interoperability is a data problem that tends to fall between the cracks.

IT departments tend to see their world through the lens of projects, whereas data is a corporate asset that must be managed in the aggregate.

Managing data in the aggregate means that data requirements must be developed independent of any single application. This perspective provides a key to identifying critical data. Again, at the application level, the determination of criticality with respect to specific data elements often turns out to be a dead end. For any single application, it is hard to emphasize some data elements over others, because the system will likely break down if any data defects occur. This is especially true for transactional and accounting systems. However, taking a bird's eye view across applications, the data most in common across systems are often the most critical.

That said, identifying common data elements across systems is not a trivial exercise. Sometimes applications use the same data labels for different things. More often, the same things are labelled differently from application to application. This problem has been exacerbated by the increased adoption of off-the-shelf solutions such as SaaS. In both cases, shared meaning is required to ensure what looks like apples-to-apples is apples-to-apples. While commonality is important in the determination of criticality, not all common data are equally critical.

Some common data domains are obvious. Take for example "customer". Every organization has customers, and most organizations have multiple interactions with customers that are recorded by different people, in different departments, using different applications, some of which have no effective data pipelines connecting them. Aside from the more obvious domains, a general approach is

to start with cross-organizational reporting that aggregate data across disparate sources, e.g., risk reporting, financial reporting, regulatory reporting. Prioritize these reports, identify data requirements common across them, then prioritize the ones that have the greatest number of disparate sources, and lastly tackle that remaining set based on aggregate defects.

The identification of, and accountability for, critical data is a core function of data governance. The chief goal of data governance is to ensure an effective data control environment is in place.

3.2 Effective data governance

Demand for high quality, well-controlled data easily over-taxes most data governance capabilities. Data governance relies heavily on manual contributions from the business that is seen as additive, making it prone to frequent de-prioritization. Consequently, data governance is always in danger of grinding to a halt without regular check-ins from senior executives and, just as importantly, adequate automation and reporting to instill confidence in data control compliance.

Over the course of the past decade, organizations have increasingly adopted and utilized hyperscalers to address the complex and computationally expensive demands of automated data control. In response to this need, the EDM Association established the Cloud Data Management Capabilities (CDMC) framework to provide guidance for establishing an automated data control environment.

CDMC identifies and defines the following automated controls:

- Data control compliance.
- Data ownership.
- Authoritative data sources.
- Data sovereignty.
- Cataloguing.
- Classification.
- Data access and entitlements.
- Data consumption purpose.
- Security controls.
- Data protection impact assessments.
- Data retention.
- Data quality.
- Cost metrics.
- Data lineage.

As important as automated data control is, it is impossible without identifying and designating authoritative data sources and their data owners. Data owners manage the people, the processes, and the technologies that produce the data in systems of record. They are accountable for the quality of the data they contribute to the enterprise.

Addressing data defects at their point of origination serves to prevent the widespread proliferation of issues across the enterprise. Without data owners, downstream consumers tend to fix defects locally rather than getting them resolved at the source. And without the subject matter expertise of data owners, downstream data “fixes” may not conform to data requirements.

Since data quality issues tend to proliferate between systems, the identification and coordination of ownership responsibilities to critical data domains is necessary for ensuring the integrity of critical data along the supply chain.

3.3 Data quality and decision confidence

Executives rely on data to make high-stakes decisions. Automated controls and accountability for the management of data assets are necessary, but not enough to ensure high-quality data at scale. Though CDMC is a standalone framework, it assumes that an organization already has a strong foundation of data management capabilities. Sadly, many organizations have not sufficiently established that foundation.

The Data Management Capability Assessment Model (DCAM®) is a structured resource that defines and describes the capabilities needed to establish and sustain a successful data management initiative in any organization. The EDM Association created DCAM based on the practical experiences and hard-won lessons of many of the world’s leading organizations. The result is the synthesis of a broad range of data management best practices from across the full spectrum of interconnected business processes. The DCAM addresses the strategies, organization-wide structures, technology, and operational practices needed to drive data management successfully. It addresses the tenets of data management based on an understanding of business value combined with the reality of implementation.

Unraveling data silos through the creation of harmonized data is a prerequisite for eliminating

redundancy, reducing reconciliation, and automating business processes across the organization. Managing this kind of fully interconnected data is essential if we are to gain insight from analytics, feed our models with confidence, enhance our service to clients, and capitalize on new, but often fleeting, business opportunities. DCAM provides the guidance needed to assess the current-state of any organization's data management and define the objectives and framework for the target-state of the data management initiative.

3.4 Data security and confidentiality

Mitigating data security and confidentiality risks requires a broad defense-in-depth strategy. It is not just about building a taller wall; it is about ensuring that even if someone gets over the wall, they cannot find your sensitive data.

3.5 Data discovery and classification

You cannot protect what you do not know you have. The first step is identifying where your sensitive data lives and tagging it based on its sensitivity levels. Many organizations have a high-level view of where their data sits in applications but struggle to understand what is most critical in those systems and what should be protected.

3.6 Implement least privilege access

The Principle of Least Privilege (PoLP) ensures that users only have access to the specific data they need to perform their jobs.

- **Identity and access management (IAM)**: use robust systems to manage user permissions.
- **Multi-factor authentication (MFA)**: this is your strongest baseline defense. Even if a password is stolen, the second factor blocks the attacker.

3.7 Encryption: at rest and in transit

Harvesting encrypted data refers to the cybersecurity threat strategy known as "harvest now, decrypt later" (HNDL). In this scenario, malicious actors, typically nation-states or advanced persistent threat (APT) groups, intercept and store large volumes of currently encrypted data with the intention of decrypting it in the future when a sufficiently powerful quantum computer becomes available.

Encryption turns your data into unreadable ciphertext, making it useless to unauthorized parties. With the emergence of new technologies, such as AI and quantum computing, utilizing the highest encryption levels for your sensitive data becomes critically important.

3.8 Network security and monitoring

Layered network defenses prevent intruders from moving laterally through your systems.

- **Firewalls and micro-segmentation**: dividing the network into smaller, isolated zones.
- **Endpoint detection and response (EDR)**: monitoring laptops and servers for suspicious behavior in real-time.
- **Logging and auditing**: keeping a record of who accessed what and when.

4. Operational resilience and data availability

While data security is about keeping people out, operational resilience and availability are about keeping the lights on. It is ensuring that your systems and data remain accessible even when hardware fails, natural disasters strike, an insider steals data. or a ransomware attack attempts to lock you out.

4.1 Implement high availability and redundancy

Resilience starts with eliminating single points of failure (SPOFs). If one server goes down, another should take its place instantly.

- **Load balancing**: distributing traffic across multiple servers so no single unit is overwhelmed.
- **Failover clusters**: configuring secondary systems to automatically take over if the primary system fails.
- **Geographic redundancy**: hosting data in multiple physical regions to protect against localized outages.

4.2 Robust backup strategy (the 3-2-1 rule)

Backups are your "break glass in case of emergency" solution. The industry standard is the 3-2-1 rule:

- 3 copies of your data (1 primary and 2 backups).
- 2 different types of media (e.g., cloud and local disk).

- 1 copy stored off-site (or in an air-gapped, immutable cloud bucket)

4.3 Disaster recovery (DR) planning

Data DR is defined not just by the ability to restore files, but by the ability to resume operations after a catastrophic event. While a backup is a copy of your data, DR is the entire engine - infrastructure, personnel, and process - that puts that data back to work.

You must define two key metrics:

- **Recovery time objective (RTO)**: how quickly do you need to be back up?
- **Recovery point objective (RPO)**: how much data can you afford to lose?

4.4 Capacity planning and scalability

If your traffic spikes and your database cannot scale, you have a data availability problem within your systems.

- **Auto-scaling**: using cloud features to automatically add more compute power during peak loads.
- **Throttling and rate limiting**: protecting your core services by limiting how many requests a single user (or bot) can make.

4.5 Regular scenario testing

A resilience plan that has not been tested is just a wish list.

- **Tabletop exercises**: walking through a disaster scenario with all the key stakeholders including senior executives.
- **Drills**: purposefully shutting down non-critical infrastructure to see if the failover actually works as it was planned.
- **Patch management**: keeping systems updated to prevent vulnerabilities caused by known software bugs or weaknesses.

5. Third-party and data sharing risk

When you share data with a third party - whether it's a SaaS provider, a law firm, or a cloud vendor - you are essentially inheriting their security posture. If they fail, your data is just as compromised as if your own servers were breached.

Mitigating this risk is about extending your circle of trust without losing control of your data.

5.1 Due diligence and vendor assessment

Before signing a contract, you must vet the third party's security maturity. You should not just take their word for it based upon the vendor completing a form; look for standardized certifications.

- **SOC 2 Type II reports**: provides independent evidence of their security controls over a period of time.
- **ISO/IEC 27001**: an international standard for information security management.
- **Security questionnaires**: use tools to ask specific questions about their encryption and access policies.
- **Security tools**: utilize security tools to assess the external vulnerability of third-party systems.

5.2 Data minimization

The most effective way to reduce risk is to share less data with third parties. Only provide the absolute minimum amount of information required for the third party to perform their function.

- **Anonymization and pseudonymization**: strip away identifying details before the data leaves your environment.
- **Tokenization**: replace sensitive data with a non-sensitive "token" that has no value if stolen.
- **Data lifecycle**: when data is no longer needed for business purposes then ensure it is deleted from third-party systems.

5.3 Contractual protections

Your legal agreements (Master service agreements or Data processing agreements) should be your safety net. Ensure they include:

- **Right to audit**: the ability for you (or a third party) to inspect their security practices.
- **Breach notification requirements**: a strict timeline (e.g., 24-48 hours) for when they must notify you if they are compromised.
- **Liability clauses**: who pays for the cleanup, legal fees, and credit monitoring if the vendor loses your data?

5.4 Continuous monitoring

Security is a moving target. A vendor who was safe in January might be vulnerable by July due to a new software exploit or a change in process.

- **Vulnerability assessments**: use platforms to get real-time risk scores for a vendor's security posture.
- **Access reviews**: regularly check if the third-party's credentials to your systems are still active and necessary. If the project is over, revoke access immediately.

5.5 Managed exit strategies

What happens to your data when you stop using the vendor? You need a "right to be forgotten" in your business relationship.

- **Data return or destruction**: require a "certificate of destruction" stating that all your data has been purged from their primary systems and backups.
- **Avoid vendor lock-in**: ensure you can get your data back in a usable format so you are not forced to stay with a risky vendor just because moving is too hard.

6. Culture, incentives, and behavior

Data risk often has a cultural element. Peter Drucker, widely regarded as the "father of modern management", once said that culture eats strategy for breakfast. Generally, an organization's culture represents tacit knowledge - it determines what is taken as unchallenged givens. Culture represents the organization's personality, shaped by its people, leadership, history, and norms.

Since most organizations struggle to mature their data management, it is safe to say data has been taken for granted, abdicating responsibility to IT. Changing culture is changing habits - habits of thinking, and habits of how to get things done. In the context of risk, the change in culture starts with a shared understanding of risk and the role data plays in threatening operational efficiency and effectiveness.

The shift towards a more effective data risk culture is tied to a shared understanding of data quality and the business responsibility that is necessary for success. Since data quality surfaces most significantly at the level of interoperability, the change of perception from blaming to cooperation is challenging without a clear understanding of the data supply chain. To borrow from lean management, what makes collaboration effective is a shared understanding of the global (and not local) maximization of throughput. As mentioned earlier, this is practically impossible without data governance.

7. Conclusion

Data risk is not a technical issue to be delegated or a compliance exercise to be endured. It is a fundamental business risk that directly shapes decision quality, operational resilience, regulatory standing, and executive credibility. When data fails, the consequences cascade - turning information from a strategic asset into a liability that erodes trust, drives financial loss, disrupts operations, and damages reputation.

The root causes of data risk are rarely isolated system defects. More often, they emerge at the seams between systems, processes, and people, where interoperability breaks down, manual workarounds proliferate, ownership is unclear, and governance is

weak. These failures sit squarely within operational risk and demand leadership attention beyond IT.

With increasing pressure to optimize AI, we would be remiss were we not to consider AI risk. Increasingly, organizations are realizing that AI risk is tightly coupled to data risk, which boils down to data quality and control.

Generative AI (GenAI) is a black box, but opaque systems are not a new challenge. The rise in the availability and adoption of third-party applications has exacerbated the challenge of addressing interoperability. Many providers either lack adequate data documentation or refuse to share it, leaving the organization (and often the business users) to make up for the challenges of keeping the flow of data working properly. We have learned quickly that AI is not immune to the old adage: garbage in, garbage out.

That said, GenAI can still hallucinate on high quality unstructured data. That is not to say that unstructured data is categorically garbage; however, what research (and common sense) demonstrates is that the better the structure and description of the data you feed AI, the better the result. The new paradigm of AI optimization raises the bar for data management to promote a new mantra: gold in, gold out.

The gold standard for data management is knowledge graph. These are data structures that represent shared concepts and the rationale for their interrelationships. Until recently, knowledge graphs were AI, while neuro AI limped along making embarrassingly wrong and even awkward assertions. And as impressive as recent developments have been, GenAI does not perform reality testing like a human - it only knows what we tell it. Consequently, the integration of the two approaches to optimizing AI is quickly gaining traction.

Organizations serious in their pursuit of AI adoption are well on the way to embracing ontologies, knowledge graphs, and ultimately engineering data assets for optimal consumption and control into reusable data products. Data products can come in many forms, but data products derived from ontologies and structured as knowledge graphs ensure interoperability as a bi-product of better AI accuracy.

While we do not have the space to unpack these innovations in this paper, you will not struggle to find evidence of how fast the pace is picking up in their adoption.

But back to broader risk, which, for senior executives, the stakes are personal as well as organizational. Regulatory regimes increasingly formalize accountability for data integrity, availability, security, and use. Boards, regulators, investors, and customers expect executives to demonstrate not just control over data, but confidence in it. Decisions made on poor data are indistinguishable from poor leadership.

Reducing data risk, therefore, requires a broad strategic holistic approach across an organization: identifying and prioritizing critical data, embedding effective data governance, ensuring data quality, protecting data confidentiality, strengthening resilience and availability, managing third-party exposure, and shaping data culture. These are not isolated initiatives but require a range of data capabilities that must be owned, controls implemented, progress and risks measured, and sustained over time.

References

Bonnie, E., 2025, "110+ of the latest data breach statistics to know for 2026 and beyond," Secureframe, September 24, https://tinyurl.com/4yvz6kay

Edjlali, R., 2025, "Lack of AI-ready data puts AI projects at risk," Gartner, February 26, https://tinyurl.com/ycjn33st

Gartner, 2026, "Gartner research signals rise of Zero Trust data governance buy-in," January 27, https://tinyurl.com/mwddzdf5

IBM, 2025, "Cost of a data breach report 2025: the AI oversight gap," https://tinyurl.com/m3kbbs4n

Kyckr, 2026, "The data blind spot: uncovering a major cause of FCA AML fines," January 29, https://tinyurl.com/y6szbcua

Nasdaq, 2019, "81% of consumers would stop engaging with a brand online after a data breach, reports Ping Identity," October 22, https://tinyurl.com/yy98rypm

New Relic, 2026, "$1.8 million per hour: New Relic report details the staggering cost of high-impact IT outages for financial services companies," January 20, https://tinyurl.com/527n3bbu

Resolve, 2025, "5 statistics indicating API downtime's cost to finance operations," August 26, https://tinyurl.com/k4h8h7bu

Snaplogic, 2023, "The state of data management – the impact of data distrust," https://tinyurl.com/af89pte9

How data and AI shape financial services

Secure **data centric** decision making

Vijay Varadharajan,
Emeritus Professor, The University of Newcastle, and former Microsoft Chair Professor of Innovation, Macquarie University Sydney, Australia

ABSTRACT

In the current technology landscape, data is central to digital transformation as it drives decisions, enables personalization, and opens innovations. In this article, we explore how data centric approach is critical for secure data driven decision making. We present a data centric approach, where data will be a fundamental tenet in the specification of systems, services, and policies. We illustrate how such an approach provides superior security capabilities, by enabling the data owners to have better control over their personal data as the data moves over the Internet and is being used on distributed web platforms and network services. Systems that adopt a data centric approach are also more amenable to the enforcement of privacy regulations.

1. Introduction

Data has become the lifeblood of contemporary technology, fundamentally shaping how society functions and businesses operate. In fact, data has become central to business transformation and innovation. Not only are we sharing data with each other, but we are also sharing data with devices and applications. Those devices and applications, in turn, sharing data with each other. In today's interconnected world, the ability to collect, analyze, and act upon data has emerged as perhaps the single most valuable capability for an organization in its digital transformation.

Every interaction with digital systems, whether it is messaging or browsing on smartphones and laptops to financial and healthcare transactions and streaming sessions on smart TVs and devices, involves creation, transfer, and consumption of data. The amount of data shared in the digital ecosystem has become phenomenal in recent times, which has been spurred by the dramatic growth in social media. For instance, on YouTube alone some 5 billion videos are watched daily and over 500 hours of videos loaded every minute. When it comes to data, we can capture everything one has ever said from the time one is born till the time one dies in less than a few percent of a petabyte. Everything one has ever done and experienced can be captured in living color with only a few petabytes, and movies made to-date amount to a petabyte or so.

In such an ecosystem, data has become the essential ingredient that makes technology intelligent and valuable. Modern technological applications learn, adapt, and personalize experiences based on the data they collect. The dramatic growth of machine learning and artificial intelligence (AI) over the past decade would be impossible without massive amounts of data. These systems learn by identifying patterns across enormous datasets, improving their accuracy and capabilities as they process more information. The large language models that power conversational AI, the computer vision systems that enable autonomous vehicles, and the recommendation engines that personalize our digital experiences all depend on training with vast quantities of diverse data.

Before the growth in smart devices and social media, if one did something in public, some people saw them. They may or may not discuss what they saw with other people, and over a period, people may forget what they saw. Now in the digital world what is public far exceeds what Orwell had imagined in 1984. After all, Orwell only imagined that we would have government operated telescreens in our houses looking at everything we did. He did not imagine that everybody in the population was a telescreen. Anyone can now use a smartphone to record what is of interest to them and upload it onto the Internet. Then it becomes available pretty much all over the world. It may also not be forgotten and can be recalled or aggregated.

Similarly, previously, the default position was data was private until one explicitly opened it. Now in the digital world, often private information is wide open. As soon as a laptop or a smartphone is connected to any form of network, one's data may not be private. That is, the presumption has changed from one of controlling the opening to controlling the closing.

2. Data centric paradigm

At the center of cyberspace lies data, which has become the universal currency in the digital economy. Yet, if one were to enter a query on a computer or a smartphone and ask something about a specific data item on the Internet, such as "who has my credit card number", the networked system will not be able to come up with an answer. Though one talks about data becoming the universal currency in this digital world, one cannot make a data specific query to the networked system (such as the Internet). What is required to address this issue is a paradigm that has a data centric focus. In such a data centric paradigm, data will be a fundamental tenet in the specification of systems, services, and policies. Such a data centric approach will also enable the data owners to better control what rights others should have on their personal data as the data moves over the Internet.

However, just because data is accessible, it does not mean it is trustworthy or reliable enough to make decisions upon, or even ethical to access and use it.

The data centric paradigm poses several issues, such as:

- Where does the data come from (data provenance)?
- How trustworthy is the data?
- How would you know where your data is?

- Do you know who can see the data and modify it without a trace?
- Who can aggregate, summarize, or embed your data for purposes other than what data owners had specified?
- How can data owners specify and enforce policies on the data as it moves over the Internet?
- How would you detect and prevent attacks on the services that operate on the data in a distributed environment?

All these issues impact on the quality of decision making and trust in the decisions being made by applications and users. For instance, machine learning algorithms are vulnerable to injection of malicious or biased data at training time (poisoning attacks) and testing time (evasion attacks).

When it comes to personal data, users typically want personal control of their data even if they do not want to exercise this control. Users are happy to allow agents (such as applications and devices) that they trust to access and process their data. On the other hand, organizations usually prefer consistent policies and agreed data related rules for compliance, which would enable them to build customer relationships. Regulators are more concerned with the rights of individuals and organizations related to the control of data.

Recently, regulations such as the E.U.'s General Data Protection Regulation (GDPR) specify controls on the collection, storage, use, and processing of personal data. GDPR specifies controls that an organization can carry out on the personal data of individuals resident in the E.U., where that processing relates to the offering of goods or services to those individuals or the monitoring of their behavior. In this context, personal data refers to any data that relates to an identified or identifiable individual. The data can range from online identifiers (such as IP addresses) to employee information, to customer services data, location data, biometric data, to health and financial information. In fact, it can even include information that does not "appear" to be personal such as a photo of a landscape without people, where that information is linked by an account number or unique code to an identifiable individual. Even pseudonymized data can be personal data if the pseudonym can be linked to a particular individual.

Systems that adopt a data centric approach are more amenable to the enforcement of regulations such as GDPR, as in these systems data becomes a fundamental tenet in the specification of services and policies. This in turn helps achieve principles of individuals' rights specified in the GDPR such as the right to know how one's data is processed, the right to see the data that is being stored about an individual, and the right to have one's data erased under certain conditions.

3. Data centric approach

Conceptually, a data centric approach requires certain fundamental mechanisms and infrastructure authorities that are not ordinarily present in current networks.

- At the basic level, there is a need for a fundamental mechanism that securely couples a data item to its policy. Typically, a policy will be captured in the form of meta-attributes and rules. There must be a security mechanism enforcing this coupling, which should be logically inseparable. That is, policy stays with the data when the data moves from place to place and when it is copied and stored.
- Next, we need data processing agents that check whether the policies are satisfied before and after data usage. Furthermore, these data processing agents must be trusted. For instance, are they trusted to enforce the policies properly at the right time and not to change the policies? There can be a range of policies such as ***pre-requisite policies***, which should be satisfied before the data can be used, ***obligation policies***, which should be undertaken following data usage (such as certain system updates), ***auditing policies***, which are related to usage and tracking of data, provenance policies, such as whether the data is coming from the correct source, and policies that check whether the data has been used for the right purpose, etc.
- Additionally, there is a need for trusted infrastructure authorities that enable mapping of policies to data, as well as from data to policies. In some sense, these authorities can be thought of as analogous authorities such as the Domain Naming System (DNS) on the current Internet, which provides the mapping between the URL and IP addresses.

1. https://tinyurl.com/5daxm68b

4. Data centric security

In terms of security, the data centric approach recognizes that data is the true asset that needs protection and operates on the principle that data should carry its own protection mechanisms. Hence, this represents a fundamental shift from traditional perimeter-based security models to an approach that focuses protection directly on the data itself, regardless of whether it is at rest or where it resides and how it moves through networked systems and used in applications.

In today's distributed environment, data-centric security can provide consistent protection across hybrid and multi-cloud environments. Traditional security models often break down when data moves between different systems, cloud providers, or organizational boundaries. Data-centric approaches maintain protection regardless of the underlying infrastructure.

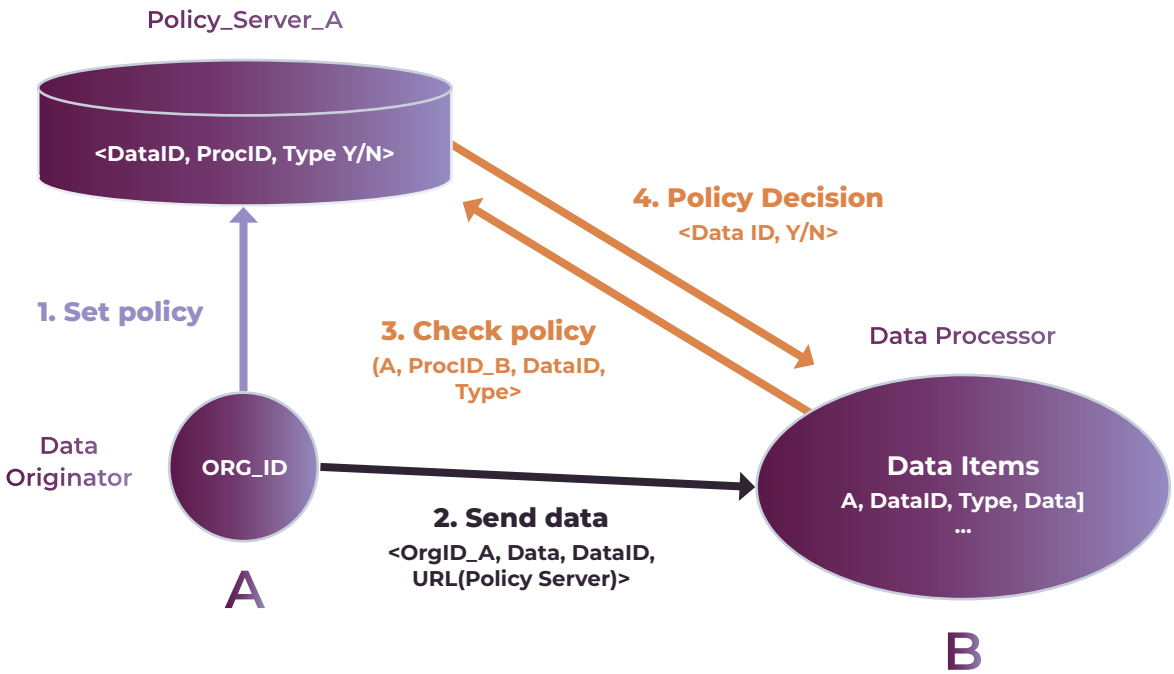
We will now examine how data centric approach enables superior security protection by considering different scenarios in the technology scenery.

4.1 Simple data centric communication scenario

Let us now start with a simple data centric communication scenario, where digital data items move from one place to another in a network. As alluded to in the credit card example above, the data owners wish to know who can have a lawful possession of their specific data item. Not only should they know who holds it, they also need to know whether these have been obtained as per the data owners' policies. Furthermore, they may wish to enforce subsequent transfer of their data to other parties as well as their deletion. These are some of the controls that are required in regulations such as the GDPR.

First, let us consider a simple case where an entity A who is the data owner of a specific data item *d*, transfers *d* to another entity B over the Internet (Figure 1). With the data centric approach, the data *d* has a policy *pol_d* coupled with it. Hence the message *{d, pol_d}* gets transferred to the entity B. Just for illustration, let us assume that the *pol_d* is a simple URL which is the address of the Policy_Server_A associated with the entity A. The Policy_Server_A holds the policies and the associated rules specified by the entity A describing who can use what data items belonging to A for what purpose and how. When B receives the message *{d, pol_d}* from A, B is trusted to contact the Policy_Server_A using the URL given in the *pol_d* before using the data *d*. This message from B to the Policy_Server_A says that B wishes to use the data *d* obtained from A. The Policy_Server_A can then check whether B is allowed to use the data *d* and how. If so, B's request is approved, and A passes the approval to B. Then B can use *d* for its purpose.

Figure 1: Simple data centric communication scenario



Even this simple protocol for the data transfer scenario illustrates several principles of the data centric approach. First, the need for secure coupling between the data *d* and the *pol_d*; if this were to be decoupled, then it will enable B to behave maliciously, by not following the rules associated with *d* specified by A in the Policy_Server_A. Also, B needs to be trusted to contact the Policy_Server_A before using the data *d*. If B is not trusted, it may bypass Policy_Server_A and use the data *d* directly. Also, B is trusted to use the data *d* only in the manner allowed by the policies specified in the Policy_Server_A and not in any other way. For instance, if the Policy_Server_A were not to allow B to make further transfer of the data *d* to another entity, say C, then B is trusted not to do so.

As such, there is no security built-in in this protocol. We can enhance this protocol with security features such as encrypting the data *d* with a key (e.g., a symmetric key) and then have this symmetric key encrypted under a public key, and then passing the encrypted symmetric key along with the encrypted data to the entity B. B can then contact the Policy_Server_A as before together with its public key certificate. Note that the public key certificate holds the verified public key of entity B. The Policy_Server_A can then send B the symmetric key (used to encrypt the data) encrypted with the public key of B. B can then use its private key to decrypt and obtain the symmetric key and then use this symmetric key to decrypt the data *d*. To counteract attacks, such as replay and masquerading, we would need to incorporate nonces and signatures thereby further enhancing the security of this protocol.

The above simple scenario can easily be extended to include multiple entities whereby A is sending data to B and then B to C, and C to D, etc. Each time the transfer occurs, the protocol like the above is followed, the policy server checking whether the transfer is allowed or not. Each entity has its own policy server containing the rules as to how its data should be processed by other entities. So, when C transfers to D, the Policy_Server_C determines whether this is allowed or not.

A typical privacy related example could involve rules associated with the data when it is being passed between various system entities. Each entity may have associated set of rules and conditions as defined by the privacy act. The rules and conditions specify the allowed sources from which data may be collected, the data categories that may be collected, and the purposes for which data that is collected may be used and the permissible uses. It also specifies a set of possible recipient entities, categories of data that may be transferred to these entities, and the set of authorized purposes for which the specified recipient entity may use the given data.

In practice, when an entity receives some data *d* from another entity, it often processes this data (say changes *d* to *d'*) before sending it to other entities. Naturally, now the question arises as to what policies are associated with the modified data *d'*. In fact, in real world applications, an entity may combine data from multiple entities to produce new data. Although the details may become complex as data is processed and passed from one entity to another, one can formulate policy language and expressions to specify the ways that the use of data may be restricted and the way in which the restrictions are transformed. For instance, such policy expressions can describe how the restrictions on the use of a particular data item may be computed from the history of its transmission: the encumbrances that are added or deleted at each step. This can be used to specify data provenance constraints and data dependent constraints as well as composition policies.

4.2 Distributed web platforms

The above data centric scenario is also relevant to distributed web platforms (such as Facebook), which commonly use third-party applications to process users' data. In such platforms, the functionality available to users is no longer the product of a single entity, but a combination of a potentially trustworthy platform running code provided by less-trusted third parties. This is the case for instance with platforms like Facebook.

As mentioned above, many applications are only useful if they can process sensitive user data such as financial or medical data. But once access to this data has been granted, there is no holistic mechanism to constrain what the application may do with it. This is where the data centric approach becomes significant, as it allows policies to follow data throughout the systems and enable these policies to be enforced *even after* the applications are given access to data. That is, data centric policies control what applications *can do* with the data. This helps to enforce privacy policies on user data preventing data breaches, while allowing untrusted applications in the web platform to deliver extended features to users.

Consider a scenario where personal medical data gathered by sensors, say from monitoring a patient at home, are used by cloud services and databases. The data centric approach allows meta-attributes to be associated with the patient data, specifying it is medical-research data. This can then in turn be used to ensure that that this data flows only to those entities conducting medical research, satisfying the policies associated with these meta-attributes. These meta-attributes can be used to satisfy requirements such as medical data must only be stored in encrypted form and can be used for research purposes only if the consent of the owner is obtained, and the data is anonymized.

5. Data centric networks

The aim of networks is to provide content of interest (such as text, images, audio, and videos) to the users that need it. However, the communication model is based on connections between hosts (devices) rather than the actual data. There is often this conflation of what content you want to access with where (on what host) that content resides, and this determines the context in which they offer network security. Security of data is tied to the host where it is stored and how (over what kind of connection) it is obtained.

The data centric approach decouples the *what* from the *where*, and secures the data itself, rather than the *containers* where it resides. This enables data-based security, rather than connection-based security, allowing users to securely retrieve *desired content* by *name* and *authenticate* the *result regardless of where it comes from*. If every data is uniquely named, a data packet is meaningful independent of where it comes from or where it may be forwarded to. Such data centric networks play a crucial role in the building of the future Internet. With content as its pivotal point, such networks can better deal with the bandwidth insufficiency in the traditional IP-based network architecture. With the Internet traffic being increasingly characterized by videos and images, the current IP-based networks have become inadequate for such data content and user intensive services.

In data centric networks, each data (content) is represented using unique content name. Such data centric networks [referred to as Named Data Networks (NDN)][2] require different routing protocols, from the traditional IP routing protocols, for sending and receiving data. To receive data, first a consumer sends out a packet that carries the name of the required data (this is referred to as the *interest* packet). The NDN router remembers the interface this request comes in and then forwards the *interest* packet by looking up the name in its *Forwarding Information Base (FIB)*, using a name-based routing protocol. Once the *interest* reaches a node that has the data, the *data* packet is sent back, which carries both the name and the content of the data. *Data* packet follows in reverse the path taken by the *interest* packet to get to the consumer. Furthermore, note as a NDN data packet is independent of where it comes from or where it may be forwarded to, the router can cache the data to satisfy future requests.

Hence, while forwarding the data packet, NDN routers keep a copy of the data packet in its cache, so that if any further interest packets arrive (either from the same consumer or from other consumers) for the same content, the routers can satisfy the requests by sending the data packets without contacting the actual Content Provider. This enables NDN to utilize efficiently network resources such as network bandwidth, routers' cache space, and lower latency content delivery to consumers.

Data is protected in such data centric accounts directly by securing each data packet. Named secured data packets form the basic building block in data centric networks such as the NDN. Data packets in NDN need to be immutable. Each data packet carries a cryptographic signature (of the data producer) binding its name to the data content at the time of creation. If required, data packets can also be encrypted. Hence security mechanisms can be provided which will enable the data receiver to determine whether it is the desired data, whether it is from the intended source, it is not changed in an unauthorized manner in transit and not disclosed to unauthorized entities.

As by default all data signed in such networks, including routing messages, this reduces spoofing attacks and allows malicious tampering to be easily detected. Note as these packets are about data and not addressed to hosts, it is more difficult to send malicious packets to a particular target.

2. https://tinyurl.com/bdbetj8y

6. Data centric cloud storage

With the rapid increase in the growth of digital data that needs to be stored, in recent times we have witnessed a tremendous growth in the storage of data in the cloud, which provides several benefits such as on-demand access and scalability. Cloud data storage raises important security issues of how to control and prevent unauthorized access to data stored in the public cloud. We have several technologies that offer access control techniques for preventing unauthorized access and use of data stored in the cloud. Once again, these techniques control the access to the containers where the data is stored. The data centric approach provides direct control over the data being stored in the cloud.

With the data centric approach, the data owners store data in the cloud in encrypted form in such a way that only authorized users can decrypt and access the data content. The authorized users are those users that satisfy the access policies specified by the data owner. At no time the data remains in plain format in the cloud and hence not even the cloud provider who stores the encrypted data has got access to the data in plain form (unless the cloud provider has been explicitly authorized by the access policies specified by the data owner). Hence such a data centric approach offers the opportunity to integrate encryption techniques for protecting data with access control for controlling who can access data, thereby enabling *enforcement of user centric security policies on encrypted data stored in the public cloud*. Zhou et al. (2013) provides a detailed description of such a data centric security architecture for cloud storage and its implementation.

Such a data centric approach is applicable in a range of business sectors such as secure storage of sensitive personal customer data in public cloud by banks. The customer data will be stored in encrypted form. The customer can specify policies that determine who is allowed to access his or her personal data. Only those users who satisfy the policies specified by the customer will be able to access the data. If the cloud provider does not satisfy these policies, then the cloud provider will not be able to access the customer data, thereby reducing the level of trust on the cloud provider. Furthermore, one can extend this approach to enable secure data sharing among several organizations such as in collaborative projects. For instance, consider sharing data between financial institutions in a consortium, where the ability to access and use the data from one financial institution by others is only allowed if the appropriate policies specified by that institution is satisfied by the others [Sultan et al. (2023b)].

Now let us combine this data centric cloud data storage with that previously outlined for named data networks (NDN). Consider a content provider (such as Netflix) streaming content over a network offering different types of content such as movies, TV shows, and music to consumers. The content such as the movies will be encrypted and stored in the cloud by the content provider, as per the data centric approach for cloud data storage. Only authorized consumers who have paid the subscription for specific content (e.g., a movie) will be able to access that specific content. Now, a consumer can request specific content using the named content, e.g., the movie name, over NDN. Data centric security mechanisms applied to the movie content will ensure that the legitimate consumer (who requested the movie) receives the desired movie and no one else. Once a movie is downloaded from the content provider cloud, the content can be cached in NDN routers, thereby enabling the provider to service further requests of the consumers from the nearby NDN routers without contacting the content provider. Sultan et al. (2023a) provides a detailed description of such a data centric security architecture for cloud storage and its implementation.

7. Conclusion

The shift in the center of gravity with the data centric approach, from applications infrastructure and processes to data as the primary asset, is becoming critical. Understanding data's central role is essential for any organization seeking to comprehend where technology is heading in the future. The role of data in business is likely to further deepen with the expansion of the Internet of Things and advances in AI and quantum computing creating entirely new approaches to data processing. In such an environment, data centric thinking will help organizations better understand the trustworthiness of their data and their reusability and the value it can create.

Furthermore, data-centric security is useful in addressing public concerns about data privacy and protection by making data security transparent and accountable. When organizations can demonstrate that sensitive personal information is encrypted

and protected at the data level, it provides concrete evidence of their commitment to privacy rather than relying on promises about secure systems. The granular controls and audit trails that data-centric security systems allow enable organizations to show regulators and the public exactly how data is being accessed and used. This level of accountability helps rebuild trust that has been eroded by high-profile data breaches and privacy violations.

Additionally, this approach enables organizations to share data more safely for legitimate purposes like research, analytics, or business partnerships while maintaining individual privacy. Moreover, data-centric security ensures that only the specific data elements needed for a particular purpose are accessible, reducing the risk of over-collection or misuse of personal information. This alignment with privacy-by-design principles demonstrates a proactive commitment to protecting individual rights.

The innovative aspect of data centric approach lies not just in the technology, but in how it enables organizations to transform their relationship with data from one of custody and control to one of stewardship and protection, ultimately fostering greater public confidence in how their sensitive information is handled.

References

Sultan, N., V. Varadharajan, S. Camtepe, S. Nepal, 2023a, "An accountable access control scheme for hierarchical content in named data networks with revocation," Proceedings of the 25th European Symposium on Research in Computer Security, ESORICS 2020, September, 569–590

Sultan, N., V. Varadharajan, L. Zhou, and F. Barbhuiya, 2023b, "A role-based encryption (RBE) scheme for securing Outsourced Cloud Data in a Multi-Organization Context", IEEE Transactions on Services Computing 16:3, 1647-1661

Zhou, L., V. M. Varadharajan, and M. Hitchens, 2013, "Achieving secure role-based access control on encrypted data in cloud storage", IEEE Transactions on Information Forensics and Security 8:12, 1947-1960

How data and AI shape financial services

Beyond the turning point: **re-forging financial institutions** for the age of agentic AI

Bill Oates,
Cloud and AI Capability Lead, Projective Group

ABSTRACT

As of 2026, the global financial services industry (FSI) remains entangled in a re-surfaced Solow Paradox, where unprecedented investment in generative AI (GenAI) infrastructure has yet to yield significant increases in Total Factor Productivity (TFP). This article investigates the measurement gap by synthesizing Carlota Perez's theory of technological revolutions with Sangeet Paul Choudary's "Reshuffle" framework. Drawing historical parallels to the transition from group drive to unit drive systems during the electrification of industry supports the idea that AI's economic impact is currently experiencing an adoption J-curve and suffering socio-institutional lag. By examining three application areas for AI in financial services, it is possible to see this lag directly and position the situation in 2026 as a turning point.

1. Introduction

As of 2026, the global economic narrative is dominated by the tension between the AI race and the productivity puzzle. In 1987, economist Robert Solow quipped that "the computer age was visible everywhere except in the productivity statistics" [Brynjolfsson (1993)]; nearly four decades later, this paradox has resurfaced with GenAI and large language models (LLMs) [Henry et al. (2024), Oxford Economics (2026)]. Despite unprecedented investment in AI infrastructure, with data centers consuming over 415 TWh globally in 2024 and projected to double by 2030, the immediate impact on total factor productivity (TFP) growth remains muted [WEF (2025)]

In this article, I focus on three critical financial services domains - risk management, retail banking, and investment banking - to determine their automation maturity. I suggest that while the former remains largely in a phase of incremental automation, both retail and investment banking exhibit signs of a genuine reshuffle, characterized by the unbundling of expertise, the implementation of AI agents at critical control points, and the commoditization of knowledge retrieval.

This article concludes that productivity gains of the second half of the 2020s depend on a systemic re-forging of institutional hierarchies. To survive the "Schumpeterian gale", firms must move beyond task-centric tools toward AI-driven engines that prioritize human curiosity, curation, and high-stakes judgment.

2. Divergent economic projections and the 2026 reality

The debate surrounding AI's contribution to growth is split between institutional optimism and academic caution. Organizations like Goldman Sachs have predicted a U.S.$7 trillion increase in global GDP and a 1.5% annual boost to U.S. productivity [GS (2025a)], while the McKinsey Global Institute have historically estimated even higher gains [MGI (2023)]. Conversely, researchers utilizing task-based frameworks, such as the Nobel Laureate, Daron Acemoglu, suggest a more restrained outlook, with productivity increases potentially as low as 0.07% annually over the next decade [Acemoglu (2024)]

By 2026, the reality aligns more closely with the more conservative estimates. In January 2026, a widely cited report from Oxford Economics stated that the "evidence of an AI-driven shakeup of job markets is patchy", leading them to conclude that there was no adjustment needed to their longer-term economic forecasts [Oxford Economics (2026)]. The Penn Wharton Budget Model indicates that while 40% of GDP is susceptible to AI's influence, the peak annual contribution to productivity growth may only reach 0.2 percentage points by 2032 [Arnon (2025)]

2.1 Implementation friction

The disconnect between technology and statistics is often explained as being a symptom of the "adoption J-curve" [Brynjolfsson (2021)]. Firms often experience an initial productivity loss following AI implementation due to the costs of redesigning workflows, retraining staff, and building data infrastructure [WEF (2025)]. This same phenomenon

Table 1: Predictions of AI's impact on productivity

Forecaster	Annual productivity growth forecast	Estimated total GDP increase (10 years)
Goldman Sachs	1.5%	$7.0 Trillion
McKinsey Global Institute	1.5% - 3.4%	N/A (Broad GDP Acceleration)
Daron Acemoglu	0.07%	0.9% - 1.8%
Wharton PWBM (2025/26)	0.2% (Peak in 2032)	1.5% by 2035; 3.7% by 2075

was observed back in the 2000s in relation to the adoption of new internet-based technologies, where it was recognized that a phase of co-invention of complementary processes and business models was necessary to achieve productivity gains [Bresnahan (2002)].

In 2026, many sectors are still navigating the trough of this curve. For instance, while 26.4% of workers utilized GenAI in late 2024, only a small fraction of organizations reported measurable returns by early 2026, with some workers reporting that AI actually creates extra work through the need for oversight [WEF (2025), Arnon (2025)]. A widely reported 2025 study from ex-MIT researchers also claimed a stark GenAI divide [Challapally et al. (2025)], observing that "just 5% of integrated AI pilots are extracting millions in value, while the vast majority remain stuck with no measurable P&L impact." The report, and the cascade of advice that followed, provided a multiplicity of remedies to how organizations could break out of "AI POC Hell".

2.2 More than adoption pains? Innovation cycles and socio-institutional lag

In the 2010s, the then Governor of the Bank of England, Mark Carney, and his Chief Economist Andy Haldane were prominent amongst those laying the foundations for thinking about AI as a "General Purpose Technology" (GPT) [Haldane (2015), Carney (2018)]. The characteristics of a GPT are classically that it is pervasive, is available at little or no marginal cost, and is a catalyst for wider innovations or spill-over effects.

Ironically, the label "GPT" is already indelibly attached to AI, thanks to OpenAI's breakthrough ChatGPT product named for its application of Generative Pre-Trainer Transformer methodology. If, however, we adhere to the earlier, General Purpose Technology definition, we can access a theoretical framework that helps us understand why Solow's Paradox is so pervasive. Specifically, we can look to the Neo-Schumpeterian economic framework put forward by Carlota Perez. In her work, Perez, posits that technological revolutions occur in distinct cycles, identified as half a lifespan of around half a century, following a predictable structure of "installation" and "deployment".

2.3 The structure of technological revolutions

According to Perez (2002), a technological revolution is a cluster of innovations that cause an upheaval in the economy and society. These revolutions undergo four phases:

1. **Irruption**: the initial appearance of the new technology.
2. **Frenzy**: a period of financial speculation and irrational exuberance, where "financial capital" leads the way.
3. **Synergy**: a "golden age" of deployment where the technology becomes the standard for the whole economy.
4. **Maturity**: diminishing returns and the search for the next big bang.

In 2026, we are arguably at the "turning point" between "frenzy" and "synergy":

- ➔ **The spending peak**: the surge to over U.S.$500 billion in infrastructure [GS (2025b), BG (2025), UBS (2025)] suggests that the "installation" is physically complete, a necessary precursor to the "turning point".
- ➔ **The quality crisis**: the widespread adoption of the term "slop" [Fisher Phillips (2026), Stewart (2026)] and the risk of "model collapse" [Gomstyn and Jonker (2025)] signify that the unrefined "frenzy" mode of production has hit a point of diminishing returns.
- ➔ **The synergy pivot**: the emergence of agentic AI and productivity beneficiaries [GS (2025b), UBS (2025)] signals the transition toward the "synergy" stage, where the focus shifts from building the technology to extracting real-world value.

The transition to synergy may, however, be hindered by institutional inertia and could carry a high change tariff – an episode of "creative destruction". Perez (2009) theorizes that each revolution brings a new "techno-economic paradigm" (TEP) - a set of best-practice principles that become the new common sense for management and organization. The Perezian framework emphasizes that productivity growth only resumes once there is a re-forging of the link between institutions and technology. Organizations that successfully undergo creative destruction or make the necessary paradigm shift at the institutional level - destroying old bureaucratic hierarchies to make room for AI-driven ecosystems - will lead the next "golden age".

3. Re-forging for AI: transformational versus incremental impact

To better understand what re-forging could mean, we can explore previous GPTs. What we see is that the impact of a GPT is felt not in the direct automation of individual tasks but in the system-wide transformations it enables.

3.1 The electrification of industry

David (1989) explored the Solow productivity paradox by drawing parallels between the impact of the dynamo (electric motor) on manufacturing with the contemporary impact (or lack of, at that time) of computers. Between 1870 and 1893, whilst electricity was widely available and indeed commoditized, the dynamo remained in an "inchoate phase", where it was merely used to replace steam engines in existing factories. These factories used a "group drive" system: one massive motor turned a central shaft, which was connected to machines by a tangle of belts and pulleys.

Productivity stagnated because factory owners treated the electric motor as a bolt-on tool rather than a transformational engine. It was only when manufacturers transitioned to the unit drive system, where each machine had its own motor, that the true power of electricity was realized. This allowed for:

- **Reconfigured factory layouts**: machines could be arranged efficiently in the order of the production process rather than their proximity to the central shaft, as exemplified by Henry Ford's moving assembly line
- **Improved working conditions**: the removal of overhead belts reduced accidents and dust, and paved the way for a new approach to industrial architecture, including art deco factory buildings that are still celebrated today.
- **Horizontal construction**: moving from multi-story vertical stacks to single-story horizontal plants, significantly lowered capital-to-output ratios over the long term.

Although it took several decades, the real impact was not the direct application of the new tool, but the wider re-configuration of the entire system of production.

3.2 AI as the new unit drive: from tool to engine

In 2026, what we see in AI can be characterized as moving from the group drive phase, where AI is treated as a generic automation tool to speed up existing tasks such as coding or writing, to the unit drive phase, where AI capabilities are embedded into new working approaches and hence restructure entire business processes. Choudary (2025) sets out these arguments in his book. Critically, he suggests that statements such as "AI won't take your job, but someone using AI will" is dangerously incomplete. He suggests that a more accurate view is that systems of work are being reconfigured so that certain roles are no longer necessary, similar in a way to how the unit drive system eventually eliminated the need for specialized belt-men in factories.

The "reshuffle" framework further provides a blueprint for the economic transformation occurring in 2026. Choudary (2025) argues that AI's primary function is not just thinking but coordination, acting as the "glue" that holds economic systems together.

Table 2: Transformations then and now: manufacturing industry versus financial services

Feature	Group drive (incremental/tool)	Unit drive (transformational/engine)
Philosophy	"Bolt-on" to old processes	Reimagined from the ground up.
Organization	Centralized, rigid hierarchies	Decentralized, modular workflows
Bottleneck	The speed of the individual worker	Coordination and risk-taking
Value Focus	Execution and output volume	Curiosity, curation, and judgment

3.3 The value chain shift: curiosity, curation, and judgment

In the pre-AI knowledge economy, value was centered on knowledge transmission, storage, and retrieval. In 2026, GenAI has "hacked" or commoditized this middle section of the value chain. When application-oriented knowledge is abundant and virtually free, value shifts elsewhere in the chain.

1. **Curiosity (the front end)**: in a world of inexpensive answers, the premium moves to asking the right questions. Having "intentional curiosity" serves as a forcing function to compress the solution space and identify the right variables.
2. **Curation (the back end)**: as AI generates vast amounts of expressive but sometimes irrelevant content (the AI slop discussed earlier), the ability to filter, exclude, and contextualize becomes vital.
3. **Judgment (the final link)**: the ultimate value sits with judgment - the ability to isolate a path of action and assume the risk and consequences of that choice.

4. Financial services in 2026: the early stages of a wider re-shuffle

The financial services industry may potentially be considered to be the canary in the coal mine for the wider AI reshuffle. Banking and finance have always been information-intensive industries, making them the most susceptible to a technology that commoditizes knowledge processing. Using the lens of Choudary's Reshuffle paradigm, we explore here three areas in the financial services where AI is a catalyst for the unbundling of existing roles and workflows, changing the basis of competition, and shifting control points – the positions in a value chain where power and value accumulate:

4.1 Compliance, risk management, and fraud detection functions

The impact: AI is fundamentally transforming the speed and accuracy of risk assessment. Financial institutions are moving from static, rule-based systems to dynamic, real-time predictive modelling:

- Operational efficiency: AI allows for the processing of vast datasets to identify compliance deviations and fraud patterns that human analysts miss. It is a winner amongst AI applications in 2025, with 56% of respondents in an MIT/EY survey identifying it as already "highly capable" [MIT/EY (2025)].
- Preventative capabilities: systems are shifting from reactive recovery to proactive prevention. For example, Visa reportedly thwarted U.S.$40 billion in fraudulent transactions in 2023 using AI [Reuters (2024), KPMG (2025)].
- Regulatory compliance: AI is being used to navigate complex regulatory environments, with tools capable of comparing regulatory changes against internal policies instantly, a task that previously took significant manual effort [KPMG (2025)].

Assessment: incremental automation or genuine reshuffle?

Verdict: primarily incremental automation (with emerging Reshuffle characteristics).

Evidence for incrementalism: much of the current application of AI in risk management fits Choudary's definition of a task-centric view, where AI is used to optimize the part rather than re-ar-

chitect the system. The fundamental workflow of a bank detecting bad actors remains structurally similar; AI simply acts as a sharper blade cutting through data faster. The focus is on efficiency and cost reduction (e.g., reducing false positives) rather than redefining the nature of risk itself.

Evidence for reshuffle: there are signs of a reshuffle regarding coordination without consensus. AI allows disparate data sources (unstructured documents, transaction histories, and external market data) to be synthesized into a unified representation of risk without requiring standardized formats across all inputs. Furthermore, as AI absorbs the tacit knowledge of compliance officers, the economic value of routine compliance roles decreases, shifting value to those who manage the AI systems (the "above-the-algorithm" employees). It is easy to conceive how these systems can orchestrate fraud detection on a perpetual basis, reconfiguring the detection rates and making it more immediate, even potentially creating tooling that moves towards prevention in the first place at the edge or transaction initiation.

4.2 The interaction layer in retail banking: the shift to AI-first engagement

The impact: the sector is moving from a branch-and-app model to an AI-first model driven by autonomous agents:

- **Hyper-personalization**: banks are deploying AI agents that act as 24/7 virtual financial advisors, capable of autonomous transaction management and hyper-personalized engagement [Sachse et al. (2025)].
- **Invisible interface**s: the interaction model is shifting toward invisible, embedded interfaces where payments and savings happen seamlessly in the background of a customer's life, orchestrated by AI.
- **Consumer sentiment**: while 68% of banking leaders are strongly in favor of this shift, consumer trust lags, with 50% of U.K. consumers reporting anxiety regarding GenAI [FIS (2025)]

Assessment: incremental automation or genuine reshuffle?

Verdict: genuine reshuffle

Evidence for reshuffle: this area strongly fits Choudary's description of shifting control points. Traditionally, banks held the control point through their branches and apps (relationship advantage). AI agents act as a wedge to capture a new control point based on "intelligence advantage". If an AI assistant (potentially owned by a tech giant rather than the bank) becomes the primary interface guiding a customer's financial decisions, the bank risks being commoditized into backend infrastructure.

Unbundling roles: the role of the human financial advisor is being unbundled. Tasks requiring data retrieval and basic advice are moving to AI, collapsing the economic value of human generalists. The human role is forced to rebundle around high-stakes judgment and empathy - managing risk-based constraints where the AI cannot be fully trusted.

Systemic change: this is not just automation; it is a structural change in how value is delivered. The transition to "outcomes-as-a-service" (e.g., "help me save for a house" rather than "open a savings account") represents a fundamental change in the economic logic of the industry.

4.3 Execution in Investment Banking

The Impact: AI is automating the highly skilled, knowledge-intensive workflows of deal-making and capital markets:

- **Deal sourcing and diligence**: AI tools are automating the analysis of earnings releases and macroeconomic data, increasing deal flow by up to 40% [McKinsey (2025)].
- **Generative creation**: companies like JP Morgan are using AI for tasks such as drafting pitch books, summarizing legal documents, and generating term sheets.
- **Algorithmic trading**: more than 60% of global transactions are expected to be algorithmic by 2025, moving execution speeds beyond human capability [McKinsey (2025)].

Assessment: incremental automation or genuine reshuffle?

Verdict: early-stage reshuffle (unbundling of expertise)

Evidence for reshuffle: this sector exhibits Choudary's concept of the "unbundling of expertise". Historically, investment banking relied on a pyramid structure where junior analysts performed the "grunt work" of data synthesis and pitch deck creation. AI turns this knowledge work into a commoditized building block, which forces a "re-bundling" of the organization. If AI generates the pitch deck, the value of the junior banker collapses, and

firms may need to restructure away from the traditional pyramid model.

Shift in competitive advantage: the "tool versus engine" tension is visible here. If banks rely on third-party AI tools (e.g. standard LLMs) for deal analysis, they lose differentiation. To survive the reshuffle, banks like Goldman Sachs and J.P. Morgan are attempting to build their own engines by training proprietary models on decades of internal deal data, thereby treating AI as a source of competitive advantage rather than a utility. This mirrors Choudary's argument that "tools can be bolted on; engines require integration" to maintain advantage.

4.4. Scaling versus agility: the decline of the "stable firm"

In retail banking, the control point is shifting from the physical bank (or call center) to the AI agent. In investment banking, the structure of the firm is being unbundled as expertise becomes commoditized. Risk management remains largely in the optimization phase but provides the data infrastructure necessary for future systemic coordination.

The winners in this reshuffle will likely be those who move from "task-centric" thinking (replacing analysts with bots) to "system-centric" thinking - redesigning their business models to manage the new constraints of risk and coordination that AI introduces. Additionally, one of the most profound insights from the 2026 reshuffle is that size, which was once an impenetrable competitive moat for banks, has become a liability. Large, legacy firms are burdened by massive sunk costs in their existing power transmission apparatus - their IT stacks and middle-management layers - much like the multi-story factories of the steam era.

Smaller, AI-native firms are using generative AI to internalize capabilities (like legal, coding, and design) that were previously outsourced, allowing them to scale without adding the overhead of a traditional firm. Workers who can use their experience to apply AI productively to complex problems have a strong role to play in this evolution. Generalist middle managers will need to respond.

While many financial institutions currently treat AI as a tool for efficiency (incremental), the retail banking and investment banking sectors show the strongest signs of a genuine reshuffle.

5. Conclusion: from measurement gap to structural transformation

We are undoubtedly at a turning point, with the conclusion on AI's impact being broadly "not yet". It is, however, poised to have just that impact. As we move into 2026, whilst we may see the resolution of Solow's paradox, it is also possible that we may see a wider impact at the whole system level. The consequences of this will be felt both in the skills we need to undertake our various tasks in the financial services organizations, and the very value attached to our roles. We are also likely to need to reform the regulatory frameworks of financial services to handle the emergence of AI-driven ecosystems.

The Reshuffle in financial services is only the beginning. AI is moving from being a "smarter brain" to a "better glue". Those who treat AI as a tool to automate the past will be destroyed by the Schumpeterian gale; those who embrace it as an engine to restack the future will lead the next technological revolution. Whilst the 2025 productivity statistics may remain sluggish for now, the infrastructure of the new economy is undoubtedly being established. The lag is not a sign of stagnation, but one of the immense complexity involved in rewriting the rules of business and society around a general-purpose engine.

References

Acemoglu, D., 2024, "The simple macroeconomics of AI," Economic Policy 40:121, 13–58

Arnon, A., 2025, "The projected impact of generative AI on future productivity growth," Penn Wharton Budget Model, https://tinyurl.com/c68v85t4

BG, 2025, "AI infrastructure construction: the next $400B boom in 2026," The Birmingham Group

Bresnahan, T., 2002, "Prospects for an information-technology-led productivity surge," NBER Innovation Policy and The Economy 2, 135-161

Brynjolfsson, E., 1993, "The productivity paradox of information technology," Communications of the ACM 36:12, 66-77

Brynjolfsson, E., 2021, "The productivity J-curve: how intangibles complement general purpose technologies," American Economic Journal: Macroeconomics 13:1, 333-372

Carney, M., 2018, "The future of work," Speech given at the Bank of England, September 14, https://tinyurl.com/yvwycskf

Challapally, A., C. Pease, R. Raskar, and P. Chari, 2025, "The GenAI divide: state of AI in business 2025, MIT NANDA, https://tinyurl.com/46kwyt62

Choudary, S., 2025, Reshuffle: who wins when AI restacks the knowledge economy, FutureFirst Books

David, P. A., 1989, "Computer and dynamo: the modern productivity paradox in a not-too distant mirror," Warwick economics research paper series no. 339, https://tinyurl.com/vbaks3sm

FIS, 2025, "Banks must educate as they innovate: over a third of UK consumers say financial services AI is moving too fast," press release, December 10, https://tinyurl.com/2u7t7kc6

Fisher Phillips, 2026, ""Slop" was the word of the year in 2025 – here's how employers can get smarter about AI use in 2026," January 8, https://tinyurl.com/nk8fykry

Gomstyn, A., and A. Jonker, 2025, "What is model collapse?" IBM Think, https://tinyurl.com/4whtpt3

GS, 2025a, "What is the US economy's potential growth rate?" Goldman Sachs, October 15, https://tinyurl.com/yzkn5uh8

GS, 2025b, "Why AI companies may invest more than $500 billion in 2026," Goldman Sachs, December 18, https://tinyurl.com/yubyyxm9

Haldane, A., 2015, "Growing, fast and slow," speech given at University of East Anglia, February 17, https://tinyurl.com/5xzzpc87

Henry, E., P-D. G. Sarte, and J. Taylor, 2024, "The productivity puzzle: AI, technology adoption and the workforce," Federal Reserve Bank of Richmond, Economic Brief no. 24-25

KPMG, 2025, "Intelligent banking - a blueprint for creating value through AI-driven transformation," https://tinyurl.com/ms8zmwcu

McKinsey, 2025, "Global banking annual review 2025: why precision, not heft, defines the future of banking," McKinsey & Co., https://tinyurl.com/bawj8wjh

MGI, 2023, "The economic potential of generative AI: the next productivity frontier," McKinsey Global Institute, McKinsey & Co., https://tinyurl.com/446uk2xw

MIT/EY, 2025, "Reimagining the future of banking with agentic AI" MIT Technology Review Insights, https://tinyurl.com/yfpj532a

Oxford Economics, 2026, "Evidence of an AI-driven shakeup of job markets is patchy," January, https://tinyurl.com/yrecfyf4

Perez, C., 2002, Technological revolutions and financial capital: the dynamics of bubbles and golden ages, Edward Elgar

Perez, C., 2009, "Technological revolutions and techno-economic paradigms" working papers in technology governance and economic dynamics no. 20, Tallinn University of Technology, https://tinyurl.com/34mkpdnm

Reuters, 2024, "Visa prevented $40 bln worth of fraudulent transactions in 2023 - official", July 23, https://tinyurl.com/374ysfdf

Sachse, H., B. Poddar, S. Tripathi, A. Kleppe, J. Uribe, and M. Schön, 2025, "From branches to bots: will AI agents transform retail banking?" Boston Consulting Group, November 5, https://tinyurl.com/2p9vhn3t

Stewart, H., 2026, "The cost of AI slop could cause a rethink that shakes the global economy in 2026," Guardian, January 4, https://tinyurl.com/56ssv8nj

UBS, 2025, "Robust fundamentals justify higher AI capex," UBS CIO daily updates, November 4, https://tinyurl.com/yc89nsx8

WEF, 2025, "AI paradoxes: 5 contradictions to watch in 2026 and why AI's future isn't straightforward," World Economic Forum, https://tinyurl.com/yc2caejb


Projective

How data and AI shape financial services

The **synthetic empathy framework**: merging AI precision with psychological insight for customer retention

Xuan Zhang,
Ph.D. Student in Computer Science, New Jersey Institute of Technology

Guiling Wang,
Distinguished Professor of Computer Science, New Jersey Institute of Technology

ABSTRACT

This paper examines how AI and psychological principles converge in customer retention, integrating data-driven systems with behavioral insights to enhance loyalty and lifetime value. Through theoretical analysis and case studies, we identify integration points where AI leverages psychological principles and misalignment risks where optimization undermines authentic connection. Three critical conclusions emerge. First, sustainable retention requires AI systems governed by psychological guardrails, preventing optimization from eroding trust. Second, organizations must leverage AI for scalable pattern recognition while maintaining human oversight in emotionally sensitive interactions. Third, competitive advantage arises from orchestrating artificial and human intelligence, where algorithms reinforce rather than replace psychological connection. The future of customer retention depends on intentionally designed AI–psychology systems grounded in transparency, ethical constraint, and strategic collaboration.

1. Introduction

In the contemporary landscape of customer retention, organizations face intensifying challenges. The cost of acquiring a new customer can be five to twenty-five times higher than retaining an existing one, while a mere 5% increase in customer retention rates can increase profits by 25% to 95% [Reichheld and Schefter (2000)]. Despite these compelling economics, industries continue to grapple with significant churn rates, with studies indicating that traditional businesses lose approximately 20-25% of their customer base annually in mature markets [Reichheld and Sasser (1990)].

This retention challenge has intensified due to several converging forces. Digital-native competitors have lowered switching barriers and heightened customer expectations for seamless, personalized experiences. Technological advances have made it easier for customers to compare alternatives and migrate across providers with minimal friction. Meanwhile, the COVID-19 pandemic accelerated digital adoption across sectors, fundamentally altering customer interaction patterns and creating new vulnerabilities in traditional relationship-based retention strategies.

Against this backdrop, two powerful forces have emerged as potential game-changers: artificial intelligence (AI) and behavioral psychology. AI technologies promise unprecedented capabilities in predicting customer behavior, personalizing interactions, and automating retention interventions at scale [Provost and Fawcett (2013), Davenport et al. (2020), Kumar et al. (2019)]. Simultaneously, advances in behavioral economics have deepened our understanding of non-rational factors driving customer loyalty. Research demonstrates that customer decisions are profoundly influenced by cog-

Table 1: Mapping psychological factors to AI capabilities and integration opportunities

Psychological factors	What AI technical can help	Integration opportunities
2.1 Trust and perceived reliability	3.1 Real-time optimization[1] 3.4 Algorithmic optimization at scale [2]	4.1.1 Trust enhancement through transparency
2.2 Emotional connection and affective experience	3.2 Hyper-personalization[3] 3.3 Discovering non-obvious patterns [d]	4.1.2 Emotionally responsive engagement
2.3 Perceived value and satisfaction	3.2 Hyper-personalization[5] 3.5 Predictive modeling for early churn detection[6]	4.1.3 Dynamic value optimization
2.4 Loss aversion and status-quo bias	3.1 Real-time optimization[7] 3.5 Predictive modeling[8]	4.1.4 Informed inertia and loss awareness
2.5 Psychological ownership and identity integration	3.2 Hyper-personalization[9] 3.3 Discovering non-obvious patterns[10]	4.1.5 Identity-relevant personalization
2.6 Social Influence and normative pressure	3.3 Discovering non-obvious patterns[11]	4.1.6 Positive social influence
2.7 Cognitive biases in decision-making	3.1 Real-time optimization[12] 3.3 Discovering non-obvious patterns[13]	4.1.7 Bias-aware choice support
2.8 Reciprocity and commitment	3.1 Real-time optimization[14] 3.2 Hyper-personalization[15]	4.1.8 Timely relationship investment

Notes: Section 4.2 discusses misalignment risks across all integration points.

(1) Singh et al. (2024), Verbeke et al. (2011); (2) Singh et al. (2024), Mamun (2025); (3) Mamun (2025), Imani et al. (2025); (4) Mamun (2025), Imani et al. (2025); (5) Mamun (2025), Imani et al. (2025); (6) Verbeke et al. (2011), Shaikhsurab and Magadum (2024); (7) Singh et al. (2024), Verbeke et al. (2011); (8) Verbeke et al. (2011), Shaikhsurab and Magadum (2024); (9) Mamun (2025), Imani et al. (2025); (10) Mamun (2025), Imani et al. (2025); (11) Mamun (2025), Imani et al. (2025); (12) Singh et al. (2024), Verbeke et al. (2011); (13) Mamun (2025), Imani et al. (2025); (14) Singh et al. (2024), Verbeke et al. (2011); (15) Mamun (2025), Imani et al. (2025).

nitive biases [Kai-Ineman and Tversky (1979)], emotional states [Schwartz (2008), Morgan and Hunt (1994)], and social contexts [Cialdini (1993)].

Customers remain loyal due to trust, emotional connection, perceived reciprocity, and psychological switching costs rather than merely product features or pricing.

Yet this convergence is neither automatic nor without tension. AI systems may perpetuate biases, optimize for short-term metrics at the expense of long-term trust [Davenport et al. (2020)], or satisfy efficiency metrics while failing to meet customers' psychological needs for recognition and empathy [Parasuraman et al. (1988)]. Understanding where AI and psychology align and where they diverge is essential for organizations seeking to harness these tools responsibly.

This paper makes several contributions. First, we synthesize psychological foundations of customer retention from behavioral economics, social psychology, and consumer research [Kai-Ineman and Tversky (1979), Thaler (1985), Watson and Kahneman (2011), Cialdini (1993), Morgan and Hunt (1994)]. Second, we analyze how AI-powered approaches fundamentally differ from traditional methods [Davenport et al. (2020), Kumar et al. (2019)]. Third, we explore critical integration points between AI and psychology, identifying where machine learning leverages psychological principles and where algorithmic approaches may undermine human needs. Fourth, we present case studies illustrating both successful AI-psychology integration and cautionary examples. Finally, we address ethical dimensions and future challenges of this convergence.

Table 1 provides an overview of how psychological factors map to AI capabilities and integration opportunities explored throughout this paper.

2. Psychological foundations of customer retention

2.1 Trust and perceived reliability

Trust emerges consistently across the literature as the cornerstone psychological factor in customer retention [Dwyer et al. (1987), Garbarino and Johnson (1999), Morgan and Hunt (1994), Zeithaml et al. (1996)]. In financial services, customers entrust institutions with their financial security and future well-being. In telecommunications, customers trust connectivity reliability and data security. In e-commerce, trust concerns product quality, delivery reliability, and payment security. Across all contexts, trust encompasses competence trust (belief in the provider's ability to deliver services effectively), integrity trust (confidence in ethical conduct), and benevolence trust (perception that the provider has customers' best interests at heart). Research demonstrates that trust significantly influences both satisfaction and loyalty, with customers exhibiting higher tolerance for service failures when baseline trust is strong [Garbarino and Johnson (1999), Morgan and Hunt (1994)].

2.2 Emotional connection and affective experience

Beyond cognitive evaluations of service quality, emotional connections play a critical role in retention. The affective dimension of customer experience encompasses feelings of comfort, safety, excitement, and enjoyment during service interactions. Research distinguishes between positive emotions (joy, trust, satisfaction) and negative emotions (frustration, anger, anxiety), with empirical evidence showing that emotional valence strongly predicts churn behavior, often more powerfully than purely functional service attributes [Gremler and Brown (1996), Oliver (1999), Wang et al. (2025)]. Emotions are particularly amplified in high-stakes contexts such as financial services. Customers experiencing negative emotions during service failures or financial stress are significantly more likely to churn, even when the institution resolves the immediate problem [Wang et al. (2025)]. Conversely, positive emotional experiences create psychological bonds that transcend rational cost benefit calculations, explaining why some customers remain loyal despite competitors offering better rates or features.

2.3 Perceived value and satisfaction

Perceived value represents the customer's overall assessment of service benefits relative to costs (monetary, time, effort, and psychological). Satisfaction reflects the degree to which service experiences meet or exceed expectations. Both constructs are fundamental predictors of retention, yet they operate through distinct psychological mechanisms [Oliver (1999), Zeithaml et al. (1996)]. Research reveals that satisfaction alone is insufficient for retention. Even satisfied customers may churn if they perceive insufficient value or encounter switching inducements [Jones et al. (2002)]. Moreover, the satisfaction-retention relationship is non-linear: moving customers from dissatisfied to satisfied yields larger retention gains than moving them from satisfied to highly satisfied [Zeithaml et al. (1996)]. Organizations must, therefore, deliver functional benefits while shaping customer perceptions of value through effective communication, personalization, and strategic framing of service offerings [Becker-Olsen (2003)].

2.4 Loss aversion and status-quo bias

Loss aversion, the tendency to feel losses more acutely than equivalent gains, creates psychological switching costs that anchor customers to existing providers [Kai-Ineman and Tversky (1979), Thaler (1985)]. Customers perceive switching as potentially losing accumulated benefits, such as relationship history, familiar interfaces, and established routines, while gains from competitors appear less certain. This asymmetry favors incumbent providers and explains why customers often tolerate suboptimal services rather than switch. Status-quo bias, the preference for maintaining current arrangements absent compelling reasons to change, further reinforces inertia through multiple mechanisms: cognitive effort required to evaluate alternatives, anxiety about unknown outcomes, and endowment effects (overvaluing what one already possesses) [Samuelson and Zeckhauser (1988), Thaler (1980)]. While these biases benefit retention in the short term, they also mask underlying dissatisfaction that may eventually trigger abrupt churn when a critical threshold is crossed.

2.5 Psychological ownership and identity integration

Psychological ownership occurs when customers develop possessive feelings toward their service relationships, viewing accounts, portfolios, or subscriptions as extensions of self-identity. This sense of ownership creates strong retention effects because switching would involve relinquishing part of one's identity [Wang et al. (2025)]. The phenomenon is particularly pronounced when customers have invested significant time, effort, or emotional resources in configuring services to their preferences.

Identity integration encompasses the degree to which a service provider becomes part of a customer's self-concept or social identity. Customers who identify strongly with a brand's values or participate in service communities develop deeper psychological bonds [Fullerton (2005), Verhoef et al. (2002)]. Research shows that these customers exhibit higher retention rates, greater word-of-mouth advocacy, and increased willingness to cross-buy additional services.

2.6 Social influence and normative pressure

Customer retention decisions are shaped by social context and normative pressures. Social influence operates through observational learning (imitating peers' choices), informational social influence (using others' decisions as signals of quality), and normative social influence (conforming to expectations of one's social group). Research demonstrates that customers are more likely to switch providers if they observe friends or family members doing so, and conversely, are more likely to stay if switching would violate social norms [Cialdini (1993), Becker-Olsen (2003)].

2.7 The role of reciprocity and commitment

Reciprocity, the norm of returning favors and maintaining balanced exchanges, creates retention effects when customers perceive their provider has invested in the relationship beyond transactional requirements. Personalized advice, proactive problem resolution, loyalty rewards, and relationship gestures trigger reciprocity obligations that psychologically bind customers [Dwyer et al. (1987), Garbarino and Johnson (1999)]. Research shows that customers who feel indebted to their provider through past acts of goodwill exhibit higher retention rates and forgive service failures more readily.

Commitment encompasses both affective (emotional attachment) and calculative (rational assessment of switching costs) dimensions. Affective

commitment arises from identification with the provider and intrinsic enjoyment of the relationship [Fullerton (2005)]. Calculative commitment stems from recognizing that switching would sacrifice accumulated benefits or incur costs [Jones et al. (2002)]. The strongest retention occurs when both forms align, though pure calculative commitment is fragile and easily disrupted when competitors eliminate switching barriers or offer superior inducements [Morgan and Hunt (1994)].

2.8 Cognitive biases in financial decision-making

Multiple cognitive biases shape customer retention across industries. The anchoring effect causes customers to fixate on initial reference points (original contract terms, promotional rates), making subsequent adjustments feel like losses [Kai-Ineman and Tversky (1979)]. The availability heuristic leads customers to overweight recent or vivid experiences when evaluating provider quality, meaning a single negative incident can disproportionately drive churn [Oliver (1999)]. Confirmation bias causes customers to selectively attend to information confirming pre-existing beliefs about their provider, either reinforcing loyalty or accelerating exit. The bandwagon effect and herding behavior create momentum in retention decisions, where individual choices are influenced by aggregate trends [Cialdini (1993)]. Hyperbolic discounting makes customers overvalue immediate benefits (switching bonuses) relative to long-term relationship value [Thaler (1980)]. Mental accounting leads customers to evaluate services in separate mental buckets, sometimes making economically suboptimal retention decisions [Thaler (1985), Thaler (1980)]. Understanding these biases is essential for designing interventions, whether traditional or AI-driven, that can identify bias-driven churn risk and deploy appropriate countermeasures [Jones et al. (2002), Verhoef et al. (2002)].

3. AI-driven approaches to customer retention

Traditional customer retention strategies have long relied on demographic segmentation, periodic surveys, and reactive interventions [Ascarza (2018), Hadden et al. (2007), Neslin et al. (2006)]. Artificial intelligence transforms retention by enabling real-time behavioral prediction, hyper-personalization at scale, proactive intervention, and continuous learning from customer interactions [Imani et al. (2025), Mamun (2025), Singh et al. (2024)]. This section examines how AI-driven approaches create new capabilities for customer retention. Table 2 summarizes the major AI methods used in customer retention, their capabilities, and typical applications.

3.1 Real-time optimization and continuous engagement

Traditional retention programs operate on fixed schedules: annual renewal campaigns, quarterly loyalty communications, and predetermined intervention points. This temporal rigidity means customers may receive irrelevant offers at inopportune moments, or miss critical support during periods of heightened churn risk [Hadden et al. (2007), Neslin et al. (2006)]. AI systems enable continuous engagement optimization through real-time decision engines that determine the next best action for each customer at each moment [Singh et al. (2024)]. Reinforcement learning algorithms learn optimal intervention timing, channel selection, and message content through trial and feedback. Stream processing architectures analyze customer interactions as they occur, updating risk scores and triggering interventions within seconds or minutes [Sana et al. (2025), Rahman et al. (2024)]. Timing is critical to psychological impact. Behavioral economics research shows that interventions are most effective when delivered at moments of heightened attention and receptivity, what psychologists call “teachable moments”. AI’s real-time capability allows institutions to engage customers when psychological conditions are favorable: immediately after positive experiences (to reinforce satisfaction), during periods of uncertainty (to provide reassurance), or just before contemplated switches (to address concerns) [Singh et al. (2024), Verbeke et al. (2011)].

3.2 Hyper-personalization through individual-level modeling

Traditional retention strategies segment customers into broad categories based on demographics, product holdings, or simple behavioral rules. AI enables true individual-level modeling where each customer has a unique retention profile based on their specific behavioral history, preferences, and predicted needs [Imani et al. (2025), Mamun (2025)]. Recommender systems identify products and services aligned with individual preferences by learning from similar customers' behaviors. Natural language processing extracts sentiment and intent from customer communications, capturing emotional states and evolving needs [Guo et al. (2024), Abdul-Rahman et al. (2024)]. Graph neural networks model customer relationships and social influence, identifying peer effects and community structures that influence retention [Anitha and Sherly (2024), Lee and Woo (2025)]. Individualization aligns with the psychological principle of perceived value, which is inherently subjective and varies across customers. Traditional segmentation assumes within-group homogeneity that does not exist psychologically. AI's individual level modeling respects the reality that retention drivers are person-specific, enabling interventions tailored to each customer's unique psychological profile [Mamun (2025), Adekunle et al. (2023)].

3.3 Discovering non-obvious patterns and continuous adaptation

Traditional retention strategies rely on explicit business rules limited by human observation capacity and struggle with complex, high-dimensional pattern spaces [Hadden et al. (2007), Neslin et al. (2006)]. Machine learning models automatically discover predictive patterns through random forests, gradient boosting, and deep neural networks that capture non-linear interactions and subtle behavioral signatures [Imani et al. (2025), Mamun (2025)]. For example, AI might reveal that customers contacting service exactly twice monthly show higher churn risk, suggesting ambivalence rather than satisfaction [Hadden et al. (2007)]. AI systems implement continuous learning through online algorithms, A/B testing frameworks, and automated monitoring that detect concept drift and trigger retraining [Ahlstrand et al. (2025), Singh et al. (2024), Li et al. (2024), Rahman et al. (2024)]. This adaptive capability aligns with the psychological reality that retention drivers evolve with market conditions and societal trends [Ascarza (2018)], ensuring strategies remain aligned with current psychological states [Imani et al. (2025), Singh et al. (2024)].

3.4 Algorithmic optimization at scale

Traditional retention depends on human judgment, which offers contextual understanding and empathy but faces critical limitations: limited scale, cognitive biases, and inconsistent quality across employees and situations [Hadden et al. (2007), Neslin et al. (2006)]. AI systems optimize retention decisions across millions of customers with perfect consistency. Mathematical optimization solves problems infeasible for humans: allocating retention budgets across thousands of at-risk customers, personalizing offers to maximize retention while minimizing costs, and sequencing multi-touch campaigns across channels [Mamun (2025), Singh et al. (2024)]. Causal inference methods identify which customers will respond to interventions versus those who would stay anyway, eliminating wasteful spending on retention efforts that provide no marginal benefit [Verbeke et al. (2011)]. AI systems properly weight evidence across thousands of outcomes, identifying true retention drivers rather than spurious correlations. This statistical rigor ensures strategies target the right customers with the right interventions at the right times, matching offers to psychological readiness [Ascarza (2018), Adekunle et al. (2023)].

3.5 Predictive modeling for early churn detection

Traditional retention efforts are predominantly reactive, responding to observable signals such as missed payments or explicit complaints. Customer satisfaction surveys provide delayed snapshots rather than real-time insights.

Machine learning models predict churn risk weeks or months in advance by analyzing hundreds of behavioral signals simultaneously [Shaikhsurab and Magadum (2024), Verbeke et al. (2011), Kabbar and Herath (2025), Li and Yan (2025)]. Gradient boosting methods and deep neural networks identify subtle pattern changes invisible to human analysts: gradual decreases in login frequency, shifts in transaction timing, or emerging service quality issues [Imani et al. (2025), Mamun (2025), Bhattacharjee et al. (2023)]. Survival analysis models predict not just whether customers will churn but when, enabling precisely timed interventions [Günther et al. (2014), Singh (2024), Zhang et al. (2025)].

This predictive capability addresses the psychological reality that churn is a process, not an event. By detecting early-stage signals (declining engagement or negative sentiment), AI systems identify customers in the contemplation stages of switching, when psychological bonds are weakening but not yet severed, the optimal window for retention intervention [Imani et al. (2025), Mamun (2025), Adekunle et al. (2023)].

Table 2: AI methods for customer retention

Model category	Representative algorithms	Key capabilities	Application scenarios
Gradient boosting models	XGBoost, LightGBM, CatBoost[1]	High-accuracy churn prediction, feature importance ranking, handling imbalanced data	Churn risk scoring, customer segmentation, early warning systems
Deep neural networks	LSTM, Transformers[2]	Sequential pattern learning, multimodal data fusion, representation learning	Temporal behavior modeling, time-series prediction, customer embeddings
Survival analysis models	Cox proportional hazards, Random Survival Forests, DeepHit[3]	Time-to-event prediction, censored data handling, intervention timing optimization	Predicting when churn occurs, optimal contact scheduling, lifecycle modeling
Reinforcement learning	Multi-armed bandits, contextual bandits, policy gradients[4]	Dynamic decision-making, online learning, exploration-exploitation balance	Next-best-action recommendation, real-time offer optimization, adaptive engagement
Natural language processing	BERT, GPT, sentiment classifiers, topic models[5]	Emotion detection, intent extraction, semantic understanding	Customer communication analysis, satisfaction monitoring, complaint detection
Graph neural networks	GraphSAGE, GAT[6]	Network effect modeling, social influence propagation, community detection	Peer effect identification, viral churn prediction, community-based retention
Recommender systems	Collaborative filtering, content-based filtering, matrix factorization, hybrid methods[7]	Personalized recommendations, preference learning, cold-start handling	Product/service suggestions, cross-sell optimization, personalized bundling
Causal inference methods	Uplift modeling, propensity score matching, heterogeneous treatment effects[8]	Treatment effect estimation, counterfactual reasoning, intervention targeting	Identifying treatment-responsive customers, ROI optimization, A/B testing

Notes: (1) Asif et al. (2025), Boozary et al. (2025), Chang et al. (2024); (2) Imani et al. (2025), Mamun (2025), Li et al. (2024); (3) Günther et al. (2014), Singh (2024), Zhang et al. (2025); (4) Ahlstrand et al. (2025), Singh et al. (2024), Verbeke et al. (2011); (5) Abdul-Rahman et al. (2024), Guo et al. (2024), Martın-Donas et al. (2024); (6) Lee and Woo (2025), Anitha and Sherly (2024); (7) Boozary et al. (2025), Imani et al. (2025), Mamun (2025) ; (8) Ljubicic et al. (2023), Verbeke et al. (2011).

4. AI-psychology integration

AI's impact on retention depends not merely on technical sophistication but on alignment with psychological principles. Note that the same capabilities that enable personalization can either enhance trust or trigger reactance.

4.1 AI-psychology integration: synergistic opportunities

4.1.1 trust enhancement through transparency

AI systems integrate naturally with trust-building mechanisms by delivering highly consistent, scalable, and error-resistant service. Predictive models and automated decision engines strengthen competence trust by providing uniform service quality independent of human fatigue, emotional variation, or information overload [Morgan and Hunt (1994)]. This integration gets significantly stronger when paired with meaningful transparency. Explainable AI techniques, such as feature attribution or attention-based explanations, support integrity and benevolence trust by clarifying why specific recommendations or retention offers are made [Doshi-Velez and Kim (2017), Ribeiro et al. (2016)]. When customers understand how their behaviors influence an AI-driven outcome, they perceive algorithmic decisions as fair and aligned with their interests. Trust-preserving AI systems explicitly communicate uncertainty, define clear boundaries for human escalation in high-stakes cases, and provide explanations calibrated to customer literacy levels.

4.1.2 Emotionally responsive engagement

AI enables emotionally responsive engagement at scale unattainable by human agents alone. By detecting affective signals in text, voice, and behavior, AI systems can time interventions, adjust tone, and personalize interactions in ways that make customers feel recognized and supported. When personalization is perceived as helpful (anticipating needs, reducing friction, or offering reassurance during stress), it strengthens emotional bonds by signaling attentiveness and care [Aguirre et al. (2016)]. Emotion-aware systems must respect personalization boundaries, prioritize customer benefit over short-term retention metrics, and incorporate human intervention when emotional distress is non-trivial [Chellappa and Sin (2005)]. When emotional data is used to genuinely support customers rather than exploit vulnerabilities, AI-driven personalization fosters authentic emotional connection.

4.1.3 Dynamic Value Optimization

AI integrates naturally with value creation by enabling individualized optimization: tailoring product bundles, pricing structures, and service enhancements to each customer's preference profile. When customers perceive that AI-driven adjustments reduce waste, lower friction, or better align offerings with actual needs, dynamic optimization strengthens perceived value and sustains satisfaction [Zeithaml (1988), Oliver (1999)]. AI also enables proactive satisfaction management by forecasting declining engagement from behavioral signals, allowing firms to intervene before dissatisfaction becomes salient, through feature education, service adjustments, or value-added gestures. When perceived as genuine improvements rather than reactive appeasement, these interventions reinforce satisfaction by exceeding expectations [Parasuraman et al. (1988)].

4.1.4 Informed inertia and loss awareness

AI integrates with loss aversion and status-quo bias by identifying when inertia reflects genuine customer benefit and reinforcing it through deeper personalization and service integration. Customized dashboards, accumulated preferences, and integrated service ecosystems increase psychological ownership, making continued engagement feel valuable rather than merely habitual. When predictive models detect early switching intent, timely interventions that highlight accumulated benefits, such as relationship history, bundled services, and loyalty rewards, help customers fully evaluate potential losses they might otherwise overlook. When framed transparently, such interventions reinforce loss awareness without coercion, aligning retention with customer welfare [Thaler and Sunstein (2009)].

4.1.5 Identity-relevant personalization

AI enables deep, individualized customization and adaptive learning that transforms services into personally meaningful extensions of the self [Pierce et al. (2001), Belk (1988)]. When customers configure dashboards, goals, and interfaces around personally salient objectives, and as systems improve through continued interaction, engagement becomes

identity-relevant rather than merely habitual. AI also facilitates identity integration through social and community mechanisms. Recommendation systems and graph-based models connect customers with similar goals or interests, embedding services within social contexts that reinforce belonging and shared identity. When communities contribute to customers' aspirational identities, retention is driven by relational value rather than technical dependence [Muniz and O'Guinn (2001)].

4.1.6 Positive social influence

AI integrates with social dynamics by identifying social influence pathways and amplifying positive peer effects. Network analysis enables firms to distinguish between customers who drive collective behavior and those primarily influenced by others, allowing interventions that reinforce supportive norms. When used constructively, AI-enhanced communities create retention through shared value, mutual support, and social belonging [Aral et al. (2009)]. AI systems also enhance informational social influence by detecting emerging collective dissatisfaction through social listening and sentiment analysis. Early identification of viral complaints enables organizations to address underlying service failures before they escalate. Authentic social proof, such as verified customer success stories or transparent community metrics, helps uncertain customers contextualize their experiences and reduce unnecessary switching [Chen et al. (2011)].

4.1.7 Bias-aware choice support

AI enables personalized choice architecture by tailoring framing, defaults, and timing to support better decision-making without eliminating freedom of choice. When aligned with customer welfare, such nudges can counteract well-documented biases. Default options that encourage beneficial behaviors leverage inertia to improve outcomes, while long-term framing can mitigate hyperbolic discounting [Thaler and Sunstein (2009)]. AI systems can also correct distorted reference points. Personalized contextual information helps recalibrate anchoring on outdated prices or isolated negative experiences by providing base-rate statistics and comparative benchmarks. Transparent framing of rewards and benefits can work with mental accounting tendencies to enhance perceived value without deception [Thaler (1985)].

4.1.8 Timely relationship investment

AI enables timely, personalized relationship investments that feel responsive rather than reactive. When AI systems identify moments of genuine customer need, such as service disruption or life transitions, and trigger meaningful assistance, reciprocity emerges naturally as gratitude rather than obligation. Such interventions strengthen affective commitment by signaling care [Morgan and Hunt (1994)]. AI can also reinforce commitment through recognition and surprise. Personalized milestone acknowledgment and unexpected benefits, which are delivered without explicit quid pro quo, create emotional resonance that strengthens relational bonds. When customers perceive these gestures as authentic appreciation rather than retention tactics, reciprocity deepens trust and willingness to continue the relationship [Cialdini (1993)].

4.2 Misalignment risks: when algorithmic optimization undermines psychological retention

While AI offers powerful integration opportunities, the same capabilities introduce systematic misalignment risks that can undermine the psychological foundations of customer retention. Across multiple mechanisms, a consistent pattern emerges: algorithmic optimization designed to maximize short-term retention metrics can paradoxically erode the long-term psychological bonds that sustain genuine loyalty.

4.2.1 Opacity and manipulation threats

High-performing AI models often remain partially opaque, and when customers encounter unexplained decisions such as credit denials and inconsistent offers integrity trust erodes quickly [Dietvorst et al. (2015)]. Hyper-personalization can trigger psychological reactance when customers perceive their vulnerabilities are being exploited rather than supported [Brehm (1966), White et al. (2008)]. Automated empathy without substantive assistance such as scripted acknowledgments followed by ineffective responses, intensifies negative emotions rather than resolving them [Gray et al. (2018)].

4.2.2 Fairness violations and expectation instability

Personalized pricing and differential treatment violate fairness norms when customers discover inconsistencies, reframing value optimization as

exploitation [Xia et al. (2004), Bolton et al. (2003)]. When AI-driven offers change frequently or appear contingent on churn threats, satisfaction becomes transactional and strategic, weakening long-term loyalty. Similarly, deliberate friction in cancellation flows or repeated retention hurdles transform psychological inertia into constraint, triggering resentment, and negative word-of-mouth [Gray et al. (2018)].

4.2.3 Lock-in and coercive dynamics

Excessive personalization can create psychological lock-in where leaving feels overwhelming rather than merely inconvenient. Filter bubbles that reinforce narrow preference profiles may maximize short-term satisfaction but constrain exploration and autonomy [Rowland (2011)]. Algorithmic manipulation of social proof, including suppressing negative feedback, manufacturing consensus, or imposing social costs on exit, transforms social influence into coercion, accelerating long-term disengagement [Cialdini (1993), Sunstein (2016)]. AI-driven experimentation that identifies bias combinations to maximize persistence through exploiting anchoring, scarcity, or loss aversion erodes informed consent and undermines retention legitimacy [Sunstein (2016), Gray et al. (2018)].

4.2.4 The core tension

These misalignment risk patterns reflect a fundamental tension: AI systems optimized purely for behavioral persistence may systematically undermine the trust, autonomy, and authentic connection that sustain durable relationships. When reciprocity becomes overtly transactional, and when customers infer that assistance is algorithmically triggered by churn risk, perceived manipulation replaces genuine appreciation [Gouldner (1960)]. Effective retention requires treating psychology as a constraint to be respected rather than a resource to be exploited, preserving customer agency through transparency, ethical guardrails, and deliberate human-AI collaboration at psychologically sensitive moments.

5. Case studies

This section examines two contrasting cases that illuminate the integration-misalignment dynamics explored in Section 3. The first demonstrates successful AI-psychology integration in financial services churn prediction, while the second reveals how algorithmic optimization can undermine psychological retention drivers when human elements are neglected.

5.1 Synergistic integration success: multimodal fusion in financial services

Rudd et al. (2023) propose a multimodal churn prediction system for financial services that integrates affective and knowledge-related signals with traditional behavioral data. The system combines three information streams: (1) estimated customer financial literacy (FL), reflecting customers' ability to comprehend financial products; (2) voice-based emotion recognition from customer service calls, capturing affective states associated with dissatisfaction; and (3) conventional CRM behavioral data.

The multimodal model achieves 91.2% test accuracy and outperforms single-modality baselines. Correlation analyses indicate that customers exhibiting negative emotional signals and lower financial literacy scores are more likely to be classified as high churn risk, patterns not fully captured by behavioral data alone.

This case illustrates how AI-based pattern recognition can be strengthened through psychologically informed feature design. Emotional connection dynamics are operationalized through voice-based affect detection, while cognitive limitations related to decision-making are indirectly captured through financial literacy estimation. Importantly, the multimodal structure supports interpretive insight into why customers are flagged as high risk by linking churn to emotional distress and limited cognitive resources, rather than treating risk as an opaque behavioral score. The study suggests that retention strategies may benefit from combining behavioral analytics with affective and knowledge-related signals, enabling more psychologically informed interventions such as financial education or empathetic human outreach while maintaining AI scalability.

5.2 Misalignment risk lessons: when AI automation undermines human connection

A 2022 study of home-sharing platforms in China (Airbnb, Tujia, Xiaozhu) investigated how users' perceived use of AI relates to customer trust, engagement, and loyalty [Chen et al. (2022)]. Drawing on survey data from 468 users, the study examines whether AI moderates the trust-engagement-loyalty pathways that underpin customer retention in the sharing economy. Contrary to expectations, results show that perceived AI functionality negatively moderates both the relationship between platform trust and customer engagement ($\beta = -0.13$, $p < 0.001$) and the relationship between engagement and loyalty ($\beta = -0.17$, $p < 0.001$) [17]. As AI becomes more salient in customer interactions, trust is less effectively converted into engagement, and engagement is less likely to translate into loyalty. This provides empirical support for the misalignment risks discussed in Section 3.

The authors suggest that AI-mediated interactions may attenuate relational processes because customers perceive algorithmic services as efficiency-oriented rather than relationally attentive. Even when customers report trust in the platform, interactions mediated by AI may feel less caring or responsive to emotional nuance. Trust in human hosts remains influential, whereas platform trust subject to strong AI mediation shows diminished effect, indicating that customers differentiate between human and algorithmic relationship elements.

This case highlights several degradation risks:

- Trust erosion: opaque AI intermediaries may raise questions about whether decisions prioritize customer welfare or platform optimization.
- Authenticity deficit: automated personalization can be recognized as transactional optimization rather than genuine care.
- Reduced reciprocity: AI-delivered service does not readily evoke norms of obligation, limiting affective commitment compared to human-provided assistance.

The findings suggest that high touch, relationship-building interactions may need to remain human-led, with AI primarily supporting routine tasks, while explicit signals of human oversight may mitigate perceptions of impersonality.

6. Strategic focus areas for AI-enabled customer retention

For organizations seeking to differentiate through AI-enabled customer retention, several strategic focus areas warrant sustained leadership attention and deliberate investment:

- **Temporal dynamics**: executives must recognize that psychological retention drivers evolve across customer lifecycles, economic cycles, and competitive contexts. Most current AI models assume static relationships, limiting their effectiveness over time. Developing dynamically adaptive models that track shifting customer expectations and emotional states represents a critical opportunity for differentiation.
- **Intervention fatigue**: organizations need to actively manage the frequency, timing, and intensity of AI-driven retention interventions to avoid diminishing returns and customer reactance. Over-optimization of engagement risks crossing the boundary from supportive personalization to perceived intrusiveness, undermining trust and long-term loyalty.
- **Human–AI collaboration**: leaders must make deliberate choices about task allocation between humans and AI. Empathy-critical moments, trust-repair interactions, and complex advisory decisions should remain human-led, while AI augments decision quality through prediction, prioritization, and recommendation. Competitive advantage increasingly depends on orchestrating hybrid interaction models rather than pursuing full automation.
- **Retention versus acquisition trade-offs**: executives should reassess traditional assumptions that maximizing retention is always optimal. AI-driven optimization may reveal non-intuitive equilibria in which allowing low-value or disengaged customers to churn improves overall profitability and service quality. Strategic focus should shift from retention quantity to retention quality.

7. Conclusion

This paper has explored the convergence of artificial and human intelligence in customer retention, examining how AI capabilities and psychological principles can be strategically integrated to enhance customer loyalty and lifetime value.

Three key insights emerge from our analysis: First, effective AI deployment in retention contexts demands built-in safeguards that protect the trust and authenticity customers expect, preventing algorithmic efficiency from compromising relationship quality. Second, success requires identifying the optimal balance point: utilizing AI's capacity for large-scale analysis and real-time responsiveness while maintaining human engagement in emotionally charged or high-stakes situations. Third, market leadership will accrue to firms that skillfully coordinate machine and human capabilities rather than simply deploying the most sophisticated technology, emphasizing augmentation over automation in customer relationships.

The path forward centers on intentional design: implementing transparency in algorithmic decisions, establishing ethical boundaries for persuasive technologies, and allocating tasks between humans and machines based on psychological sensitivities. Organizations that embrace customer wellbeing as a primary goal, and not merely as a consideration, will build sustainable advantages in an increasingly digital marketplace.

References

Abdul-Rahman, S., M. F. Akif Md Ali, A. Abu Bakar, and S. Mu talib, 2024, "Enhancing churn forecasting with sentiment analysis of steam reviews," Social Network Analysis and Mining 14:1, 178

Adekunle, B. I., E. C. Chukwuma-Eke, E. D. Balogun, and K. Olu sola Ogunsola, 2023, "Improving customer retention through machine learning: A predictive approach to churn prevention and engagement strategies," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 9:4, 507-523

Aguirre, E., A. L. Roggeveen, D. Grewal, and M. Wetzels, 2016, "The personalization privacy paradox: implications for new media," Journal of Consumer Marketing 33:2, 98-110

Ahlstrand, J., A. Borg, H. Grahn, and M. Boldt, 2025"Using transformers for b2b contractual churn prediction based on customer behavior data," presented at the International Conference on Enterprise Information Systems (ICEIS) 2025, Apr 4-6

Anitha, M. A., and K. K. Sherly, 2024, "Customer churn prediction using graphsage model with degree based sampling and max pooling aggregation," presented at the International Conference on Computing, Communication, Security and Intelligent Systems, Springer, 103-117

Aral, S., L. Muchnik, and A. Sundararajan, 2009, "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks," Proceedings of the National Academy of Sciences 106:51, 21544-21549

Ascarza, E., 2018, "Retention futility: targeting high-risk customers might be ineffective," Journal of marketing Research 55:1, 80-98

Asif, D., M. S. Arif, and A. Mukheimer, 2025, "A data-driven approach with explainable artificial intelligence for customer churn prediction in the telecommunications industry," Results in Engineering 26, 104629

Becker-Olsen, K. L., 2003, "And now, a word from our sponsor – a look at the effects of sponsored content and banner advertising," Journal of Advertising 32:2, 17-32

Belk, R. W., 1988, "Possessions and the extended self," Journal of consumer research 15:2, 139-168

Bhattacharjee, S., U. Thukral, and N. Patil, 2023, "Early churn prediction from large scale user-product interaction time series," presented at the 2023 International Conference on Machine Learning and Applications (ICMLA), IEEE, 2079-2086

Bolton, L. E., L. Warlop, and J. W. Alba, 2003, "Consumer perceptions of price (un) fairness," Journal of consumer research 29:4, 474-491

Boozary, P., S. Sheykhan, H. GhorbanTanhaei, and C. Magazzino, 2025, "Enhancing customer retention with machine learning: a comparative analysis of ensemble models for accurate churn prediction," International Journal of Information Management Data Insights 5:1, 100331

Brehm, J. W., 1966, A theory of psychological reactance, Academic Press

Chang, V., K. Hall, Q. A. Xu, F. O. Amao, M. A. Ganatra, and V. Benson, 2024, "Prediction of customer churn behavior in the telecommunication industry using machine learning models," Algorithms 17:6, 231

Chellappa, R. K., and R. G Sin, 2005, "Personalization versus privacy: An empirical examination of the online consumer's dilemma," Information technology and management 6:2, 181-202

Chen, Y., C. Prentice, S. Weaven, and A. Hisao, 2022, "The influence of customer trust and artificial intelligence on customer engagement and loyalty–the case of the home-sharing industry," Frontiers in psychology 13, 912339

Chen, Y., Q. Wang, and J. Xie, 2011, "Online social interactions: a natural experiment on word of mouth versus observational learning,"
Journal of marketing research 48:2, 238-254

Cialdini, R. B., 1993, Influence: the psychology of persuasion (revised edition), Morrow

Davenport, T., A. Guha, D. Grewal, and T. Bressgott, 2020, "How artificial intelligence will change the future of marketing," Journal of the Academy of Marketing Science 48:1, 24-42

Dietvorst, B. J., J. P. Simmons, and C. Massey, 2015, "Algorithm aversion: people erroneously avoid algorithms after seeing them err," Journal of experimental psychology: General 144:1, 114

Doshi-Velez, F., and B. Kim, 2017, "Towards a rigorous science of interpretable machine learning," arXiv:1702.08608

Dwyer, F. R., P. H. Schurr, and S. Oh, 1987, "Developing buyer-seller relationships," Journal of marketing, 51:2, 11-27

Fullerton, G., 2005, "The impact of brand commitment on loyalty to retail service brands," Canadian Journal of Administrative Sciences/ Revue Canadienne des Sciences de l'Administration 22:2, 97-110

Garbarino, E., and M. S. Johnson, 1999, "The different roles of satisfaction, trust, and commitment in customer relationships," Journal of marketing 63:2, 70-87

Gouldner, A. W., 1960, "The norm of reciprocity: a preliminary statement," American sociological review, 25:2, 161-178

Gray, C. M., Y. Kou, B. Battles, J. Hoggatt, and A. L. Toombs, 2018, "The dark (patterns) side of ux design," in Proceedings of the 2018 CHI conference on human factors in computing systems, 1-14

Gremler, D. D., and S. W. Brown, 1996, "Service loyalty: its nature, importance, and implications," Advancing service quality: A global perspective 5:1, 171-181

Günther, C-C., I. F. Tvete, K. Aas, G. I. Sandnes, and Ø. Borgan, 2014, "Modelling and predicting customer churn from an insurance company," Scandinavian Actuarial Journal 2014:1, 58-71

Guo, Y., Y. Li, D. Liu, and S. X. Xu, 2024, "Measuring service quality based on customer emotion: an explainable ai approach," Decision Support Systems 176, 114051

Hadden, J., A. Tiwari, R. Roy, and D. Ruta, 2007, "Computer assisted customer churn management: state-of-the-art and future trends," Computers & Operations Research 34:10, 2902-2917

Imani, M., M. Joudaki, A. Beikmohamadi, and H. R. Arabnia, 2025, "Customer churn prediction: a review of recent advances, trends, and challenges in conventional machine learning and deep learning," PrePrints.org

Jones, M. A., D. L. Mothersbaugh, and S. E. Beatty, 2002, "Why customers stay: measuring the underlying dimensions of services switching costs and managing their differential strategic outcomes," Journal of business research 55:6, 441-450

Kabbar, K., and R. Herath, 2025, "Customer churn prediction to enhance customer retention strategies in the banking industry: a study using seven machine learning algorithms,"

Kai-Ineman, D., and A. Tversky, 1979, "Prospect theory: An analysis of decision under risk," Econometrica 47:2, 363-391

Kumar, V., B. Rajan, R. Venkatesan, and J. Lecinski, 2019, "Understanding the role of artificial intelligence in personalized engagement marketing," California Management Review 61:4, 135-155

Lee, M., and J. Woo, 2025, "Tempodegraphnet: predicting user churn using dynamic social graphs and neural odes," PLoS One 20:6, e0321560

Li, Y., and K. Yan, 2025, "Prediction of bank credit customers churn based on machine learning and interpretability analysis," Data Science in Finance and Economics 5:1, 19-34

Li, Y., G. Xia, S. Wang, and Y. Li, 2024, "A deep multimodal autoencoder-decoder frame work for customer churn prediction incorporating chat-GPT," Multimedia Tools and Applications 83:41, 89563-89589

Ljubicic, K., A. Mercep, and Z. Kostanjcar, 2023, "Churn prediction methods based on mutual customer interdependence," Journal of Computational Science 67, 101940

Mamun, M. N. H., 2025, "Advancements in machine learning for customer retention: A systematic literature review of predictive models and churn analysis. Journal of Sustainable Development and Policy," 1:01, 250-284

Martın-Donas, J. M., A. L. Zorrilla, M. deVelasco, J. C. Vasquez-Correa, A. Alvarez, M. I. Torres, P. Delgado, A. Lazpiur, B. Romero, and I. Alkorta, 2024, "Speech emotion recognition for call centers using self-supervised models: A complete pipeline for industrial applications," in Proceedings of the 7th International Conference on Natural Language and Speech Processing (ICNLSP 2024), 119-128

Morgan, R. M., and S. D. Hunt, 1994, "The commitment-trust theory of relationship marketing," Journal of Marketing 58:3, 20-38

Muniz, A. M., and T. C. O'Guinn, 2001, "Brand community," Journal of consumer research 27:4, 412-432

Neslin, S. A., S. Gupta, W. Kamakura, J. Lu, and C. H. Mason, 2006, "Defection detection: measuring and understanding the predictive accuracy of customer churn models," Journal of Marketing Research 43:2, 204-211

Oliver, R. L., 1999, "Whence consumer loyalty?" Journal of Marketing 63(4 suppl1), 33-44

Parasuraman, A., V. A. Zeithaml, and L. Berry, 1988, "Servqual: a multiple-item scale for measuring consumer perceptions of service quality," Journal of Retailing 64:1, 12-40

Pierce, J. L., T. Kostova, and K. T. Dirks, 2001, "Toward a theory of psychological ownership in organizations," Academy of Management Review 26:2, 298-310

Provost, F., and T. Fawcett, 2013, "Data science and its relationship to big data and data-driven decision making," Big Data 1:1, 51-59

Rahman, H., R. P. Afrianto, F. Mohammad, D. A. Tajriyaani, and R. A. N. Azhar, 2024, "IBM telco customer churn prediction with survival analysis," in Proceedings of the International Conference on Advanced Technology and Multidiscipline (ICATAM 2024) 245, 386, Springer Nature

Reichheld, F. F., and W. E. Sasser, 1990, "Zero defections: quality comes to services," Harvard Business Review Sept-Oct, 13

Reichheld, F. F., and P. Schefter, 2000, "E-loyalty: your secret weapon on the web," Harvard Business Review 78:4, 105-113

Ribeiro, M. T., S. Singh, and C. Guestrin, 2016, ""Why should I trust you?" Explaining the predictions of any classifier," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144

Rowland, F., 2011, "The filter bubble: what the internet is hiding from you," portal: Libraries and the Academy, 11:4, 1009-1011

Rudd, D. H., H. Huo, M. R. Islam, and G. Xu, 2023, "Churn prediction via multimodal fusion learning: integrating customer financial literacy, voice, and behavioral data," in 2023 10th International Conference on Behavioural and Social Computing (BESC), IEEE, 1-7

Samuelson, W., and R. Zeckhauser, 1988, "Status quo bias in decision making," Journal of Risk and Uncertainty 1:1, 7-59

Sana, J. K., M. S. Rahman, and M. S. Rahman, 2025, "Privacy-preserving customer churn prediction model in the context of telecommunication industry," Engineering Applications of Artificial Intelligence 162, 112514

Schwartz, H., 2008, "Predictably irrational: the hidden forces that shape our decisions," Business Economics 43:4, 69-72

Shaikhsurab, M. A., and P. Magadum, 2024, "Enhancing customer churn prediction in telecommunications: an adaptive ensemble learning approach," arXiv preprint arXiv:2408.16284

Singh, C., M. K. Dash, R. Sahu, and A. Kumar, 2024, "Artificial intelligence in customer retention: a bibliometric analysis and future research framework," Kybernetes 53:11, 4863-4888

Singh. K., 2024, "Survival-based predictive analytics for customer churn: evaluating classical and neural network models," thesis University of Uppsala

Sunstein, C. R., 2016, The ethics of influence: government in the age of behavioral science, Cambridge University Press

Thaler, R., 1980, "Toward a positive theory of consumer choice," Journal of Economic Behavior and Organization 1:1, 39-60

Thaler, R., 1985, "Mental accounting and consumer choice," Marketing Science 4:3, 199-214

Thaler, R. H., and C. R. Sunstein, 2009, Nudge: improving decisions about health, wealth, and happiness, Penguin

Verbeke, W., D. Martens, C. Mues, and B. Baesens, 2011, "Building comprehensible customer churn prediction models with advanced rule induction techniques," Expert Systems with Applications 38:3, 2354-2364

Verhoef, P. C., P. H. Franses, and J. C. Hoekstra, 2002, "The effect of relational constructs on customer referrals and number of services purchased from a multiservice provider: does age of relationship matter?" Journal of the Academy of Marketing Science 30:3, 202-216

Wang, Y., Z-J. Jin, C-H. Jin, and C. Kan, 2025, "Building customer loyalty through emotional connection: How service provider rapport drives sustainable business," Sustainability 17:6, 2396

Watson, K., and D. Kahneman, D., 2011, "Thinking, fast and slow. New York, NY: Farrar, Straus and Giroux. 499 pages," Canadian Journal of Program Evaluation 26:2, 111-113

White, T. B., D. L. Zahay, H. Thorbjørnsen, and S. Shavitt, 2008, "Getting too personal: reactance to highly personalized email solicitations," Marketing Letters 19:1, 39-50

Xia, L., K. B. Monroe, and J. L. Cox, 2004, "The price is unfair! A conceptual framework of price fairness perceptions," Journal of Marketing 68:4, 1-15

Zeithaml, V. A., 1988, "Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence," Journal of Marketing 52:3, 2-22

Zeithaml, V. A., L. L. Berry, and A. Parasuraman, 1996, "The behavioral consequences of service quality," Journal of Marketing 60:2, 31-46

Zhang, X., D. Mehta, Y. Hu, C. Zhu, D. Darby, Z. Yu, D. Merlo, M. Gresle, A. Van Der Walt, H. Butzkueven, and Z. Ge, 2025, "Adaptive transformer modelling of density function for nonparametric survival analysis," Machine Learning 114:2, 31

How data and AI shape financial services

The case for material ESG data in corporate finance

Rodrigo Tavares,
Invited Full Professor, NOVA School of Business and Economics (Nova SBE)

ABSTRACT

Demand for environmental, social, and governance (ESG) data has grown exponentially over the last decade, but this expansion has also generated measurement noise, weak comparability, and contested interpretations of what ESG is for. This paper argues that ESG integration becomes financially defensible only when it is governed by financial materiality, so that the metrics used in valuation, underwriting, and portfolio decisions have a credible linkage to cash flows, access to finance, or cost of capital. The paper shows that separating enterprise value disclosure from impact/ethical narratives improves decision-usefulness, and it sets out the resulting implications for governance, internal controls, reporting design, investor modelling, and stewardship. For executives, the operating implication is simple, govern material ESG like financial reporting, by selecting a small set of decision-critical ESG indicators, and linking them explicitly to capital allocation, financing strategy, and incentives.

1. Introduction

The proposition that ESG is now a financial imperative is gaining traction partly because it offers a pragmatic reframing of a concept that has become politically charged over the last few years. This is an argument about financial mechanics, not moral preference. Sustainability related risks and opportunities increasingly affect expected cash flows and discount rates, and the information infrastructure around those channels is becoming more standardized, and more usable, by capital markets. This reframing has proceeded amid a visible political backlash in the U.S. associated with President Trump's anti-ESG stance, which has helped shift parts of the policy and market discourse from questions of financial materiality to claims that ESG constitutes ideological interference in capital allocation.

The central thesis advanced here is that ESG integration becomes credible and durable only when it is governed by the same decision rule that governs financial reporting and investment analysis, namely materiality. But what is materiality? In finance, materiality is the threshold that determines whether information is important enough that a reasonable investor's or creditor's decision could change if that information were omitted, misstated, or obscured. In practice, something is financially material when it is reasonably expected to affect enterprise value, meaning it can change expected cash flows, access to finance, or cost of capital over relevant horizons. It is not the same as "socially important" or "ethically significant"; those can be true without being financially material for a particular firm at a particular time. Materiality is also context dependent. The same issue can be material in one industry or company, and immaterial in another, and it can become material as regulation, technology, litigation, or market conditions change.

The practical meaning of this thesis is tangible. Corporations and investors should use only financially material ESG data when the purpose is financial decision-making such as valuation, underwriting, portfolio construction, risk management, capital budgeting, executive incentives, and stewardship aimed at protecting enterprise value. The restriction is not a rhetorical preference for fewer metrics. It is a necessary condition for decision-usefulness because the inclusion of immaterial indicators dilutes informational content, increases false precision, and expands the opportunity set for strategic disclosure.

This article distinguishes impact accountability from financial decision making, and cautions against mixing their objectives within the same disclosure and analytical system. The desire to document social and environmental impacts that may not be priced is legitimate, but it corresponds to a different disclosure function and a different evaluative standard. Illustrative impact-oriented indicators include employee volunteer hours, philanthropic donations, and hectares restored. When impact-oriented indicators are introduced into enterprise value models, as if they were value drivers, the result is often mispricing, politicization, and confusion about fiduciary purpose. Conversely, when enterprise value disclosure is forced to serve all stakeholder audiences simultaneously, reporting volume expands, while decision-usefulness weakens. A materiality discipline, therefore, requires a structural separation between enterprise value disclosure and impact reporting, even when both are produced by the same organization.

2. Materiality and enterprise value as a decision rule

Investor oriented sustainability disclosure guidance has emerged out of a longer corporate reporting trajectory, in which voluntary sustainability reports developed largely outside mainstream financial filings, and gradually moved toward investor decision-usefulness, and standardization. Early corporate sustainability reporting was shaped by multi stakeholder frameworks, most prominently the Global Reporting Initiative, founded in 1997, with guidelines released in 2000, which encouraged broad coverage of environmental and social impacts, and prioritized accountability to a wide set of stakeholders, rather than an explicit enterprise value lens [GRI (2022)].

As sustainability reporting proliferated, capital markets also demanded a more finance facing architecture, reflected in the rise of investor focused initiatives, including the Sustainability Accounting Standards Board, founded in 2011 to develop an industry-based language for the financial impacts of sustainability, and the Task Force on Climate related Financial Disclosures, established by the Financial Stability Board, whose 2017 recommendations sought decision-useful climate risk disclosure structured around governance, strategy,

risk management, and metrics, and designed for inclusion in mainstream financial reporting [SASB (2024), FSB (2017)]. This gradual convergence created the conditions for the IFRS Foundation to establish the International Sustainability Standards Board in 2021, with a mandate to develop a global baseline of high quality sustainability disclosure standards focused on investors' information needs, and to consolidate parts of the fragmented disclosure landscape [IFRS Foundation (2021)].

Within this institutional arc, IFRS S1, and IFRS S2, issued by the ISSB in June 2023, represent a decisive step toward embedding sustainability related disclosures within the logic of general-purpose financial reporting, and toward making corporate disclosures more comparable, and more usable, for valuation, credit analysis, and capital allocation [IFRS Foundation (2023a)]. IFRS S1 sets the general requirements for disclosure of sustainability related financial information, centered on providing information useful to primary users, namely existing and potential investors, lenders, and other creditors, in decisions about providing resources to the entity, and it anchors relevance in the entity's ability to generate cash flows over the short, medium, and long term [IFRS Foundation (2023b)]. IFRS S2 is the climate specific companion standard, building on S1, and requiring detailed disclosures on climate related risks, and opportunities, including governance, strategy, risk management, and metrics and targets, to support consistent pricing of transition, and physical, risks in capital markets [IFRS Foundation (2023b)]. The importance of this architecture for corporate disclosures lies in its attempt to shift sustainability reporting away from expansive narrative volume, and toward investment grade information, defined by materiality, comparability, and connectivity with financial reporting, thereby strengthening trust, and reducing the scope for selective disclosure.

The ISSB educational material, published in November 2024, further operationalizes this shift by clarifying how materiality judgements should be made under IFRS S1, and IFRS S2, defining material information as information whose omission, misstatement, or obscuring could reasonably be expected to influence the decisions of primary users, and emphasizing that disclosures should focus on sustainability related risks, and opportunities, that could reasonably be expected to affect cash flows, access to finance, or cost of capital, over relevant horizons, while warning against obscuring material information within large volumes of additional stakeholder oriented content [IFRS Foundation (2024)]. The conceptual implication is that sustainability information becomes financially relevant, not because it is ethically compelling, but because it changes the distribution of future economic outcomes for the entity and, therefore, belongs within the evidentiary discipline of corporate financial disclosure, rather than within a general-purpose narrative of corporate virtue.

It is also important to note that materiality is inherently entity-specific and industry-specific. The same indicator can be material for one business model and immaterial for another because the mechanism linking the indicator to cash flows or discount rates differs. Materiality is also dynamic because regulation technology litigation and social preferences evolve and can change what is priced. The materiality discipline proposed here accommodates emerging issues through continuous horizon scanning, and scenario analysis, while limiting the variables used in valuation, underwriting, and portfolio decisions to those with a defensible linkage to enterprise value at the time of decision.

A strict materiality doctrine also reduces ambiguity about what ESG is meant to accomplish in finance. If the purpose is enterprise value, then the relevant question becomes whether an ESG metric maps into a valuation channel, such as expected operating margins, growth, required investment, working capital, risk of disruption, legal costs, insurance costs, financing spreads, tail risk, or terminal value assumptions. If such a mapping cannot be articulated and defended, then the metric may still be meaningful for impact accountability or stakeholder dialogue, but it should not be treated as an investment grade input.

3. The problem of immaterial ESG data measurement noise and incentives

The growth of ESG datasets has created an environment in which analysts can easily confuse breadth with insight. Many corporate reports and ESG vendor products aggregate large sets of indicators into summary scores that appear quantitative and comparable, but embed subjective decisions about inclusion, weighting, and the treatment of missing values. Consistent with this concern, Benuzzi et al. (2025) show that methodological design choices in ESG score construction can drive a substantial share of observed score variation, independently of underlying corporate disclosure. A principal consequence is informational dilution. When models average financially relevant and financially irrelevant information, the composite can become a weak predictor of financial outcomes, even if some underlying components contain decision-useful content. A second consequence is false precision. Scores often compress heterogeneous constructs into single rankings and thereby invite over interpretation. A third consequence is incentive distortion. If firms are evaluated on broad composites, they may allocate resources toward improving measurable, but immaterial, indicators, rather than toward managing the sustainability related risks that most affect resilience and cost of capital. These dynamics contribute to greenwashing, not primarily through outright falsification, but through selective emphasis and the substitution of immaterial disclosures for material performance improvements.

Recent empirical evidence supports the proposition that materiality filtering matters. Ben Amor and Kooli (2025 examine Canadian firms and report a positive association between a firm level ESG materiality score and financial performance while immaterial ESG issues show either no significant effect or a negative relation with performance. The structure of this result is important because it suggests that treating all ESG issues as equally relevant is analytically wrong and that the performance association is concentrated in the subset of issues that are material in the enterprise value sense. Yang et al. (2025) construct a SASB aligned ESG score and find that it predicts future financial performance and is associated with lower risk while a broad ESG disclosure score fails to predict performance or risk outcomes. This again suggests that disclosure volume alone is not the mechanism and that the content and materiality alignment of disclosure matter.

At the same time, the empirical literature warns against attributing portfolio outperformance to ESG signals without careful control for fundamentals. Ahn et al. (2024) replicate evidence of outperformance for a portfolio of firms with improving ratings on material ESG issues but provide evidence that the performance can be explained by exposure to profitability and growth factors and that alpha dissipates once fundamental determinants are comprehensively controlled. The implication is not that material ESG issues are irrelevant, but that material ESG scores may proxy for fundamentals and that the evidentiary bar for claiming incremental ESG driven alpha is high. This is consistent with the broader argument that ESG becomes financially credible when it is treated with the same econometric skepticism and benchmarking discipline as that applied to other candidate factors.

4. Reporting architectures, and the limits of disclosure volume

The movement toward investor-oriented sustainability disclosure guidance increases the feasibility of a materiality-centered approach. Evidence from event studies highlights why disclosure volume without materiality discipline may not move markets. Haley et al. (2025) examine market reactions to announcements of corporate sustainability reports that incorporate SASB metrics intended to be financially material and find limited evidence that these reports provide a significant amount of new information to investors. They further report that common metrics may be financially immaterial in magnitude or preempted by traditional financial disclosures and that many firms target sustainability reports at a broad set of stakeholders rather than a narrow set of financially-oriented investors. This evidence reinforces two claims advanced here. First, the mere inclusion of ostensibly material metrics is insufficient, if the information is not timely, decision relevant, and incremental, relative to existing disclosures. Second, mixed audience

reporting can weaken the financial information function, because narrative, and breadth, can dominate, at the expense of analytic clarity.

A strict materiality doctrine, therefore, implies an architectural separation. Firms can and often should publish broader impact reporting, but sustainability-related financial disclosures used for enterprise value decisions should be clearly identifiable, governed by internal controls, and integrated into the reporting ecosystem in a way that supports connectivity with financial statements and management discussion. The objective is not minimal disclosure but decision-useful disclosure with verifiable metrics and explicit linkages to value drivers.

5. Implications for corporate governance, internal controls, and metric selection

The corporate implication of strict financial materiality is that the core task is not to maximize the number of ESG indicators reported, but to maximize the reliability, comparability, and decision-usefulness of the subset of sustainability metrics that are material for the entity. This requires a disciplined materiality assessment, rooted in the business model. The assessment begins by identifying how sustainability-related conditions map into revenue durability, cost structure, capital intensity, asset life, legal exposure, and financing conditions. It then selects a limited set of indicators whose measurement boundaries and estimation methods are defensible, and whose variation plausibly affects enterprise value. The emphasis on limited sets is not aesthetic. It is necessary, because measurement quality is costly. Concentrating on fewer material indicators enables investment in data governance systems that create audit trails, consistent definitions, reliable time series, and assurance pathways.

Operationally, this discipline benefits from a workflow that begins with an industry-based materiality map, tests each candidate metric against explicit valuation channels, and then hardens definitions, boundaries, and assurance. Platforms such as Finland-based Upright can accelerate initial screening, and peer benchmarking, by translating business activities into structured sustainability profiles. They can also map ESG exposures to financial materiality through expected cash flow drivers, such as revenue durability, cost structure, capex needs, and risk premia. Yet, the resulting outputs still require management judgement, and explicit integration into budgeting, valuation, and remuneration systems.

The materiality doctrine also implies that if a metric is financially material, it should be treated operationally as a financial metric. That entails clear ownership, internal controls, documented methodologies, change management for definitions and boundaries, and board oversight focused on the integrity and decision relevance of the metrics. The ISSB educational material explicitly links materiality to the needs of primary users, and warns against obscuring material disclosures with additional information, implying that governance must enforce focus and clarity [IFRS Foundation (2024)]. Boards should, therefore, resist metric sprawl, and ask whether disclosed indicators change capital allocation decisions, risk tolerances, or financing strategy, and whether management has constructed credible causal narratives connecting the indicators to financial outcomes.

Strict materiality also clarifies how to handle emerging issues. Firms should scan broadly and develop scenario analysis for issues that could plausibly become material under regulatory, technological, or social change. But until the expected effect on cash flows or cost of capital becomes defensible, such issues should be discussed as uncertainties, and monitored, rather than embedded as mechanical inputs into valuation and incentive systems.

6. Implications for investors modelling, underwriting, and stewardship

For investors, the strict materiality doctrine shifts ESG integration away from generic scoring and toward underwriting of sustainability-related value drivers. The empirical evidence suggests that materiality aligned measures contain more predictive content than broad disclosure measures for performance and risk [Yang et al. (2025)] and that financial performance attributed to material ESG improvement can be explained by exposures to

fundamental factors [Ahn et al. (2024)]. Together, these findings imply that investors should treat ESG indicators as candidate signals that must pass the same tests of incremental predictive power robustness and interpretability as those applied to other financial variables. It also implies that analysts should explicitly specify whether a sustainability metric enters the model through expected cash flows or through risk premia, and should avoid double counting the same risk in multiple places.

In practice, investment teams can operationalize this by combining issuer disclosures with independent, activity-based datasets, using tools such as the Upright Platform for rapid cross-sectional screening, while enforcing a strict filter that admits only indicators with a defensible effect on cash flows, downside risk, or cost of capital, for the relevant sector, and issuer.

In equity valuation, a material sustainability metric should affect either forecasted operating outcomes, or discount rates, depending on the channel. If the metric captures regulatory costs, operational disruption probability, insurance costs, or litigation exposure, the effect may be primarily on expected cash flows, and downside distributions. If the metric captures uncertainty, tail risk, or financing constraints, it may affect the discount rate, or scenario weights. The materiality discipline requires that such linkages be explicit, and evidence-based, rather than narrative. Again, the analytic question is not whether a topic is socially salient, but whether it affects covenant headroom, liquidity resilience, collateral quality, and refinancing risk.

The same logic applies to credit. Using European repo transactions from 2019 to 2022, Giuzio et al. (2026) find that banks with higher financed emis-

sions pay higher repo borrowing rates, with a one standard deviation increase associated with roughly 7% to 12% higher rates, consistent with transition risk premia being priced even in core funding markets. At the borrower level, Carroll et al. (2025) show that Irish firms located in flood risk areas face higher loan rates, by roughly 7 to 13 basis points, and are more likely to post collateral, indicating that physical risk affects credit conditions, and, therefore, expected cash flows, and financing constraints. In a global syndicated loan sample, Kempa (2026) reports that higher climate vulnerability is associated with higher borrowing costs, reinforcing that physical climate risk can enter discount rates, and credit spreads, in a manner consistent with strict financial materiality.

In stewardship, strict materiality encourages a narrower, and more effective, engagement agenda. If engagement is intended to protect enterprise value, it should focus on governance, risk management, and disclosure of the material sustainability drivers for the sector, and firm. The evidence that sustainability reports may be oriented toward broad stakeholder audiences, and may contain limited incremental investor information, suggests that investors who want decision-useful disclosure should clearly communicate what they need, and should reward timeliness, connectivity, and verifiability [Haley et al. (2025)]. Engagement built around long checklists of immaterial indicators risks becoming performative, and incentivizes disclosure management, rather than risk management.

.

7. Mandatory disclosure and information environment effects

Disclosure regimes shape information environments, and can, therefore, influence liquidity, and price discovery. Krueger et al. (2024) compile a global dataset on mandatory ESG disclosure, and document positive effects on firm level stock liquidity, and show that effects are stronger when requirements are implemented by government institutions and coupled with stronger enforcement. This finding supports the view that disclosure can improve information environments and generate capital market benefits. It also strengthens the case for materiality discipline, because the benefits of disclosure are likely to be larger when the disclosed information is comparable, enforceable, and focused on decision-useful content. This is particularly salient in sectors where a small set of indicators is recurrently material, such as electric utilities where emissions intensity per MWh and generation mix map into regulated investment needs and compliance costs, banks where financed emissions and sectoral credit exposure map into expected losses and capital requirements, and mining where tailings facility risk and remediation liabilities map into operational continuity and contingent costs. If mandatory regimes are met with floods of immaterial indicators, the intended improvement in comparability, and price discovery, may be weakened by information overload, and strategic obfuscation. The ISSB educational material explicitly recognizes the risk of obscuring material information, and, therefore, indirectly supports a reporting design that privileges clarity, and focus [IFRS Foundation (2024)].

8. Discussion legitimacy and the depoliticization of ESG in finance

The strict financial materiality doctrine also has a political economy advantage, because it clarifies fiduciary purpose. A frequent critique of ESG is that it smuggles contested social preferences into capital allocation. When investors rely on metrics untethered to enterprise value, that critique becomes difficult to rebut. When investors rely only on financially material sustainability information, the justification becomes an extension of standard risk and return analysis. The relevant disagreements shift from ideology to empirical questions about mechanisms, magnitudes, horizons, and model specification. This shift is compatible with the claim that ESG becomes a financial imperative, not through moral persuasion, but through the maturation of standards, and analytics, that make sustainability data usable in the language of finance [Tavares (2025)].

At the same time, strict materiality can improve corporate behavior by changing incentives. If markets reward broad score maximization, firms have incentives to optimize for optics. If markets reward credible management, and disclosure, of material sustainability risks, firms have incentives to invest in resilience, and risk mitigation. This reduces the space for greenwashing, through substitution of

immaterial disclosures for material improvements. It also protects impact reporting from being dismissed as irrelevant, simply because it is not immediately priced. By separating enterprise value disclosure from impact narratives, the doctrine allows each to be assessed against its appropriate objective.

9. Conclusion

This article has argued that ESG becomes financially credible only when governed by strict financial materiality. Investor-oriented sustainability disclosure guidance emphasizes that sustainability related financial disclosures should provide material information, useful to investors, lenders, and other creditors, and that material information should not be obscured by additional stakeholder-oriented information [IFRS Foundation (2024)]. Recent empirical evidence suggests that materiality-aligned ESG measures predict financial outcomes more reliably than broad disclosure measures [Ben Amor and Kooli (2025), Yang et al. (2025)], while also cautioning that apparent ESG portfolio outperformance can reflect exposures to fundamentals, rather than incremental ESG information [Ahn et al. (2024)]. Evidence from sustainability report announcement studies suggests that even reports incorporating financially-oriented metrics may provide limited incremental information to investors and may be targeted to broader stakeholder audiences [Haley et al. (2025)]. Evidence from mandatory disclosure regimes suggests that enforced disclosure can improve information environments, and liquidity [Krueger et al. (2024)], which further motivates a focus on decision-useful material content, rather than reporting volume.

The practical implication is clear. Corporations should concentrate measurement resources on a limited set of financially material sustainability drivers, governed by internal controls, and connected to financial reporting, and strategy. Investors should replace generic ESG scoring with underwriting of material sustainability value drivers and apply econometric discipline to claims of ESG alpha. Stewardship should focus on governance, and disclosure, of material metrics, rather than exhaustive checklists. In this way, ESG can be understood, not as an ideological overlay, but as a disciplined extension of financial analysis of risks, and opportunities, that affect enterprise value.

References

Ahn, B. H., P. N. Patatoukas, and G. S. Skiadopoulos, 2024, "Material ESG alpha: a fundamentals based perspective," The Accounting Review 99:4, 1-27

Ben Amor, S., and M. Kooli, 2025, "Does materiality matter in ESG investing," Research in International Business and Finance 79

Benuzzi, M., K. Bax, S. Paterlini, and E. Taufer, 2025, "Chasing ESG performance: how methodologies shape outcomes," International Review of Financial Analysis 104

Carroll, J., M. Mahony, B. Morando, C. O'Sullivan, and S. Shahabi Ahangarkolaee, 2025, "Firm credit conditions and flood risk: Evidence from Ireland," The Economic and Social Review 56:4, 449-483

FSB, 2017, "Recommendations of the Task Force on Climate-related Financial Disclosures," Financial Stability Board, https://tinyurl.com/zpjwdtmp

Giuzio, M., B. Kahraman, and J. Knyphausen, 2026, "Climate change, bank liquidity and systemic risk," European Central Bank working paper series no. 3168, January

GRI, 2022, "25 years as the catalyst for a sustainable future 1997–2022," Global Reporting Initiative, https://tinyurl.com/3bv582w8

Haley, S., M. Shaffer, and R. Sloan, 2025, "Do sustainability reports contain financially material information," Review of Accounting Studies, 1-36

IFRS Foundation, 2021, "IFRS Foundation announces International Sustainability Standards Board, consolidation with CDSB and VRF, and publication of prototypes," November 3,

IFRS Foundation, 2023a, "ISSB issues inaugural global sustainability disclosure standards," June 26

IFRS Foundation, 2023b, "IFRS S1 General Requirements for Disclosure of Sustainability-related Financial Information,"

IFRS Foundation, 2024, "Sustainability related risks and opportunities and the disclosure of material information," Educational material, November

Kempa, K., 2026, "Physical climate risk and the pricing of bank loans," Journal of Environmental Economics and Management, 137

Krueger, P., Z. Sautner, D. Tang, and R. Zhong, 2024, "The effects of mandatory ESG disclosure around the world," Journal of Accounting Research 62:5, 1795-1847

SASB, 2024, "About SASB" IFRS Foundation, https://tinyurl.com/mpkjr7b6

Tavares, R., 2025, "Why ESG is now a financial imperative," World Economic Forum, March

Upright Project, 2026, "Upright Platform, model and methodology," https://www.uprightproject.com/

Yang, S. S., J. W. Huang, H. Y. Chen, and M. H. Tsay, 2025, "Detecting corporate ESG performance: the role of ESG materiality in corporate financial performance and risks," The North American Journal of Economics and Finance, 76

How data and AI shape financial services

Tokenization and financial market inefficiencies*

Itai Agur,
Senior Economist, Macro-Financial Division, Research Department, International Monetary Fund (IMF)

Germán Villegas-Bauer,
Economist, Macro-Financial Division, Research Department, International Monetary Fund (IMF)

Tommaso Mancini-Griffoli,
Assistant Director and Chief of Payments, Currencies, and Infrastructure Division, Monetary and Capital Markets Department, International Monetary Fund (IMF)

Maria Soledad Martinez Peria,
Assistant Director and Chief of Macro-Financial Division, Research Department, International Monetary Fund (IMF)

Brandon Joel Tan,
Economist, Payments, Currencies, and Infrastructure Division, Monetary and Capital Markets Department, International Monetary Fund (IMF)

ABSTRACT

This paper provides a conceptual framework centered on market inefficiencies to investigate the potential implications of the tokenization of financial markets. Tokenization is the creation of assets or asset representations on a digital ledger that is shared among, and trusted by, market participants and that enables self-executing contracts to be programmed onto it. Such a ledger could retire asset record keeping services, reducing costs related to asset issuance, servicing, and redemption. Tokenization may also decrease asset trading costs by addressing certain forms of counterparty risk more efficiently, speeding up settlement cycles, and lowering search frictions on specialized markets. If tokens are on-ledger assets rather than asset representations, asset custodianship costs could be saved too. Tokenization can also affect shock transmission externalities by influencing financial institutions' funding liquidity and leveraging incentives, increasing interconnectedness, and automating trading. Potential impacts on externalities arising from networks, innovation, and market infrastructure are discussed as well. Additionally, shared ledgers can lower broker-switching costs, harnessing competition to lower investors' fees. However, if tokenization facilitates direct trading without brokers, this could amplify internalities from limited investor understanding of risks.

* The views expressed are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

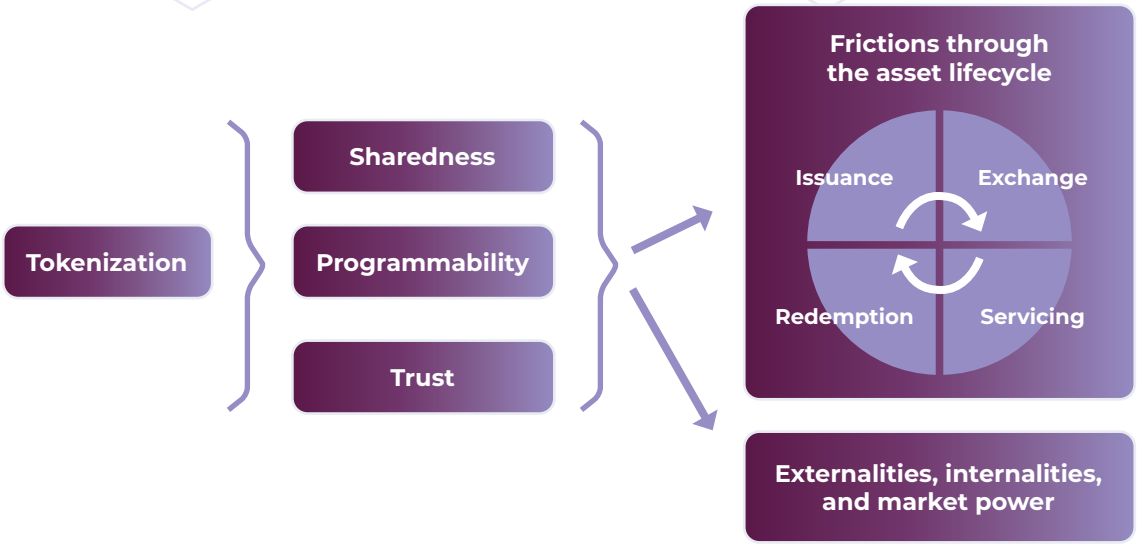
1. Introduction

Recent advances in digital technologies are transforming financial markets. Improvements in computer hardware, networking infrastructure, and software development have facilitated new financial activities such as high-frequency trading [Brogaard et al. (2014), Jacob Leal et al. (2016)], mobile banking and payments [Shaikh and Karjaluoto (2015), Tam and Oliveira (2017)], and crypto-asset issuance and trading [IMF (2023)]. Tokenization is considered one of the recent financial innovations facilitated by technological change [Aldasoro et al. (2023), FSB (2023), Banerjee et al. (2023)]. Several government-sponsored initiatives, involving central banks and private financial institutions, have been recently announced or are underway to test its potential effects.[1] The Boston Consulting Group estimates that, by 2030, asset tokenization could reach U.S.$16 trillion or 10 percent of global GDP [Kumar et al. (2022)]. Given the significant potential growth of asset tokenization in the coming years, it is key to analyze its potential effects, and determine which specific features of tokenization give rise to these effects.

We present a conceptual framework rooted in economic first principles to evaluate the potential effect of tokenization on market inefficiencies. Inefficiencies are grouped into two broad categories (Figure 1). The first is frictions, which include information asymmetries, search problems, transaction costs, and counterparty risks. The effects on financial markets frictions can be usefully analyzed along the lifecycle of an asset (issuance, exchange, servicing, and redemption). The second category includes externalities, internalities, and market power.[2] We take existing market structures (including the role of intermediaries) and regulation as given.

Tokenization refers to the creation of assets or representations of assets (known as "tokens") on a digital ledger that is shared, trusted, and programmable. Sharedness refers to the capacity of transacting parties to possess, acquire, and transfer assets on the ledger. Trust depends on the accuracy of asset ownership and the predictability of transaction orders. Trust is needed for agents to willingly transact on the ledger. Programmability means that assets, the financial applications associated with them - such as repos and swaps -

Figure 1: Overview of conceptual framework on tokenization and market inefficiencies



1. On April 3, 2024, the Bank for International Settlements (BIS), together with seven central banks (Bank of France, Bank of Japan, Bank of Korea, Bank of Mexico, Swiss National Bank, Bank of England, and the Federal Reserve Bank of New York), announced plans to join forces with a large group of private financial firms convened by the Institute of International Finance to explore how tokenized commercial bank deposits can be integrated with tokenized wholesale central bank money in a public–private programmable core financial platform (Project Agorá) [BIS (2024)]. Project Global Layer 1 and Drex are initiatives involving, respectively, the Monetary Authority of Singapore and the Central Bank of Brazil, aiming to facilitate the issuance, redemption, and transfer of tokenized financial assets. The Swiss National Bank (SNB) has been evaluating tokenized securities settlement in Swiss franc wholesale central bank digital currency for financial market participants (Helvetia Pilot) [SDX (2024)].
2. The effects of tokenization on this set of inefficiencies have repercussions along multiple stages of the lifecycle, which are interrelated. For this reason, it is more appropriate to discuss their effects throughout the entire lifecycle in subsections specific to each inefficiency.

and the conditions for transactions, including the investors permitted to hold certain assets, can all be hard coded into, and executed by, the ledger. Moreover, transactions can be contingent on one another and bundled to occur as one. We note that the features of token ledgers can be achieved to varying degrees; they are continuous rather than dichotomic. One ledger may exhibit greater sharedness than another or may support a broader set of programs.

The effects of tokenization on financial market inefficiencies emanate primarily from improvements in the sharedness and programmability features of ledgers. Trust is a necessary condition for their utilization. More shared and programmable ledgers have the potential to reduce market frictions in asset issuance, trading, servicing, and redemption. For example, such ledgers could mitigate certain forms of counterparty risk through simultaneous settlement, enable faster settlement, and reduce search frictions. More shared and programmable ledgers may also affect the degree and costs associated with externalities, internalities, and market power. In particular, they may facilitate the spread of shocks across financial institutions, increase the cost of operational risk events, improve liquidity, facilitate positive knowledge externalities, exacerbate retail investor internalities, and affect market power. While tokenization may, in some instances, lessen the need for certain intermediaries, it is more likely to reduce the costs of the processes performed by intermediaries rather than rendering them obsolete. How much of these cost savings will be passed on to investors will depend on the degree of market competition.

Increases in sharedness are not unique to tokenization and thus some of the implications for financial markets analyzed in this paper may apply to changes to other (nonprogrammable) ledgers as well. Unifying disparate ledgers increases sharedness, as more agents come to share the same trading venue. However, such unification is not necessarily predicated on tokenization: that is, the resulting unified ledger need not be programmable. This underlines the importance of considering tokenization not as a singular innovation but rather as a package composed of distinct dimensions, some of which can be attained by other means as well. Arguably, the presence of some inefficiencies in financial markets originates from the failure to upgrade legacy systems or market structures, because of, for example, coordination problems, vested interests, or political economy constraints. To the extent that tokenization attains what could have been attained through other means (namely, increases in sharedness), its purpose in current

Table 1: Estimated quantitative benefits from tokenization

Lifecycle category	Article	Benefit topic	Improvement
Issuance	Leung et al. (2023)	Bond underwriting fees	-25.80%
	Aldasoro et al. (2025)	Bond underwriting fees	0.00%*
	Liu et al. (2023)	ABS yield spread	-9.52%
Exchange	Leung et al. (2023)	Bond transaction costs	-5.30%
	Aldasoro et al. (2025)	Bond transaction costs	-36.67%
	Plepi and Schwendner (2024)	Bond transaction costs	0.00%
	Tewes et al. (2023)	Transaction costs	-43.00%
	Van Gysegem and De Patoul (2021)	Equity transaction costs	-22.02%
	Onyx (J.P. Morgan) and Apollo (2023)	Portfolio management fees	-22.00%
Servicing	Van Gysegem and De Patoul (2021)	Costs	-38.46%

Note: * The estimated change is not statistically different from zero.

policy initiatives could be more as a coordination device (to push through reforms) than a genuine technological innovation. Regarding advances in programmability, however, the "package of tokenization" contains more novelty.

We study the potential effects of tokenization across a range of financial markets, for some of which several recent empirical studies estimate their potential gains. Based on a global sample, Leung et al. (2023) estimate that tokenizing bonds can alleviate issuance and trading frictions: underwriting fees and bid-ask spreads can fall by 25.8 and 5.3 percent, respectively. Utilizing a similar approach, Aldasoro et al. (2025) estimate that bond tokenization can reduce bid-ask spreads by around 37 percent, while they find no significant effect on underwriting fees. On the other hand, Plepi and Schwendner (2024) do not find reductions in transaction costs from bond tokenization in Switzerland. Studying asset-backed securities (ABS), Liu et al. (2023) estimate that tokenization lowers issuance costs: yield spreads decline by an average of 25 basis points. Onyx (J.P. Morgan) and Apollo (2023) estimate that the automated portfolio deployment of cash, facilitated by programmability, could lead to a 24-basis point reduction in portfolio management fees to investors. Private sector estimates also include Tewes et al. (2023) from Porsche Consulting, who expect tokenization to result in a 43 percent reduction in asset transaction costs; and Van Gysegem and De Patoul (2021) from Roland Berger, who predict equity tokenization to result in declines of 22 and 38 percent in transaction and servicing costs, respectively. The main findings from these studies are summarized in Table 1.

A number of studies have also conceptually analyzed the potential effects of tokenization across a set of financial markets [OECD (2020), Aldasoro et al. (2023), CPMI (2024)].[3] Carapella and others (2023) and FSB (2024) discuss potential financial stability implications. We complement these studies by evaluating the potential effects of tokenization employing a different conceptual framework, which centers on financial market inefficiencies. We consider how these inefficiencies are currently mitigated or exacerbated by existing financial market intermediaries. Our approach then highlights the current challenges in addressing inefficiencies and identifies which changes to financial market infrastructures have the potential to better address them. Additionally, we contribute by linking each potential effect to the specific features that define tokenization. As mentioned above, this facilitates the understanding of which potential benefits can be achieved through other (nonprogrammable) ledgers as well.

2. What is tokenization?

A digital token is an asset or a representation of an asset on a digital ledger that is shared, trusted, and programmable.

To transfer tokens on a ledger, the ledger must be shared, which means that transacting parties can own assets recorded on the ledger and instruct the ledger to update ownership. That is, asset ownership rights are embedded and transferable. For the purposes of this paper, separate but interoperable ledgers are equivalent to a ledger to which all transacting parties have access.

To be willing to use the ledger, transacting parties need to trust it for ownership and order purposes. Trust in ownership stems from a token not being transferable without its owner's consent and, once transferred with such consent, the recipients' confidence that ownership is lasting and cannot be suddenly reversed.[4,5] Trust in orders means that instructions on the ledger will be predictably executed, and that confirmation will be accurate and truthful, as well as final.

3 The definition of tokenization used in this paper is akin to that considered in studies by Aldasoro and others (2023) at the Bank for International Settlements, the Committee on Payments and Market Infrastructures [CPMI (2024)], and the Financial Stability Board [FSB (2024)]. Similar to this research, we discuss distributed ledger technology (DLT) as one (prominent) technological implementation for tokenization, but not as part of what defines a token. This stands in contrast to various other studies, including in the private sector, which tie distributed ledgers to the definition of digital tokens [Banerjee et al. (2023), Carapella et al. (2023)]. Our focus on the features of tokenization, which can be implemented in different models (Subsection 2.A), guides the choice for a technology-neutral definition. Although we describe different models of tokenization, the analysis of the effects on market inefficiencies is not specific to a particular model, unless stated explicitly.

4. The reversal of illicit and fraudulent transactions can constitute an exception to this: a ledger that is overseen by an authority with a capacity to reverse illicit/fraudulent transactions (and with a credible ability to identify cases of fraud, money laundering, and other financial crimes, and only reverse transactions in such cases) may be trusted to a larger degree by bona fide users than a ledger where owners have the sole authority over their assets. Currently, trading platforms possess quasi-regulatory powers to oversee trading activity, as they maintain detailed rulebooks for trading, market abuse monitoring, and risk management systems. These platforms are typically overseen by a regulator themselves.

5. For tokens that are representations of assets outside the ledger, ownership trust also depends on confidence in the parity between the off-ledger asset and its on-ledger representation.

Programmability means that the ledger stores code-based instructions (smart contracts) that can be used to create an asset or financial application (also called a protocol). For instance, a smart contract can be used to create a transferable bond that pays coupons, or it can be used to escrow and exchange assets. Importantly, smart contracts can include rules that are automatically enforced every time an asset is transferred, such as to comply with regulations (for example, only allowing qualified investors to transact riskier assets). Furthermore, applications can be combined and reused (referred to as composability). Multiple steps to a transaction, such as receiving one payment before initiating another, can be executed in a single inseparable transaction (known as atomicity).

An example of a digital token currently in use is Ether, deployed on Ethereum, a network that includes a distributed ledger that is shared and trusted by its users and is programmable [Agur et al. (2023)]. The ledger contains assets other than Ether, can be programmed to ensure simultaneous exchanges, and offers "composability."

The creation of assets or representations of assets on a digital ledger that is shared, programmable, and trusted is called tokenization. Tokenization can take place in different forms [Lavayssière (2023)]. Digital tokens can be created through the issuance of new tokens directly on a shared, programmable, and trusted ledger. This paper calls such tokens "native" because they exist solely on the ledger. Digital tokens can also be created as virtual representations of existing assets, such as financial assets held by a custodian. We refer to such tokens as "non-native", since they are representations of assets that exist outside the ledger. "Migration" refers to the case where a non-native token and the off-ledger asset that it represents are replaced by a native token on the ledger [Lavayssière (2023)]. A non-native token could potentially represent an asset that is a native token on another ledger. Digital tokens can, therefore, migrate from one ledger to another. Table 2 summarizes our definition of tokenization and the forms by which it can be attained.

The features that characterize the ledgers where digital tokens are created are not dichotomic variables: sharedness, programmability, and trust can be attained to different degrees. A ledger is considered shared once it has more than one user. Its degree of sharedness increases as more users join, either through onboarding or through merging or attaining interoperability with previously separate ledgers. Similarly, ledgers may support a broader or narrower set of programs. Trust can also be attained to different degrees and a ledger that engenders more trust is likely to attract a larger number of users.[6] As the three features are part of a spectrum, there can be incentives to migrate digital tokens from one ledger to another that enables more sharedness, programmability, and/or trust.

2.1 Models to attain sharedness, programmability, and trust

There are different models for tokenization, that is, models which can attain sharedness, programmability, and trust on a digital ledger.

There are three main models to obtain sharedness: (1) a single ledger, (2) a common ledger, and (3) compatible ledgers. Model (1) entails issuing (native) assets on a ledger on which all (potential) transacting parties (or their intermediaries) have accounts. Model (2) entails creating representations of assets, initially existing on separate ledgers, on a ledger accessible to all parties (non-native assets). This may be accomplished by placing the assets in escrow and requesting the common ledger operator (or third party) to issue escrow certificates that can be owned and transferred among the ledger's participants. Model (3) involves standardizing tech-

Table 2: What is tokenization?

Tokenization	**Token creation**
Creation of assets or representations of assets on a digital ledger that is: → Shared → Programmable → Trusted	Can be attained through: → New issuance (native) → New virtual representation of an existing asset (non-native) → Destruction of an existing asset and reissuance on a ledger (migration)

6. This highlights how the features are interrelated, as higher levels of trust would lead to increased sharedness.

nology across ledgers so they can run the same programs (smart contracts) and thus offer coordinated transfers (that is, making ledgers interoperable) [Mancini-Griffoli et al. (2024)].

Programmability on token ledgers differs from more traditional forms of programmability. Traditional financial systems separate value from logic. For example, bank balances are entries in a ledger/database, while logic - payment instructions, compliance checks, and business rules - resides in outside systems, scripts, and human processes. Token ledgers are built to integrate both value and logic within the ledger itself. Code (smart contracts) and data (the tokens/transactions) are tightly integrated: the logic that governs behavior sits on the same platform that transfers value. This removes the need for manual initiation, verification, or approvals. Merging the code with the database (the ledger that records transactions) facilitates a greater degree of programmability because it guarantees coherence between them, increasing the efficiency of piecing together transactions into smart contracts that are bundled or contingent on other transactions [Lavayssière (2023)].

The models to achieve trust can be classified into three categories: single-operator systems, permissionless DLT, and permissioned DLT. These models achieve consensus on the state of the ledger - ownership of assets - and its programming functions through different mechanisms.[7] Single-operator systems are centralized in the sense that a single entity is responsible to keep records of ownership, execute transactions, and confirm them. Trust in these systems stems from a mix of credibility of the ledger operators, supervision, and legal recourse in case of any wrongdoing. Permissionless DLT relies on a decentralized network of nodes to verify and validate transactions. Information on the state and programming of the ledger is shared across the network; hence no single operator can unilaterally alter the state or programming functions of the ledger. Whether such decentralization of governance allows users to build trust at a lower cost remains a topic of debate.[8] Validator nodes are incentivized to propose valid transactions through rewards based on newly minted coins and transaction fees. Bitcoin and Ethereum are examples of ledgers based on permissionless DLT. Permissioned DLT is a hybrid between single-operator and permissionless DLT systems, wherein the ledger's validators are onboarded by a central authority.[9]

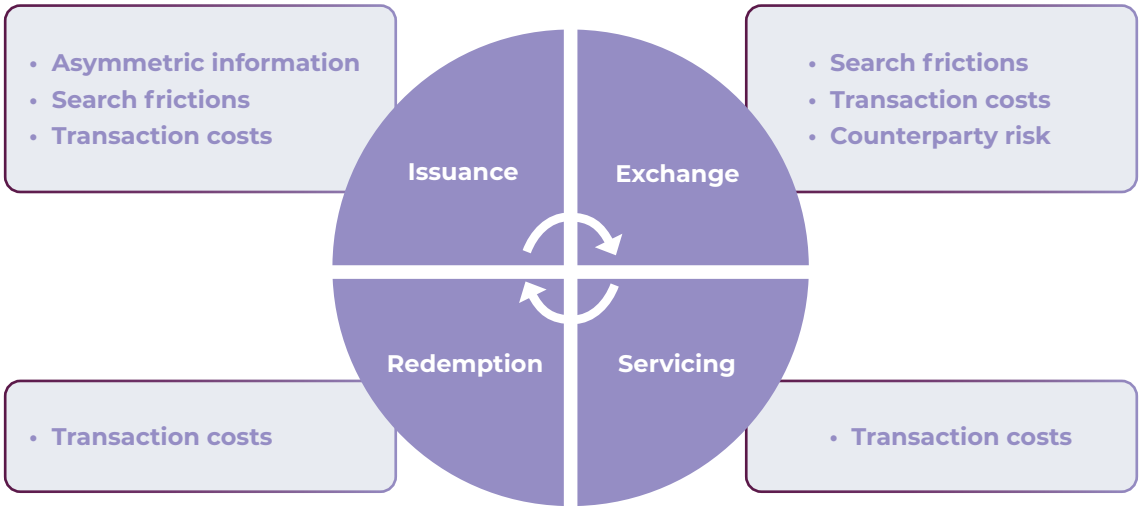
3. Inefficiencies in financial asset markets and the role of intermediaries

This section provides an overview of the types of inefficiencies common in financial markets and the role that financial intermediaries play in addressing some of these. The aim of this section is to set the stage for subsequent sections' examination of how tokenization may affect these inefficiencies in comparison to currently prevailing market structures.

The lifecycle of a financial asset that is traded on a market can be categorized by the processes involved in its issuance, exchange, servicing, and redemption. Issuance is the process of creating and selling new financial assets, such as stocks and bonds, to investors. The issuer may sell these assets to investors through various mechanisms such as initial public offerings (IPOs) or private placements. Exchange refers to the buying and selling of financial assets on the market. Servicing refers to the payment of interest or dividend on financial assets. Redemption is the process of returning these assets to the issuer in exchange for cash or other assets.[10]

7. In DLT systems, such a mechanism is known as a "consensus mechanism", which is an algorithm by which various parties reach an agreement. A consensus mechanism allows nodes (which are the transaction validators in a DLT system) to agree on the status of the network and to maintain consistent copies of a single data set. This underlies the ability of a DLT system with multiple nodes to settle transactions.
8. For example, Budish (2025) estimates that for systems that operate at scale, centralization in a single operator reduces the costs of creating trust. This relates to the so-called Blockchain Trilemma, coined by Ethereum co-founder Vitalik Buterin, wherein the scalability of a DLT system trades off with its degree of decentralization and/or its secureness. However, if there are concerns about the credibility of the central authority, decentralization could potentially enhance trust.
9. The models to achieve trust and programmability are interrelated. Databases that rely on external programmability are not well suited for ledgers based on DLT systems because, in the presence of multiple operators, determining which operator should execute the external code and update the database information becomes a complex issue. By contrast, with databases that enable internal programmability, instructions are sent as smart contracts that run directly on the ledger [Lavayssière and Zhang (2024)]. They are pre-validated by all the nodes on the network and provide an automatic confirmation that is generated once the ledger is updated. These databases are well suited to both single-operator and DLT systems.
10. Not all financial assets undergo the four stages of the lifecycle. For instance, shares of common stock in a company or perpetuities do not have a maturity or redemption date.

Figure 2: Frictions along the lifecycle of an asset



Note: This figure does not aim to be fully comprehensive of all possible market frictions present in the lifecycle of financial assets. It only displays certain frictions that have the potential to be affected by tokenization.

The issuance, exchange, servicing, and redemption of assets on financial markets are subject to market frictions (Figure 2). Market frictions are a subcategory of market inefficiencies and include information asymmetries, search problems, transaction costs, and counterparty risks.[11] Asymmetric information is present when one party to a transaction has more information relevant to the transaction than the other. For instance, an entrepreneur taking a company public through an IPO knows more about the prospects of the company than potential buyers. Search frictions are impediments to a match, or agreement, between two parties to a transaction, which may result in unrealized gains from trade. For example, an asset issuer might not know which investors may be interested in purchasing its financial assets and how to reach them. Transaction costs include costs of trade and the opportunity cost of the time involved in financial market activity.[12] Counterparty risk refers to the possibility that one of the parties to a contracted trade does not deliver its part. Digital transactions typically involve a single immediate trade, such as buying a stock on a digital platform. For these transactions, counterparty risk results from the potential delay between one party's payment and the other's delivery of a financial asset (or vice versa), which could allow the second mover to cancel its leg of the exchange after the other leg has settled. Derivative transactions involve more than just a single immediate trade. For example, a call option includes the initial trade of the derivative for the payment, along with the option for the asset holder to purchase the underlying asset at a specified price in the future. For derivative assets, there is an additional layer of counterparty risk stemming from the possibility that one party to the contract may fail to meet its obligations between the writing of the contract and the time that future transactions are supposed to occur. The nature of both subcategories of counterparty risk is similar, as both emanate from the possibility that a counterparty does not honor its part of the transaction.

Specialized intermediaries reduce market frictions in financial markets at different stages of the asset lifecycle. For instance, investment banks conduct due diligence to assess the creditworthiness and financial health of a company engaging in its first stock or bond issuance. By underwriting the issuance, and thereby putting their established reputations on the line, investment banks act as a "credibility bridge" between a new issuer and investors, mitigating frictions related to asymmetric information.[13] Another example is central securities

11. A market inefficiency occurs when there exists a feasible reallocation of resources that would improve the welfare of at least one agent without harming any other.
12. Transaction costs can be related to other inefficiencies (like asymmetric information) or could result from intermediary market power.
13. Intermediaries also play key roles in providing risk management, supplying depth to markets by deploying their balance sheets, and ensuring regulatory compliance throughout the issuance process.

depositories (CSDs) and centralized clearinghouses (CCHs), which help address counterparty risk. CCHs perform the clearing process, which includes ensuring that both parties have sufficient funds to meet their obligations. CSDs centrally record assets and manage settlement functions, only delivering the assets to the buyer once the payment is made. Such actions address counterparty risks involving immediate trades. Central counterparties (CCPs), a subcategory of CCHs, also mitigate the counterparty risk stemming from the possibility that one party to a derivatives contract may fail to meet its obligations between the writing of the contract and the time that future transactions are supposed to occur. CCPs take on this risk by becoming the counterparty to all transacting parties of a transaction. If one of the parties fails to meet its obligations, the CCP pays the affected party using the margin and any collateral collected from the defaulting party, and if these prove insufficient, it uses its guarantee fund.[14]

Financial markets are also affected by externalities, internalities, and market power. Externalities arise when an agent does not consider how its actions affect other financial market participants or agents in the economy more broadly. Externalities can be positive or negative, and both types of externalities can lead a market to socially inefficient outcomes. An example of a positive externality is network effects. For instance, every agent that joins a financial market helps to deepen that market, which can lead to improved outcomes for all participants, such as higher liquidity. However, the joining agent only considers its own costs and benefits from joining, not the benefits to other market participants; hence the number of agents in the market may be lower than socially optimal. An example of a negative externality is the socialization of losses for certain intermediaries, particularly those that are systemically important. An intermediary that has implicit or explicit guarantees of a government bailout can have incentives for excessive risk-taking because it internalizes the full upside of risk (that is, higher profits) but not the full downside, some of which is borne by the public safety net. Internalities arise when financial market participants are not well informed or not perfectly rational. This raises the potential for an agent to take an action that is not in its own interest. An example is a retail investor who does not fully grasp the risk associated with a financial asset. Market power refers to the ability of a firm to influence the price of an instrument or the terms of a transaction. For instance, if there exist barriers for users to move their assets across financial institutions, this will provide such institutions with an increased ability to charge higher-than-competitive fees.

The remaining sections of this paper discuss how tokenization may affect financial market inefficiencies relative to the status quo. The next section analyzes how tokenization could influence financial market frictions throughout the lifecycle of an asset. Tokenization may be able to reduce some market frictions to a greater extent than intermediaries can currently achieve. Alternatively, it may be cheaper to address market frictions in a tokenized setting. The subsequent section examines how the features of tokenization may affect externalities, internalities, and inefficiencies from market power. The effects of tokenization discussed in this paper emanate from changes in the features that define tokenization, that is, the degree of sharedness, programmability, and trust of digital token ledgers. Changes in these features may take place under any of the tokenization models discussed in Subsection 2.1. The potential effects of tokenization apply to all the models, unless otherwise noted.

4. Tokenization and frictions in the lifecycle of an asset

This section discusses the potential impact of the features of tokenization on frictions along the lifecycle of a financial asset. It focuses on a qualitative discussion, which considers, at each stage of the lifecycle, how tokenized ledgers would change frictions as compared with the prevailing financial market structures.

4.1 Asset issuance

Asset issuance requires the creation of records of a firm's bond and shareholders. In presently prevailing market structures, this process involves intermediaries known as "registrars". These are often banks or trust companies. The need for such intermediaries arises because of economies of scale: by managing the recordkeeping and large volumes of transactions toward investors of many firms, these intermediaries incur lower transaction

14. Guarantee funds [ICE (2016)] are collected by CCHs from clearing members to cover losses that may arise in the case that a member defaults.

costs compared to a company that would choose to keep the records on its own. However, these intermediaries have operational costs and charge transaction fees.

A shared and programmable ledger can reduce some of the transaction costs from asset issuance. On a shared ledger, no separate record of asset owners is needed because every asset is linked to the account of the owner that purchased it. The services of a registrar are, therefore, not required, saving on issuers' transaction costs [Cohen et al. (2018)]. Similarly, in the case of the issuance of native tokens, cost savings could arise from the fact that custodian services would not be required. Furthermore, programmability could reduce certain asset issuance costs. For instance, smart contracts could be used to bring greater efficiency to auctions for new shares that are currently distributed by underwriters to their institutional clients [Omar et al. (2021)]. Smart contracts could automatically execute auction rules once predefined conditions are met, reducing the time and resources needed to manage and execute the auctions. Programmability could also enable data to be wrapped into the token asset, which could aid transparency (that is, reduce information search costs) toward potential investors by facilitating access to information on the asset or issuing company, particularly for small-value issuances. For example, in 2022, BNP Paribas distributed a renewables project finance bond in tokenized form [Russell-Walling (2022)], embedding bond term sheet and environmental, social, and governance data in the token, which the bank argued improved transparency for this small-scale project finance initiative.[15]

4.2 Asset trading

Asset trading is subject to several frictions. These frictions originate from the need to overcome counterparty risks and limit search frictions, as well as from data processing costs.

Although financial intermediaries, including CCHs and CSDs, currently mitigate counterparty risk, they impose fees, thereby increasing the costs of conducting transactions on financial markets. Even if these financial intermediaries do not strive for profit, their cost recovery will necessitate transaction fees. These costs include the staff, equipment, and office space, and for a CCH that also acts as a central counterparty to derivatives trades (that is, a CCP) there can be costs associated with absorbing counterparty risk.

Shared and programmable ledgers offer a route to overcome counterparty risk on immediate trades through simultaneous settlement, without reliance on a financial intermediary.[16] When funds and assets are placed on a shared ledger where programmability enables parties to write smart contracts, a smart contract for the exchange of funds and an asset can be written, where the condition for executing the trade is that both the funds and the asset have been locked in place, ready for automated exchange, or otherwise the assets revert to the original owner. This ensures the simultaneity of delivery versus payment.[17] Such automation could result in transaction cost savings [Benos et al. (2022)]. Relatedly, a shared and programmable ledger can facilitate further ways in which transactions can be bundled, thereby minimizing counterparty risks from consecutive transactions. For instance, a transaction to lend an asset against local currency liquidity, then swapping the proceeds for another currency, and purchasing a foreign asset can be bundled into one so that it happens only if and when all currencies and assets are available.[18]

Despite the aforementioned benefits, sharedness and programmability cannot mitigate counterparty risk stemming from the possibility that one party to the contract may fail to meet its obligations between the contract's initiation and the time when future transactions are expected to occur. For example, if one party writes a derivative contract for the future delivery of an underlying asset conditional on certain events (such as a price level being reached), there is a risk that this party

15. There have been other examples of digital technologies improving transparency on financial markets, which have had beneficial impacts on bond transaction costs. These include electronic trading [Hendershott and Madhavan (2015)] and the Trade Reporting and Compliance Engine (TRACE) [Bessembinder et al. (2006), Edwards et al. (2007), Goldstein et al. (2006), Jacobsen and Venkataraman (2018)].
16. Note that a CCH may have additional functionalities beyond contributing to address counterparty risk, such as dispute arbitration [Perino (2002)] and the monitoring and management of collateral of open positions to ensure sufficient coverage for the associated risks [ECB (2005)]. Such functionalities could necessitate the continued presence of an intermediary, even if its scope could be more limited under tokenization (that is, if ensuring that transacting parties have sufficient funds to meet their obligations is no longer required). The same could be the case for other intermediaries.
17. Simultaneity of exchange and settlement could still fail in a shared and programmable ledger because of operational risk events, such as cyberattacks. These risks are also present when trades are performed through traditional financial intermediaries.
18. Package trade, a transaction that involves the execution of two or more component transactions in financial instruments where the execution of each component is simultaneous and contingent upon the execution of all the other components, does take place in current financial markets, with settlement managed by broker-dealers, centralized clearinghouses, and trading platforms such as stock exchanges.

reneges on its obligations between the writing and the potential future execution of the contract, leaving the counterparty with losses. Most derivatives markets employ a CCP that ensures the execution of trades even if one of the parties reneges. There are two potential routes to address this form of counterparty risk on tokenized ledgers: either an intermediary (that is, a CCP or a trusted financial institution that guarantees its clients' repayments) is employed or all assets are locked in (by a program) for the duration of the contract. The latter would imply that a party that wishes to sell, for example, an option on a stock, would need to provide the stock at the time that the option contract is sold and have this collateral locked in place until the option expires or is executed.[19] Although this method would allow for trading without counterparty risk, it could be perceived as costly by investors (as a result of collateral lock-in) and reduce the incentives to write and trade derivatives contracts.

A shared ledger can also enable instantaneous settlement, potentially reducing frictions to transaction speed (constituting a transaction opportunity cost). Whereas simultaneous settlement refers to trades in which both legs are settled at the same time, instantaneous settlement means that settlement takes place immediately once a trade has been agreed upon. Even in cases where counterparties desire to perform an instantaneous trade, for a CCH to clear trades with broker-dealers, it can currently take up to several business days (shortened to one day in the U.S. since May 2024) [OCC (2024)]. Delays in the current system are the outcome of having to synchronize multiple ledgers, including asset/money accounts at brokers and asset registration at a central custodian. Synchronization typically involves manual processes, which translate into transaction costs. A ledger that is shared by all participants may be programmed not to require manual synchronization, allowing for the reduction of time delays. The availability of faster settlement options can improve capital allocation. For instance, for transactions involving collateral, shortened settlement cycles can reduce the amount of time that collateral is locked, enabling better liquidity management.[20] In this sense, transaction settlement speed constitutes a friction that is part of the transaction cost category (namely, an opportunity cost).

A shared and programmable ledger can reduce search frictions on specialized markets with limited liquidity. An example is certain over-the-counter (OTC) asset markets, on which many types of bonds, derivatives, and foreign exchange are traded. OTC markets are facilitated through networks of broker-dealers that intermediate the market. The broker-dealers act as market-makers by quoting prices at which they are willing to buy and sell a financial asset. A shared digital ledger to which investors have direct access could facilitate finding a counterparty to perform a trade, potentially reducing the need to rely on financial intermediaries, such as broker-dealers. This could reduce the cost of trading on such markets (for example, the bid-ask spread charged by broker-dealers).[21]

4.3 Asset servicing and redemption

Transmitting payments to the owners of a company's financial assets currently involves a transaction cost friction. When firms need to make interest, principal, or dividend payments, they need to know who the bond and shareholders are. As discussed, the recordkeeping of a firm's bond and shareholders is typically performed by registrars, which have operational costs and charge transaction fees.

The services of collection and distribution of dividends and interest payments could be automated on a shared and programmable ledger, reducing the associated costs. As already mentioned, on a shared ledger no separate record of asset owners is needed because every asset is linked to the account of the owner that purchased it. Moreover, through the application of smart contracts on a programmable ledger, tokenized assets can embed asset servicing. A smart contract could

19. For certain types of derivative transactions, there is no specific asset to be locked away, and therefore this form of addressing counterparty risk may not be feasible. For example, in the case of an interest rate swap, which is an agreement to swap income streams referencing unknown future rates, it would not be possible to calculate a collateral amount that would fully cover the exposure.

20. The benefit discussed in this paragraph regarding faster settlement pertains to the length of time between the moment the buyer (seller) submits the payment (asset) and the moment the buyer receives the asset (payment). This is distinct from the interval between the time the contract is agreed upon and the moment the asset or payment has to be provided. Shortening this interval comes with pros and cons. On the one hand, a shorter time frame can reduce the gap between exposure calculations and the settlement of variation margin [Choudhury et al. (2023)]. On the other hand, a larger time window enables investors to sell assets that they do not yet own or buy assets before they have the needed funds. With instantaneous settlement in this respect, assets and funds have to be "prepositioned" for a trade to occur [Lee (2021), Lee et al. (2022)], potentially by intermediaries rather than the contracting parties, and potentially in the form of collateral. In addition, when investors are limited to only selling financial assets that they already hold, there can be unintended consequences on the informational environment, through the partial revelation of a trading history, which can have adverse consequences on market liquidity [Lee et al. (2024)].

21. Note, however, that OTC markets may facilitate trading in, for example, non-standardized instruments or complex derivative transactions requiring capital allocation from broker-dealers.

specify the amount and timing of payments that an asset token should service, and this could be made conditional on the funds made available (such as a dividend distribution decision) by the company that issued the assets. Such automation would bring about a reduction in the costs of asset servicing activities.

5. Tokenization and externalities, internalities, and market power

This section considers how the features of tokenization affect the extent of market inefficiencies caused by externalities, internalities, and market power. As mentioned above, existing market structures and regulations are taken as given and the section asks what change in market inefficiencies (compared to those prevalent in the current status quo) would come about by the adoption of tokenized ledgers.

5.1 Shock transmission externalities

High leverage, low funding liquidity, and interconnectedness amplify the transmission of shocks among financial intermediaries, exacerbating externalities and adversely affecting financial stability. Financial institutions do not have incentives to internalize the transmission of shocks to each other and related financial stability implications when deciding on their funding structure and connections (within regulatory boundaries). When a financial institution is highly leveraged, it has greater incentives to take risks. Possessing a small equity stake relative to its issued debt allows the institution to fully benefit from the potential profits that such risk-taking may render, while if the institution goes bankrupt, some of the losses would be borne by creditors (and/or the public sector in the case of bailouts), thereby imposing negative externalities on other market participants and/or taxpayers. Moreover, highly leveraged institutions, whose equity is small relative to the size of their balance sheets, have fewer buffers to withstand shocks. Similarly, institutions with low levels of funding liquidity may be forced to sell illiquid assets in the event of a shock, potentially leading to declines in the values of such assets and to financial losses. Risks to the financial system are amplified when leveraged financial institutions and/or institutions with low levels of liquidity are significantly interconnected with each other. A domino effect may occur when a shock pushes one institution into bankruptcy, leading it to default on its obligations to other interconnected financial institutions. These, in turn, may themselves be highly leveraged or lack sufficient liquidity and thus have limited capacity to withstand the shock.

Tokenization may increase the spread of shocks across financial institutions, undermining financial stability. This relationship depends on whether and how tokenization affects leverage, funding liquidity, and interconnectedness. It also depends on tokenization's influence on the speed with which shocks are able to spread, because greater speed increases the difficulty for policymakers in mitigating the effects and costs associated with externalities.

A shared and programmable ledger can reduce the cost of leverage, consequently strengthening the incentives for financial institutions to lever up. As discussed in Section 4.1, tokenization has the potential to decrease the cost of asset issuance. Both Leung et al. (2023) and Liu et al. (2023) find lower costs associated with the issuance of tokenized bonds as compared to conventional bonds. A permanent reduction in the cost of debt issuance could render leverage more appealing for financial institutions (holding capital requirements and other regulations to mitigate excessive leverage constant and assuming these are not binding). However, the impact will depend on how sensitive leverage incentives are to issuance cost reductions. Programmability could also facilitate the leveraging of intermediary balance sheets through so-called rehypothecation, whereby assets that are used as collateral to secure a transaction are further used by the intermediary arranging the transaction: tokens received as collateral could be automatically rehypothecated using smart contracts [FSB (2024)]. Should leverage increase, this could raise the potential for externalities and financial stability risks. This scenario illustrates how mitigating one inefficiency, in this case, transaction costs, could inadvertently amplify another, such as externalities.

Tokenization may affect funding liquidity and increase the spread of shocks if financial institutions' reliance on retail deposits becomes harder to sustain. Financial institutions generally fund themselves either with (insured) retail deposits or through market-based (wholesale) funding. The

former are considered more stable than the latter. If a shared and programmable ledger reduces the costs of investing in financial assets other than bank deposits, retail depositors may increase their share of funds invested in such securities. Under these conditions, the availability of retail deposits would decline, and financial institutions could find themselves more reliant on wholesale funding. Such funding is more likely to dry up at the same time that a shock hits the market [Brunnermeier and Pedersen (2009)], exacerbating the likelihood of default and associated knock-on effects.

A shared and programmable ledger can lead to increased interconnectedness, resulting in heightened shock transmission externalities. A larger number of investors having cheaper and faster access to a wider set of markets and assets (as a result of the removal of frictions that underpin transaction costs) can lead to greater interdependencies within the financial system. Higher interconnectedness may magnify the externalities of market participants on each other and/or the public sector stemming from economic shocks (the domino effect discussed previously).[22] Programmability facilitates the construction of complex assets, whose returns may depend on several underlying assets or on the settlement of several preceding transactions. Composed assets (whose label reflects that they are composed of other assets) are, by their nature, interconnected to other assets. Increased sharedness facilitates access to these interconnected assets by a larger set of investors.

One way that increased interconnectedness can amplify shocks is by expanding the number of assets with values pegged to other assets or currencies. A pegged instrument or an intermediary offering a fixed-value claim has broader interlinkages when an instrument is used as collateral in other transactions or when an intermediary has borrowed funds from other intermediaries. In such cases, a run on one instrument or intermediary can precipitate a broader crisis of confidence in financial markets or institutions.[23]

A widely shared ledger can increase trading speed, potentially amplifying the velocity and intensity of shock transmission. Leveraged intermediaries have the potential to face runs, which involve many investors quickly and simultaneously trying to remove their funds or sell their assets. These developments display cascading dynamics, as the observation of such events induces even more investors to rapidly withdraw or sell. Under such circumstances, separate ledgers can function as inadvertent dams to the flow of funds, as time is needed to transfer information between ledgers that are not interoperable. Shared ledgers may remove such dams.[24,25]

By facilitating automated trading behavior, a programmable ledger can increase trading speed and as a consequence the velocity and intensity of shock transmission. Automated trading entails the use of systematic and rule-based algorithms to execute trades.[26] Extreme market volatility induced by automated trading behavior, often referred to as a flash crash, has already been observed in present-day markets [Kirilenko et al. (2017)]. By enhancing the programmability of ledgers, tokenization could exacerbate the risk of such events. This could happen, for example, by having chains of smart contracts written to be contingent on one another across a broader range of assets and extending access to these contracts to a larger user base.

5.2 Market infrastructure investment externalities

Increased sharedness and programmability have the potential to raise the costs associated with market infrastructure investment externalities if they make trading faster or lead to increased interconnectedness. Counteracting operational risk on a ledger requires significant investment, including a multilayered security approach that includes

22. There exists a large literature on the role of interconnectedness in financial stability. See, for example, Acemoglu et al. (2015), Roncoroni et al. (2021), and Ghio et al. (2023). Moreover, Camera and Gioffré (2025) use numerical simulations to investigate the extent to which increased interconnectedness from tokenization can amplify financial stability risks.
23. One example is stablecoins that offer a fixed redemption value against, for instance, U.S. dollars. These could experience runs because of uncertainty about the credibility of the redemption promise (for instance, if they lack verifiable reporting on their underlying assets). For case studies of stablecoin runs, see, for example, IMF (2021), Uhlig (2022), Briola et al. (2023), Kosse et al. (2023), and Anadu et al. (2024).
24. One example is the transfer of funds from one bank to another in the U.S., as traditional bank transfers are not instantaneous. Although funds can be instantaneously transferred (from the users' perspective) through financial intermediaries such as Zelle or Venmo, such transfers have amount limits. Instantaneous bank transfers can also be made through FedNow, an instant payment infrastructure developed by the Federal Reserve. However, uptake by banks is not complete.
25. Depending on their design, central bank digital currencies may also affect run risks for the financial sector [Mancini-Griffoli et al. (2018), Agur et al. (2022), Gross et al. (2025)].
26. Automated trading is common in stock exchanges (for example, New York Stock Exchange), futures and options exchanges (for example, Chicago Mercantile Exchange), and foreign exchange markets, among others.

encryption, authentication, access controls, and monitoring. If a ledger is operated by one or multiple private parties, who bear only part of the total cost incurred by all ledger participants from operational risk events, then they could have incentives to underinvest in operational security. The costs associated with such events could rise under tokenization. Any digital system faces a certain probability of failing. Programs can fail because of software errors, unexpected human behaviors, or the malicious intent of illicit or fraudulent actors [Lavayssière and Zhang (2024)]. The aggregate effect of a malfunction in one transaction depends on how many other transactions are affected by it. The increased interconnectedness and number of users (sharedness) who can be associated with tokenized ledgers (discussed in Section V.A) can magnify the effects of failures. Moreover, increased sharedness can come with a concentration risk because a widely shared ledger can become a single point of failure to cyberattacks. A dominant central role for a single financial market infrastructure implies a more widespread potential disruption should that infrastructure experience a disruptive cyberattack.[27]

5.3 Network externalities

Market liquidity and innovation incentives are subject to externalities that can be affected by the sharedness and programmability of a ledger. A network externality (also known as network effect) occurs when the value to an individual agent of taking an action depends on how many other agents take the same action. Participation in a ledger constitutes such an externality when the extent of market liquidity, which is of value to all market participants, depends on how many users the ledger has. That is, an individual investor or issuer's decision to join a ledger can depend on how many other investors and issuers have decided to join.

Sharedness and market liquidity are closely related as the increased use of a shared ledger implies that more buyers and sellers are brought together. For instance, if an OTC market is moved to a shared ledger, the liquidity of that market could improve as a result of the offering of standardized contracts, the increased clarity of market pricing information (that is, having the buy and sell offers centrally quoted rather than bilaterally contracted), and the lower search frictions caused by all agents trading on the same ledger. Liquidity, in turn, is closely related to transaction costs.[28] For instance, bid-ask spreads tend to shrink as market liquidity increases. Lower transaction costs may also incentivize more users to join the ledger, carrying the potential for a virtuous cycle of increased sharedness.

The fractionalization of assets enabled by tokenized ledgers can further enhance incentives for retail investors to join such ledgers, increasing positive market liquidity externalities. The price of a single stock or bond can be hundreds of dollars, which can make it hard for small retail investors to access, particularly if they desire to diversify. Asset fractionalization refers to the splitting of assets into smaller pieces and can be provided to retail investors either by ledger design or by financial intermediaries. Currently, only a few intermediaries, such as Robinhood and Fidelity, enable asset fractionalization.[29] While Robinhood and Fidelity offer U.S.-domiciled equities to investors in fractionalized form, they do not offer fractionalized bonds or fractionalized international equities.[30] Moreover, where intermediaries do provide such services, this may come with an added risk to investors: even if the fractional offering is fully collateralized in terms of underlying assets, the bankruptcy of a broker could imply a lengthy period to recover the value of fractionalized assets.[31] Fractionalization can also be attained by design in new ledgers where assets can be issued in tokens of arbitrarily small value. If the tokens are offered directly by the issuer, there is also no intermediary risk involved in holding a fractionalized asset on such a ledger.

However, should interest in tokenization cause a proliferation of noninteroperable ledgers, this could lead to a decrease in market liquidity and, therefore, reduce the beneficial network externalities. If several competing and noninteroperable new asset trade ledgers are launched, and assets and users end up segmented across ledgers, then

27. In permissionless ledgers, however, a larger number of users generally increase the value of the (settlement) asset used to reward transaction validators, attracting more validators, and increasing the network's security.
28. For example, joining primary and secondary markets on the same ledger may facilitate trading [OECD (2020)].
29. To an extent, exchange traded funds (ETFs) perform a similar function (allowing retail investors to buy a complete portfolio of assets). However, an ETF has a pre-determined allocation, which cannot be established by the retail investor. Securitization also shares some features with the fractionalization of assets enabled by tokenization [Chang (2020)].
30. For a discussion and empirical analysis of equity fractionalization in the U.S., see Bartlett et al. (2024) and Da et al. (2025). These studies also point to the risks inherent in fractionalization, including meme-stock-like trading frenzies.
31. In the U.S., the Securities Investor Protection Corporation (SIPC) insures retail investors' securities and cash at brokerage firms up to a limit, and would be involved in the resolution process of a failed broker.

fragmented trading could result. This risk is particularly pertinent when ledgers are set up by private entities and there is a lack of coordination and regulatory requirements to ensure interoperability. The segmentation of market participants across several ledgers could lead to lower liquidity (and, relatedly, higher transaction costs and search frictions) than in current financial markets [Allen and Wittwer (2023)]. A proliferation of noninteroperable ledgers would represent a decrease in sharedness. Consequently, the risks discussed in this paragraph emanate from a weakening rather than a strengthening of the features that define tokenization.

5.4 Knowledge externalities

The benefits from the creation of new knowledge are not always fully appropriated by the innovator, resulting in positive externalities to other agents. For instance, open code developments can be used by anyone, allowing agents to benefit from new programs without incurring development costs. For this reason, innovators have incentives to invest in new knowledge to an extent which is lower than socially desirable. The strength of knowledge externalities also depends on the size of the market. The larger the number of users of the platform (sharedness), the greater the social benefits from new knowledge creation, and the higher the potential for the innovator to profit from creating new knowledge.

More shared and programmable ledgers provide a common infrastructure for developers to build on and ease the path for innovations to attain critical mass. A larger shared user base strengthens private companies' incentives to invest in such innovation.[32] Moreover, when more investors and asset issuers share a programmable ledger, there is a broader set of contingent actions that can be automatically executed, which could expand the universe of contractual outcomes [Townsend and Zhang (2023)].

Whether or not the code of a ledger is publicly accessible can affect knowledge externalities related to innovation.[33] Open code access can be either a boon or a bane for innovation incentives. Open code access facilitates building improvements on top of work produced by others, which can reduce the cost involved in developing new products and services. However, open access also reduces the potential benefit of investing in research and development because companies foresee that their own code could quickly be copied or improved upon by others, leaving less scope to monetize the innovation.[34]

5.5 Retail investor internalities

To the extent that direct trading by retail investors is allowed on these ledgers, it has the potential to exacerbate internalities. Retail investor internalities arise when they do not fully comprehend the risks associated with an asset. When trading with brokers, retail investors generally have the option to request the services of a financial adviser.[35] With direct trading, investors could still seek the advice of independent financial consultants, but this may entail an additional search effort, reducing the incentives to hire such services. Should retail investors resort to direct trading without advice, there could be increased costs associated with internalities from purchasing financial products that advertise high returns and whose risks they do not fully comprehend. Moreover, unadvised retail investors may be more likely to insufficiently diversify away idiosyncratic risk in their portfolio. While ill-advised or unadvised trade exists on traditional brokered markets too, its occurrence could be greater if direct trading occurs on tokenized ledgers, unless appropriate regulations are put in place to guarantee adequate consumer protection.

Internalities are likely to increase when the ledger offers a greater ease of constructing complex products through programmability. Costs associated with retail investor internalities are especially likely to occur when transacting with complex assets, such as derivatives, whose market value can depend on multiple variables and states of the world. Instances have occurred where investors faced sizable and unexpected losses from derivative contracts that they did not fully comprehend [Storbeck (2023)]. Increased programmability can make it easier to create complex assets, thereby compounding the costs associated with internali-

32. One example would be a new smart contract that streamlines the implementation of auctions for stock issuances. A ledger with a limited number of users might not offer sufficient incentives for the development of such a smart contract.
33. On permissionless DLT systems, the code is necessarily publicly accessible.
34. Licensing or patenting might prevent users from copying or using the code developed by others to create enhancements. However, these measures would also limit the benefits associated with open-source code.
35. Financial advisors can provide unsound advice too. Nevertheless, the odds of appropriate investment strategies when engaging the average financial advisor are likely higher than when the average retail investor deploys their own strategy.

ties, particularly if retail investors engage in direct trading on tokenized ledgers.[36]

Tokenization on permissionless distributed ledgers can come with larger costs associated with internalities. Regulatory and legal rulings are particularly difficult to enforce in permissionless distributed ledgers because of their decentralized governance structures, for instance, regarding the reversal of fraudulent/illicit transactions. This implies that mitigating risks to retail investors could be more challenging.

5.6 Market power inefficiencies

Switching costs and barriers to asset trade can uphold broker rents by stifling broker competition. Established intermediaries in asset trade can profit off the market power derived from separate ledgers that make it costly (in terms of time, effort, and sometimes required fees) for retail investors to switch to the services of another intermediary. For instance, if a retail investor currently wants to switch brokers in the U.S., the services of a special clearinghouse that allows in-kind transfers of assets are required. This clearinghouse is called an automated customer account transfer service (ACATS). The investor needs to fill out a transfer initiation form with the receiving broker, with information that matches the records held at the old broker, and share the latest account statement. The new broker contacts the old broker, and they initiate the transfer through ACATS. The process typically takes between three and six business days. The time and effort involved discourages retail investors from changing brokers in search for lower fees. In turn, this subdues the incentives for brokers to reduce fees with the aim of attracting new clients.

On a shared ledger, the transfer and reconciliation of data between brokers could become instantaneous, reducing the costs involved in switching and harnessing competition to bring about lower brokerage fees. On a shared ledger, every broker could immediately verify every other broker's record, because the information about the assets on these records could be made instantly accessible. A transfer of assets could then be implemented directly and without the need for a clearinghouse. This could lead to two types of cost reductions as compared to the current financial market structures. First, the costs of operating a clearinghouse can be saved, and such savings may find their way to lower transaction costs for retail investors. Second, the increased ease of switching would result in competitive pressure on brokerage fees. For instance, new entrants in the brokerage business could attempt to quickly gain market share by undercutting the fees charged by established brokers.

Competitive pressure to limit brokerage fees could also arise if the shared ledger enables direct trading by retail investors or allows them to purchase assets directly from issuers. In the currently prevailing structures for financial asset markets, this is generally not possible. Although there are some exceptions, such as TreasuryDirect, which allows U.S. residents to purchase debt directly from the U.S. Treasury, retail trading is generally required (by regulation) to be done through brokers.[37]

Even when direct trading is possible, brokers may well continue to execute most trades, although possibly at lower fees. An empirical case study that speaks to the continued role of brokers in directly accessible markets is provided by Bergquist et al. (2024). They study the introduction of a mobile-phone accessible trading ledger offered to farmers in Uganda. The ledger can directly match the buyers and sellers of agricultural commodities. They find that, despite the ease of access, very few small farmers choose to directly trade on the ledger. Instead, the same agents that intermediated trade before the introduction of the ledger continue to do so afterward. Nevertheless, the trading fees charged to the small farmers do decline: their outside option to forego the intermediaries and trade directly raises the competitive pressure on the intermediaries, which leads to the decline in fees.[38]

36. Regulated tokenized ledgers could enable direct trading by retail investors subject to certain constraints, whose enforcement may be facilitated trough programmability, in the same way that today, for example, there are restrictions on who can sell option contracts (currently, retail investors can buy stocks and may be able to buy simple options, but they are not allowed to sell options).

37. Access to direct trading on the New York Stock Exchange (NYSE), which enables investors to execute trades by directly interacting with the trading platform's electronic order book, is restricted to the broker-dealers registered with the Securities and Exchange Commission with self-regulatory organization status and established clearing firm connections. Retail investors are explicitly excluded from direct NYSE membership and trading capabilities. Instead, individual investors must rely on brokerage firms to execute trades on their behalf, with these firms acting as intermediaries between retail clients and the exchange. Some brokers may offer more advanced routing options or trading platforms, but true direct access to NYSE trading systems remains unavailable to typical retail investors.

38. Furthermore, despite the presence of a centralized ledger with direct access, intermediaries can nevertheless sometimes retain market power derived from client relations and continue to conduct bilateral trades with other intermediaries on their clients' behalf outside of the ledger.

Tokenization can also affect market power inefficiencies by a different route: if the ledger is privately owned, the owner(s) could extract monopoly rents. The more shared and dominant a ledger is, the larger the market power of its owner(s) may be. The ledger owner(s) would have privileged access to data, which could be used to increase profits, for instance, through data sales or direct credit provision [Agur et al. (2025)].[39] Market power could also be monetized by charging fixed or variable (for example, per transaction) fees for firms or households that wish to be active on the ledger. One strategy could be for a private provider to initially establish the ledger with free access and, once dominant, begin to charge fees. Network effects, setup costs, and returns to scale can function as barriers to entry that prevent other private ledgers from challenging the dominant position of the fee-charging ledger by undercutting its fees.

6. Conclusion

We adopt a broad definition of tokenization, the key features of which are sharedness to allow ownership transfer, programmability to execute contingent transactions, and trust by participants in the system's integrity for users to willingly transact on the ledger.

We provide a conceptual framework based on market inefficiencies to analyze how tokenization may affect financial markets, and we link each effect to changes in the specific features of token ledgers. We take existing market structures and regulations as given when investigating the changes that tokenization could bring to market inefficiencies compared to the current status quo. The analysis is not circumscribed to a particular model of tokenization but rather applies to all, except when specifically noted.

The potential effects of tokenization stem from improved sharedness and programmability. These effects include further mitigation and reduced costs associated with easing frictions throughout the asset lifecycle by reducing the need for certain intermediaries, automating procedures, mitigating counterparty risk through simultaneous settlement, enabling faster settlement, and reducing search frictions. In addition, more shared and programmable ledgers may affect the degree and costs associated with externalities, internalities, and market power. On the positive side, they may result in increased innovation and liquidity externalities that could benefit the functioning of financial markets. At the same time, they may facilitate the spread of shocks across financial institutions, increasing financial stability risks. They could also augment the costs of operational risk events, exacerbate retail investor internalities, and affect market power. Further research is needed to provide a quantification of the benefits and risks associated with the impact of tokenization on market inefficiencies.

Regulations play a key role in addressing financial market inefficiencies, and they might need to be adapted to both harness the benefits and mitigate the adverse effects of tokenization. Such regulatory changes are beyond the scope of this paper and are the remit of standard-setting bodies and financial sector regulators and supervisors around the world.

The authors would like to thank Parma Bains, Marianne Bechara, Eugenio Cerutti, Luis Brandao-Marques, Yaiza Cabedo, Alexander Copestake, Giovanni Dell'Ariccia, Jose Garrido, Kazuhiro Hiraki, Dong He, David Hofman, Xavier Lavayssière, Junghwan Mok, Murad Omoev, Marco Reuter, James Roberts, Fabian Schär, Nadine Schwarz, Felix Simione, Nobuyasu Sugimoto, Dmitry Vasilyev, Mauricio Villafuerte, Torsten Wezel, Nicholas Zhang, and participants at the IMF's Fintech Brownbag Seminar for their helpful comments.

39. Traditional trading platforms such as exchanges face regulatory requirements on data usage. Examples include the General Data Protection Regulation (GDPR) in Europe and the California Consumer Privacy Act in the U.S. Moreover, regulatory bodies such as the SEC and the Commodity Futures Trading Commission (CFTC) impose requirements for reporting trades to ensure transparency in the United States.

References

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi, 2015, "Systemic risk and stability in financial networks," American Economic Review 105:2, 564-608

Agur, I., A. Ari, and G. Dell'Ariccia, 2022, "Designing central bank digital currencies," Journal of Monetary Economics 125, 62–79

Agur, I., A. Ari, and G. Dell'Ariccia, 2025, "Bank competition and household privacy in a digital payment monopoly," Journal of Financial Economics 166: 104019

Agur, I., X. Lavayssière, G. Villegas-Bauer, J. Deodoro, S. Martinez Peria, D. Sandri, and H. Tourpe, 2023, "Lessons from crypto assets for the design of energy efficient digital currencies," Ecological Economics 212, 107888

Aldasoro, I., S. Doerr, L. Gambacorta, R. Garratt, and P. K. Wilkens, 2023, "The tokenisation continuum," Bank for International Settlement bulletin no. 72

Aldasoro, I., G. Cornelli, J. Frost, P. K. Wilkens, U. Lewrick, and V. Shreeti, 2025, "Tokenisation of government bonds: assessment and roadmap," Bank for International Settlement bulletin no. 107

Allen, J., and M. Wittwer, 2023, "Centralizing over-the-counter markets?" Journal of Political Economy 131:12, 3310-3351

Anadu, K., P. Azar, M. Cipriani, T. M. Eisenbach, C. Huang, M. Landoni, G. La Spada, M. Macchiavelli, A. Malfroy-Camine, and J. C. Wang, 2024, "Runs and flights to safety: are stablecoins the new money market funds?," Federal Reserve Bank of New York staff report no. 1073

Bains, P., 2022, "Blockchain consensus mechanisms: a primer for supervisors," International Monetary Fund fintech note no. 2022/003

Banerjee, A., I. De Bode, J. Sevillano, M. de Vergnes, and M. Higginson, 2023, "Tokenization: a digital-asset déjà vu," McKinsey & Co., https://tinyurl.com/yknttraaf

Bartlett, R. P., J. McCrary, and M. O'Hara, 2024, "Tiny trades, big questions: fractional shares," Journal of Financial Economics 157: 103836

Benos, E., W. Huang, A. J. Menkveld, and M. Vasios, 2023, "The cost of clearing fragmentation," Management Science 70:6, 3381-4165

Bergquist, L. F., C. McIntosh, and M. Startz, 2024, "Search costs, intermediation, and trade: experimental evidence from Ugandan agricultural markets," National Bureau of Economic Research working paper no. 33221

Bessembinder, H., W. Maxwell, and K. Venkataraman, 2006, "Market transparency, liquidity externalities, and institutional trading costs in corporate bonds," Journal of Financial Economics 82:2, 251-288

BIS, 2024, "Project Agorá: central banks and banking sector embark on major project to explore tokenisation of cross-border payments," Bank for International Settlement, April 3, https://tinyurl.com/49fh3zdk

Briola, A., D. Vidal-Tomás, Y. Wang, and T. Aste, 2023, "Anatomy of a stablecoin's failure: the Terra-Luna case," Finance Research Letters 51, 103358

Brogaard, J., T. Hendershott, and R. Riordan, 2014, "High-frequency trading and price discovery," Review of Financial Studies 27:8, 2267–306

Brunnermeier, M. K., and L. H. Pedersen, 2009, "Market liquidity and funding liquidity," Review of Financial Studies 22:6, 2201-2238

Budish, E., 2025, "Trust at scale: the economic limits of cryptocurrencies and blockchains," Quarterly Journal of Economics 140:1, 1–62

Camera, G., and A. Gioffre, 2025, "Tokenization and interconnectedness: a numerical exploration," Economic Science Institute at Chapman University working paper no. 25-08

Carapella, F., G. Chuan, J. Gerszten, C. Hunter, and N. Swem, 2023, "Tokenization: overview and financial stability implications," Board of Governors of the Federal Reserve System finance and economics discussion series no. 2023-060r1

Chang, C., 2020, "From securitization to tokenization," in Pentland, A., A. Lipton, and T. Hardjon (eds.), Building the new economy, MIT Press

Choudhury, R., K. Jhanji, H. Samd, S. Gleeson, and S. Bennett, 2023, "Impact of distributed ledger technology in global capital markets," Boston Consulting Group, Clifford Chance, Cravath, and GFMA, https://tinyurl.com/zab6dcu6

Cohen, R., P. Smith, V. Arulchandran, and A. Sehra, 2018, "Automation and blockchain in securities issuances," Butterworths Journal of International Banking and Financial Law 33, 144-150

CPMI, 2024, "Tokenisation in the context of money and other assets: concepts and implications for central banks," Report to the G20, Committee on Payments and Market Infrastructures

Da, Z., V. W. Fang, and W. Lin, 2025, "Fractional trading." The Review of Financial Studies 38:3, 623–660

ECB, 2005, "Central counterparty clearing houses and financial stability," European Central Bank, Financial Stability Review, Section IV F, December

Edwards, A. K., L. E. Harris, and M. S. Piwowar, 2007, "Corporate bond market transaction costs and transparency," The Journal of Finance 62:3, 1421-1451

FSB, 2023, "The financial stability risks of decentralised finance," Financial Stability Board

FSB, 2024, "The financial stability implications of tokenization," Financial Stability Board

Ghio, M., L. Rousova, D. Salakhova, and G. Villegas-Bauer, 2023, "Derivative margin calls: a new driver of MMF flows," International Monetary Fund working paper no. 2023/061

Goldstein, M. A., E. S. Hotchkiss, and E. R. Sirri, 2007, "Transparency and liquidity: a controlled experiment on corporate bonds," Review of Financial Studies 20:2, 235–273

Gross, M., M. Miccoli, J. Verrier, and G. Villegas-Bauer, 2025, "Evaluating the implications of CBDC for financial stability," International Monetary Fund fintech note no. 2025/008

Hendershott, T., and A. Madhavan, 2015, "Click or call? Auction versus search in the over-the-counter market." Journal of Finance 70:1, 419-447

ICE, 2016, "Manage your risk: how clearing works," Intercontinental Exchange, Inc., https://tinyurl.com/3w9asksr

IMF, 2021, "The crypto ecosystem and financial stability challenges," (Chapter 2), Global Financial Stability Report, International Monetary Fund

IMF, 2023, "Elements of effective policies for crypto assets," International Monetary Fund policy paper no. 2023/004

Jacob Leal, S., M. Napoletano, A. Roventini, and G. Fagiolo, 2016, "Rock around the clock: an agent-based model of low- and high-frequency trading," Journal of Evolutionary Economics 26, 49-76

Jacobsen, S. E., and K. Venkataraman, 2018, "Does trade reporting improve market quality in an institutional market? evidence from 144A corporate bonds," SSRN working paper no. 3171056

Kirilenko, A., A. S. Kyle, M. Samadi, and T. Tuzun, 2017, "The flash crash: high-frequency trading in an electronic market," Journal of Finance 72:3, 967-998

Kosse, A., M. Glowka, I. Mattei, and T. Rice, 2023, "Will the real stablecoin please stand up?," Bank for International Settlements working paper no. 141

Kumar, S., R. Suresh, D. Liu, B. Kronfellner, and A. Kaul, 2022, "Relevance of on-chain asset tokenization in 'crypto winter'," Boston Consulting Group

Lavayssière, X., 2023, "The design of programmable ledgers," SSRN working paper 4797032

Lavayssière, X., and N. Zhang, 2024, "Programmability in payment and settlement systems: concepts and implications," International Monetary Fund working paper no. 2024/177

Lee, A., 2021, "What is programmable money?" Board of Governors of the Federal Reserve System FEDS notes

Lee, M., A. Martin, and B. Müller, 2022, "What is atomic settlement?" Liberty Street Economics, Federal Reserve Bank of New York

Lee, M., A. Martin, and R. M. Townsend, 2024, "Optimal design of tokenized markets," Federal Reserve Bank of New York staff report no. 1121

Leung, V., J. Wong, A. Ying, and W. Wan, 2023, "An assessment on the benefits of bond tokenization," Hong Kong Institute for Monetary and Financial Research (HKIMR) research paper no. 17/2023

Liu, J. I. Shim, and Y. Zheng, 2023, "Absolute blockchain strength? Evidence from the ABS market in China," Bank for International Settlement working paper no. 1116

Mancini-Griffoli, T., M. S. M. Peria, I. Agur, A. Ari, J. Kiff, A. Popescu, and C. Rochon, 2018, "Casting light on central bank digital currencies," International Monetary Fund staff discussion note no. 2018/08

Mancini-Griffoli, T., et al., 2024, "G20 note on financial platforms: what are they and what are their macro-financial implications?" International Monetary Fund staff paper to the G20

OCC, 2024, "Securities operations: shortening the standard settlement cycle," Office of the Comptroller of the Currency Bulletin 2024-3, January 17, https://tinyurl.com/mt95p88x

OECD, 2020, "The tokenisation of assets and potential implications for financial markets," Organization for Economic Co-operation and Development blockchain policy series

Omar, I. A., H. R. Hasan, R. Jayaraman, K. Salah, and M. Omar, 2021, "Implementing decentralized auctions using blockchain smart contracts," Technological Forecasting and Social Change 168, 120786

Onyx (J.P. Morgan) and Apollo, 2023, "The future of wealth management," https://tinyurl.com/4adpzv24

Perino, M. A., 2002, "Report to the Securities and Exchange Commission Regarding Arbitrator Conflict Disclosure Requirements in NASD and NYSE Securities Arbitrations," United States Securities and Exchange Commission

Plepi, A., and P. Schwendner, 2024, "A method for uncovering tokenisation archetypes and their effects: thus spoke Switzerland," SSRN working paper no. 5017084

Roncoroni, A., S. Battiston, M. D'Errico, G. Hałaj, and C. Kok, 2021, "Interconnected banks and systemically important exposures," Journal of Economic Dynamics and Control 133, 104266

Russell-Walling, E., 2022, "BNP Paribas embraces tokenisation for project finance bonds," The Banker, November 14, https://tinyurl.com/yvy6w8sw

SDX, 2024, "Project Helvetia III successfully completed, wCBDC pilot on SIX Digital Exchange to continue over the next two years," SIX Digital Exchange, June 20, https://tinyurl.com/mrwy4cxt

Shaikh, A. A., and H. Karjaluoto, 2015, "Mobile banking adoption: a literature review," Telematics and Informatics 32:1, 129-142

Storbeck, O., 2023, "Deutsche Bank continued to push risky derivatives years after probe found mis-selling," Financial Times, September 12, https://tinyurl.com/3tu37sx2

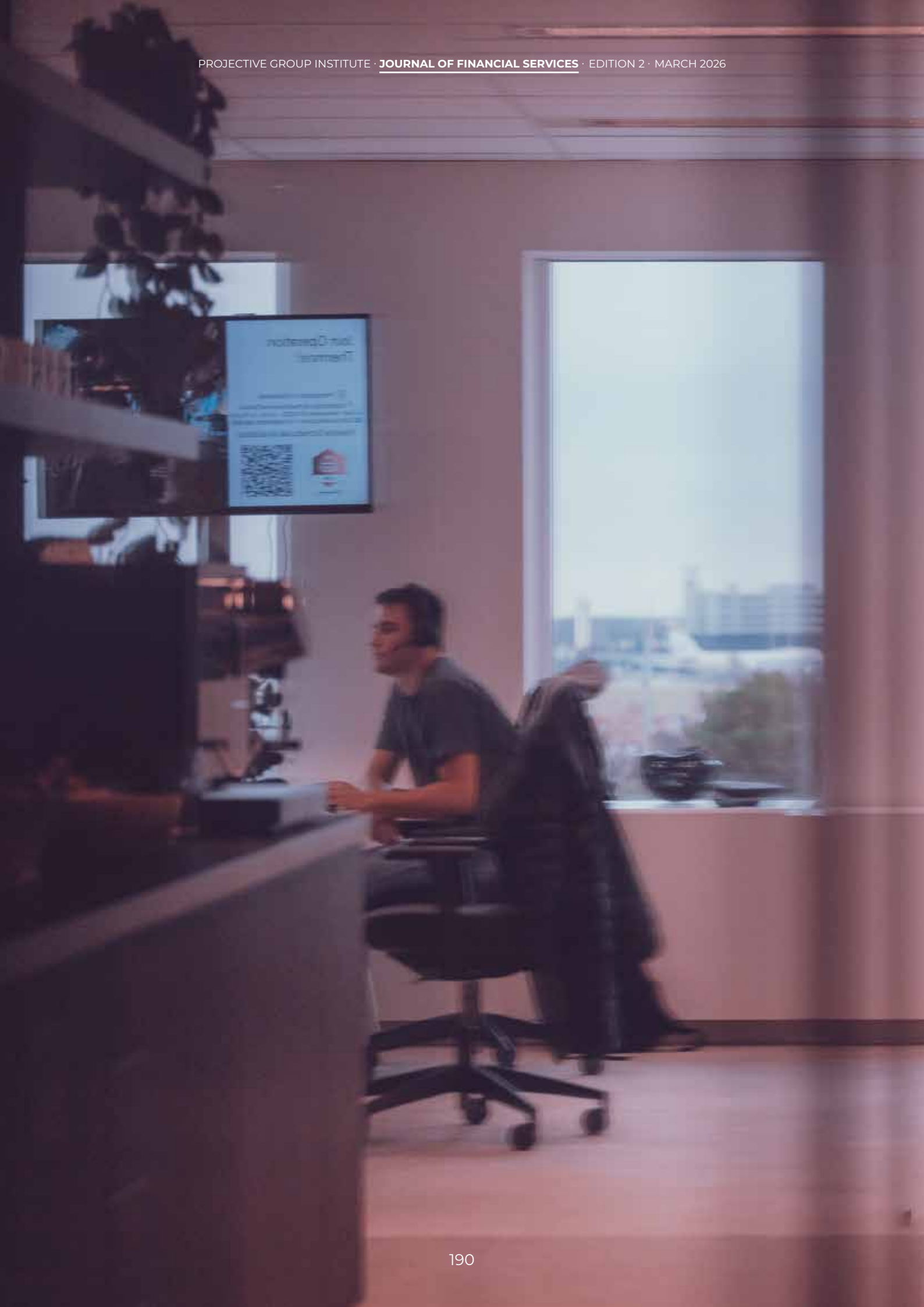
Tam, C., and T. Oliveira, 2017, "Literature review of mobile banking and individual performance," International Journal of Bank Marketing 35:7, 1044-1067

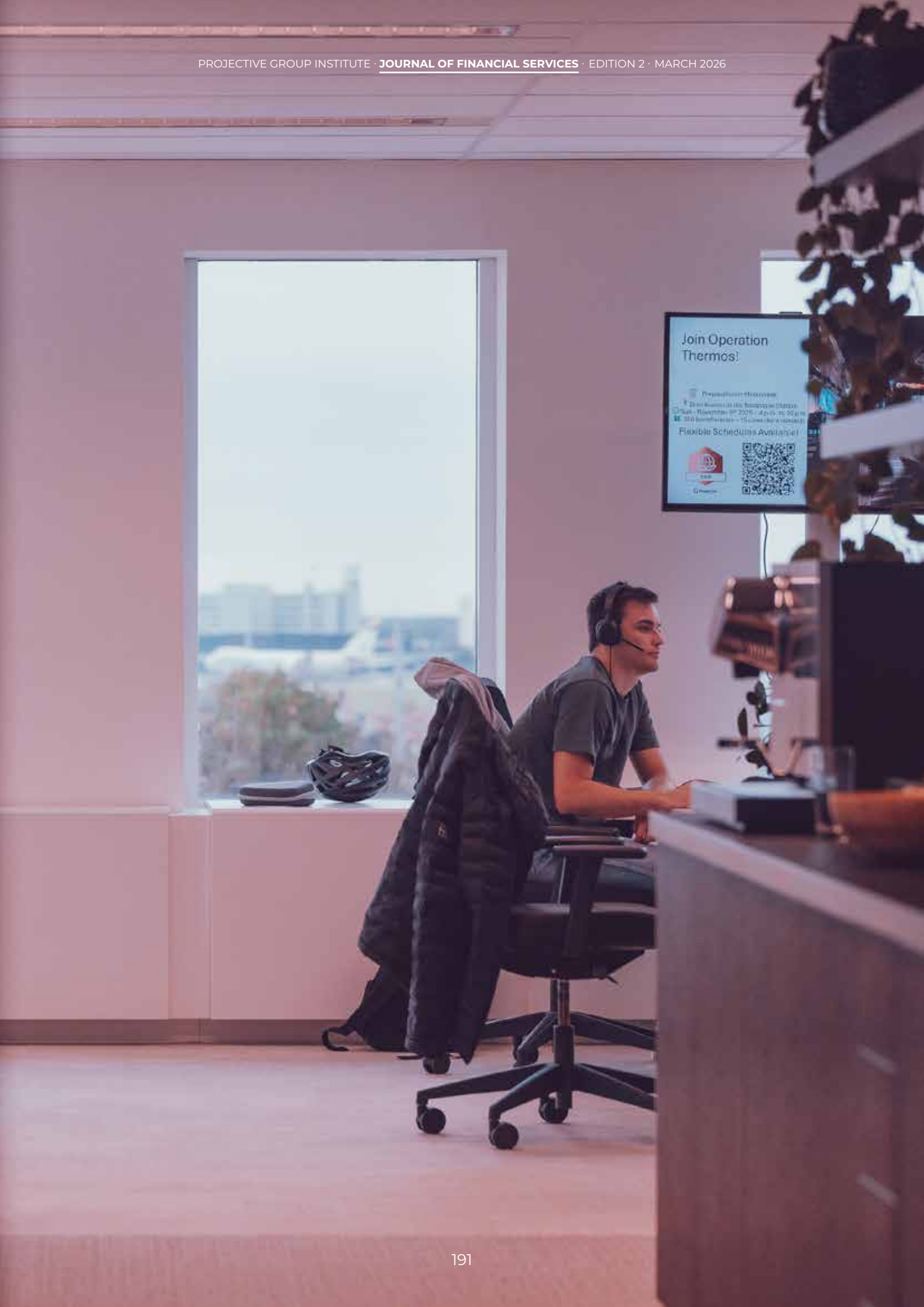
Tewes, M., M. Bauer, and G. Holz, 2023, "Security tokenization: how to unlock hidden value by moving stocks, bonds, and funds on the blockchain," Porsche Consulting

Townsend, R. M., and N. X. Zhang, 2023, "Technologies that replace a 'central planner'," AEA Papers and Proceedings 113, 257-262

Uhlig, Harald, 2022, "A luna-tic stablecoin crash," National Bureau of Economic Research working paper no. 30256

Van Gysegem, F., and K. De Patoul, 2021, "The tokenization of the economy and its impact on capital markets and banks," Roland Berger




Join Operation
Thermos!
Flexible Schedules Available!

Future of Data Management

Efficient underwriting using **agentic AI**

Mohammad Asif Ali,
Technical Lead,
PNC Financial Services Group, Inc.

ABSTRACT

Loan underwriting remains a critical yet resource-intensive function within financial services institutions, often constrained by manual workflows, rigid rule-based systems, and limited adaptability to evolving risk factors. This study investigates the application of an agentic artificial intelligence (AI) architecture for end-to-end underwriting automation, combining large language models (LLMs), retrieval-augmented reasoning (RAG), and robotic process automation (RPA). The proposed framework enables autonomous data ingestion, contextual policy evaluation, and dynamic decision support while maintaining human oversight at key control points. Experimental observations indicate meaningful improvements in processing efficiency, decision consistency, and operational scalability when compared to conventional underwriting approaches. The paper also discusses governance considerations, including explainability, regulatory alignment, and bias mitigation, positioning agentic AI as a practical and extensible foundation for next-generation underwriting systems.

1. Introduction

An agentic underwriting model extends beyond conventional automation by combining contextual reasoning, policy retrieval, and autonomous task execution. By integrating large language models (LLMs) with retrieval-augmented knowledge bases, the system evaluates applications against institutional guidelines and regulatory constraints in a consistent and explainable manner.

The architecture supports continuous orchestration of document analysis, data validation, credit evaluation, and recommendation generation. Robotic process automation (RPA) complements this reasoning layer by interfacing with external systems, extracting structured inputs, and executing downstream actions. Together, these components reduce manual intervention, minimize variability in decisioning, and enable underwriting teams to focus on exception handling and higher-value analysis rather than routine processing.

In this article, I will (1) examine the influence of agentic AI on the automation of underwriting processes; (2) compare AI-driven underwriting with traditional approaches regarding accuracy, processing time, and risk profiling; (3) analyze the contribution of Large Language Models (LLMs) and retrieval-augmented generation (RAG) in the decision-making process of underwriting; and (4) explore the aspects of AI explainability, compliance, and future improvements.

2. Background of agentic AI

Agentic AI is an emerging technology that is revolutionizing a wide range of industries. In order to create autonomous AI agents that can analyze data, set goals, and take actions with less human supervision, it combines enterprise automation, machine learning (ML), and LLMs. With each interaction, these agents can make decisions, solve problems dynamically, learn, and get better.

With its constant learning and output optimization, agentic AI is ideal for dynamic environments. In contrast to AI applications, which are frequently task-specific and excel in specialized fields like data analysis or image recognition, agentic AI manages intricate, multi-step workflows that require real-time contextual understanding and decision-making.

AI agents, RPA robots, and humans work together in a symbiotic relationship in agentic automation. People give the agents their objectives, maintain governance, and intervene when human review and judgment are needed (human in the loop). By gathering the data needed for AI agents to make decisions (such as logging in, connecting, and comprehending information across multiple systems), RPA robots make AI agents more accurate, productive, and successful. They can also carry out a variety of other predetermined tasks for agents.

2.1 Agentic AI levels

Agentic AI operates at a number of levels, with each level indicating a varying degree of independence and intricacy in the actions and learning capabilities of AI systems.

1. **Reactive agents**: operating under a fundamental sense-and-respond paradigm, reactive agents respond to present inputs without the ability to retain information or engage in planning.
2. **Proactive agents**: proactive agents surpass the act of responding by utilizing lessons learned from past experiences and formulating targeted actions to reach their goals.
3. **Adaptive agents**: through dynamic learning from their interactions, adaptive agents enhance their responses and strategies, allowing them to operate efficiently within complex environments.
4. **Fully agentic systems**: fully independent AI systems are designed to autonomously

determine their objectives, make choices, and perform activities in environments that are not structured.

2.2 Future of agentic AI

With the advancement of agentic AI, its possibilities will broaden into more complex applications:

1. **Advanced enterprise autonomy**: AI agents that go beyond mere execution to include strategic planning and process optimization.
2. **Cross-disciplinary integration**: merging domains like banking and finance, healthcare, and legal research to achieve more profound solutions to challenges.
3. **Collaborative human interaction**: AI functioning as an active collaborator, working in harmony with human teams.

3. Literature review

3.1 Traditional underwriting versus AI-driven underwriting

Traditional underwriting methods are based on rule-based decision-making, statistical analysis, and human oversight. Conversely, AI-enhanced underwriting utilizes ML models, natural language processing (NLP), and automation to boost efficiency and accuracy. Studies have demonstrated that AI underwriting can shorten loan approval times by 40-60% and increase the consistency of risk assessments by 30% [Bonnet (2025)].

3.2 Role of LLMs in Underwriting

The use of LLMs, like Mistral 7B, is on the rise in areas such as document analysis, information extraction, and automated support for decision-making. By leveraging RAG, these models facilitate adherence to underwriting policies and regulatory requirements during the decision-making process [Giulio et al. (2024), Pederson (2025)].

3.3 AI bias, explainability, and compliance in financial decision-making

One of the primary challenges associated with AI-driven underwriting is the presence of bias in model predictions, which can result from imbalances in the training datasets. Research underscores the necessity for interpretable AI models to comply with Fair Lending Practices and the Equal Credit Opportunity Acts. This study examines mitigation approaches, such as human-in-the-loop (HITL) oversight and federated learning, aimed at securing credit assessments [Lehigh University (2024), FinRegLab (2023), Vats et al. (2025)].

3.4 Problem statement

Traditionally, underwriting in the financial services industry has been reliant on human expertise, rule-based evaluations, and statistical models to determine loan eligibility, the ability to repay the borrowed funds, as well as evaluate insurance risks. While human underwriters provide critical domain knowledge, judgment, and ethical considerations, they are limited by factors such as subjectivity, inefficiency, extended processing times, and scalability challenges. Additionally, human-driven underwriting is at risk of biases, inconsistencies, and fatigue, which can negatively influence financial inclusion and risk management strategies.

With the advent of agentic AI, there exists a promising opportunity to develop autonomous, adaptive, and self-improving underwriting agents that can carry out real-time risk assessments with superior efficiency, consistency, and scalability compared to traditional human underwriters.

4. Why choose an agentic AI underwriter

The agentic AI underwriter has a very strong framework that leverages RAG with LLMs for faster indexation and knowledge retrieval. By harnessing Mistral AI services (Mistral 7B and MoE 8x7B), it creates a robust knowledge-driven AI agent that can generate intelligent suggestions and handle complex analysis and decision making. The data can be sourced from historical loan data, API integration with financial data sources, websites, policy documents, and regulatory guidelines.

By simulating human judgment, agentic AI underwriter reduces human intervention. AI agents prioritize tasks, allocate resources, and predict outcomes - implementing the decisions they make to move the process forward and achieve the desired outcome. It addresses the discrepancies like missing data or unexpected formats without human intervention.

Advantages:

- **Enhanced efficiency and accuracy**: conventional underwriting processes are characterized by their labor-intensive nature and susceptibility to human error. By automating the phases of data collection and analysis, the AI agent greatly enhances the efficiency of underwriting, enabling human underwriters to evaluate applications at an unparalleled speed while maintaining the integrity of risk assessment.
- **Automated risk profiling**: the agentic AI underwriter significantly influences the assessment of risk factors. By utilizing AI platforms, companies can swiftly gather and evaluate extensive datasets, including personal information and financial records, which enables them to develop a thorough understanding of an applicant's risk profile.
- **Uniformity in underwriting practices**: the agentic AI underwriter also promotes the uniformity of underwriting standards, guaranteeing that each application is assessed according to identical criteria. This consistency not only improves the equity of the underwriting process but also reduces the likelihood of human bias.
- **Reductions in operational costs**: the benefits of using the agentic AI underwriter go beyond the improvements in efficiency and accuracy, it can also help reduce operational costs by mitigating extensive manual intervention. Additionally, the improved precision in risk evaluations creates a more competitive, equitable pricing and robust environment.

5. Methodology

The agentic AI underwriter methodology integrates RPA, sophisticated AI, and human oversight to enhance and refine the loan underwriting process. The procedure commences with the gathering of applicant information through document submissions and email interactions. A software robot is responsible for downloading these documents and forwarding them to Mayan EDMS for data extraction, utilizing the Mistral 7B LLM to process and extract essential details such as the applicant's name, address, and income.

Concurrently, the bot accesses the applicant's credit score from credit bureaus via APIs. Additionally, it analyzes the sales representative's email to extract pertinent loan information. This collected data is then input into the Mistral 7B loan recommendation AI agent, which assesses the loan application in accordance with the bank's policies and guidelines stored within a RAG system, producing a recommendation based on compliance and risk evaluation.

The decision-making process bifurcates into two outcomes: approval or denial. In either scenario, a human-in-the-loop task enables a loan underwriter to review and confirm the AI-generated recommendation. Following the final decision, the bank's loan management system is updated through API, and automated email notifications are dispatched to all relevant parties, ensuring transparency and effective communication.

By merging automation, AI, and human oversight, this methodology significantly improves efficiency,

Table 1: The technology stack

Component	Technology/tool	Purpose
Programming language	Python	Core development, AI integration
RPA	Automation Anywhere/any other RPA tool	Download documents from the portal and email automation
Document automation system (DAS)	Mayan EDMS	Document storage, metadata extraction, and processing
AI LLM, Generative AI (GenAI)	Mistral 7B	Document extraction, email extraction, loan document analysis, risk assessment, and natural language understanding.
	RAG System	Stores bank policies and loan criteria used for evaluation by AI Agent.

shortens processing times, and upholds compliance, resulting in a robust and scalable loan underwriting solution.

5.1 Design flow

The agentic AI underwriter follows the following steps:

1. **Data collection**
   - → The loan applicant uploads the required documents (W2, pay stubs, bank statements, address and ID proof, photograph, etc.) to the loan application portal.
   - → The sales representative emails the bank loan department with key loan details, including purchase price, taxes, down payment, and the applicant's income.
2. **Document extraction**
   - → A software robot (bot) downloads the documents from the loan application portal.
   - → The documents are sent to Mayan EDMS, a DAS for data extraction (e.g., applicant's name, address, income, etc.) using Mistral 7B – an LLM.
3. **Credit score**
   - → The bot pulls the applicant credit score from credit bureaus via APIs (e.g. Experian API, Equifax API) using Python.
4. **Email extraction**
   - → A Mistral 7B GenAI-powered email extractor processes the sales representative's email.
   - → It extracts the relevant loan details (loan amount, income, down payment, taxes, etc.).
5. **Loan recommendation**
   - → The extracted data (loan amount, income, credit score, etc.) is sent to the Mistral 7B loan recommendation AI agent.
   - → The AI agent evaluates the application against bank policies, and loan criteria stored in a RAG system using Mistral 7B.
   - → Generates a loan recommendation based on compliance and risk assessment.
6. **Decisioning**
   - → Approval process: (1) if the AI agent recommends approval, it highlights risk and proposed loan terms; (2) a human-in-loop task is triggered, allowing a loan underwriter to review and validate the recommendation; and (3) the underwriter submits the final decision, initiating further action.
   - → Denial process: (1) if the AI agent detects any non-compliance policies, it recommends rejection/denial; and (2) similar to the approval task, a human review task is triggered to validate the denial decision.
7. **Communication**
   - → The agentic AI system updates the bank's loan management system with the final loan status using API.
   - → The bot sends automated emails to the respective stakeholders (applicant, sales rep, loan department) regarding the approval/denial.

5.2 System flow summary

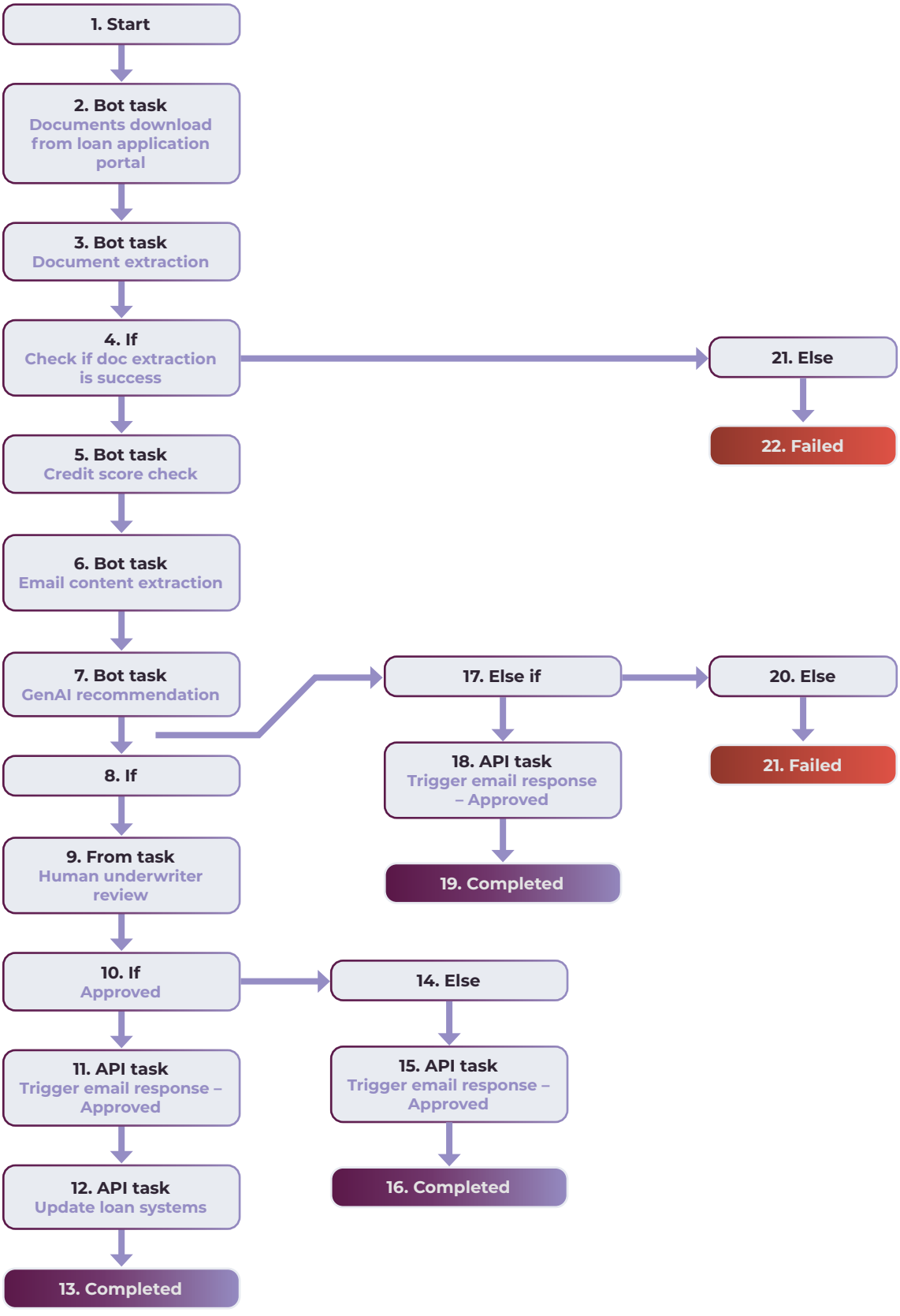
1. **Data collection** → loan application submission.
2. **Document extraction** → OCR and structured data extraction.
3. **Credit score retrieval** → API calls to credit bureaus using Python.
4. **Email extraction** → Mistral 7B GenAI for email processing.
5. **Loan recommendation** → AI-driven risk assessment and recommendation using Mistral 7B and RAG.
6. **Decisioning** → Human review triggered for approval/denial.
7. **Communication** → Automated notifications and system updates using API.

6. Conclusion and future work

This study illustrates how agentic AI can be leveraged to modernize underwriting operations by integrating autonomous reasoning, intelligent automation, and human oversight within a unified framework. The findings indicate improvements in operational efficiency, decision consistency, and system scalability, while preserving regulatory compliance and transparency.

The adoption of agentic AI enables financial institutions to develop more agile, transparent, and scalable underwriting systems, supporting broader loan accessibility and contributing to improved financial inclusion. By automating routine decision workflows and enabling contextual policy evaluation, such systems can reduce processing delays while maintaining responsible risk controls.

Figure 1: Process flow diagram



Despite these benefits, several challenges remain, including model explainability, bias mitigation, and regulatory validation at scale. Future research should focus on enhanced interpretability methods, privacy-preserving learning approaches such as federated learning, and long-term evaluations across diverse underwriting portfolios. As agentic systems continue to evolve, they are likely to play a central role in intelligent financial operations, balancing innovation with ethical and governance requirements.

Addressing these areas will further strengthen the applicability of agentic AI in underwriting, enabling sustainable automation that aligns efficiency gains with fairness, accountability, and responsible AI governance.

References

Acharya, D. B., B. Divya, and K. Kuppan, 2024, "Explainable and fair AI: balancing performance

in financial and real estate machine learning models," IEEE Access, https://tinyurl.com/46kp6zjt

Acharya, D. B., K. Kuppan, and B. Divya, 2025, "Agentic AI: autonomous intelligence for complex goals – a comprehensive survey," IEEE Access 13, https://tinyurl.com/um25uwbc

Ashok, P., T. P. Godwin James, G. Malathi, A. Kavinilavu, and L. Sharmila, 2022, "Development of truly intelligent autonomous agents based on classical AI using complex representations," Published in: 2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), https://tinyurl.com/tf6v48kf

Auomation Anywhere, 2025, "What is agentic AI?," https://tinyurl.com/yjuu3b5n

Bonnet, O., 2025, "Comparative study of traditional versus AI-driven underwriting efficiency," working paper, University of Florida, https://tinyurl.com/5yaud7yr

Caballar, R. D., 2024, "What Are AI agents? Here's how AI agents work, why people are jazzed about them, and what risks they hold," IEEE Spectrum, November 19, https://tinyurl.com/yvnzmn5w

FinRegLab, 2023, "Explainability and fairness in machine learning for credit underwriting: policy analysis," December, https://tinyurl.com/46ay7sdp

Fourney, A., et al., 2025, "AutoGen v0.4: reimagining the foundation of agentic AI for scale, extensibility, and robustness," Microsoft research blog, January 14, https://tinyurl.com/4uvrpjjx

Glozman, R., 2020, The digitally-enabled underwriter, Wiley

Giulio, B., B. Marie, G. Claudia, M. Carsten, and S. Alpay, 2024, "Leveraging large language models in finance: pathways to responsible adoption," European Securities and Markets Authority (ESMA)/Alan Turing Institute/Institut Louis Bachelier, https://tinyurl.com/yc6yrn2z

Ingram, W. A., B. Banerjee, and E. A. Fox, 2024, "Agentic AI for improving precision in identifying contributions to sustainable development goals," https://tinyurl.com/2fwbzvuf

Lehigh University, 2024, "AI exhibits racial bias in mortgage underwriting decisions, researchers find," August 21, https://tinyurl.com/yf7nvdhf

Kissinger, H. A., E. Schmidt, and D. Huttenlocher, 2021, The age of AI: and our human future, Little, Brown and Company

Merritt, R., 2025, "What is retrieval-augmented generation, aka RAG?," Nvidia blog, January 31, https://tinyurl.com/msz6h95w

Pederson, J., 2025, "Retrieval-augmented generation (RAG)," Pinecone, June 12, https://tinyurl.com/mryx6a7p

Royce, T., 2024, Agentic AI: harnessing the power of autonomous systems in the modern workplace, independently published

Safransky, T. H., 2024, The AI advantage: how to use AI to underwrite multifamily development, independently published

Singh, A., A. Ehtesham, S. Kumar, and T. Talaei Khoei, 2024, "Enhancing AI systems with agentic workflows patterns in large language model," presented at 2024 IEEE World AI IoT Congress (AIIoT), https://tinyurl.com/bdzfmzkm

UiPath, 2025, "What is agentic AI?," https://tinyurl.com/44vh6tds

Vats, R., S. Agrawal, and S. S. Chippada, 2025, "Bias detection and fairness in large language models for financial services," International Journal of Scientific Research in Computer Science Engineering and Information Technology 11:2, 1329-1345

Appendix
A1 Technical design

This section explains the implementation approach for the agentic AI underwriter.

A1.1 Step 1:
Download documents from loan application portal using Python API integration

1. **Setup and configuration**
   - → Define the base_url for the API.
   - → Set up the API_KEY for authentication.
   - → Create a directory for storing downloaded documents.
2. **Initialize API session**
   - → Create a session object with headers containing the Authorization token.
3. **Fetch loan applications**
   - → Send a GET request to the /applications endpoint.
   - → Parse the response to get a list of application IDs.
4. **Fetch documents for each application**
   - → For each application ID (1) send a GET request to /applications/{application_id}/documents and (2) parse the response to get a list of document metadata (e.g., document ID, type).
5. **Download Documents**
   - → For each document in the list (1) send a GET request to /documents/{document_id}/download and (2) save the document to the local directory with a unique filename (e.g., including application ID, document type, and timestamp).
6. **Send Documents to Mayan EDMS – Document Automation System (DAS)**
   - → For each downloaded document (1) open the file and prepare it for upload, (2) send a POST request to Mayan EDMS via its REST API to the upload endpoint with the document file and metadata (e.g., application ID, document type), and (3) log the response from the automation system.
7. **Logging**
   - → Log all activities, including successful downloads and errors.
8. **Create document index**
   - → Generate a CSV file containing metadata for all downloaded documents (e.g., application ID, document type, filename).
9. **Error handling**
   - → Handle network errors, authentication failures, and invalid responses gracefully.

A1.2 Step 2: Documents extraction using Mistral 7B LLM

Load the Mistral 7B LLM model	model = load_mistral_7b_model()
Preprocess the document for input (text extraction)	text = preprocess_document(document)
Send the document text to Mistral 7B model for data extraction	extracted_data = model.extract_information(text)
Extract specific data points from the model output	applicant_name = extracted_data["name"] applicant_address = extracted_data["address"] applicant_income = extracted_data["income"] applicant_id = extracted_data["id"] applicant_photo = extracted_data["photo"]
Return the extracted data	return { "name": applicant_name, "address": applicant_address, "income": applicant_income, "id": applicant_id, "photo": applicant_photo }
Load the pre-trained Mistral 7B model (from local storage or server)	model = load_model("Mistral_7B") return model
Step to extract text from document if it's in PDF, image, or other formats	if document is image: text = extract_text_from_image(document) else if document is PDF: text = extract_text_from_pdf(document) else: text = document.text
Clean and normalize the extracted text	clean_text = clean_text(text) return clean_text
Remove unnecessary characters and normalize the text	cleaned_text = remove_special_characters(text) cleaned_text = correct_ocr_errors(cleaned_text) return cleaned_text
Use an OCR tool like Tesseract to extract text from an image	return ocr_tool.extract_text(image)
Extract text from the PDF using a PDF parser	return pdf_parser.extract_text(pdf)

A1.3 Step 3: Credit score pull from credit bureaus' API using Python

Get access token

```
import requests
import json
from datetime import datetime

def authenticate_equifax(api_key, client_id):
    """
    Authenticate with Equifax API and retrieve an access token.
    """
    print("Authenticating with Equifax API...")
    auth_url = "https://api.equifax.com/v1/auth"
    headers = {
        'client_id': client_id,
        'api_key': api_key
    }

return "mock_equifax_token"
```

Using the access token,
get the Equifax score
	def get_equifax_credit_score(access_token, applicant_info):
	“””
	Retrieve credit score from Equifax API.
	“””
	print(“Retrieving credit score from Equifax API…”)
	endpoint = “https://api.equifax.com/v1/credit-score”
	headers = {
	‘Authorization’: f’Bearer {access_token}’,
	‘Content-Type’: ‘application/json’
	}
	
	payload = {
	‘consumerPII’: {
	‘name’: {
	‘firstName’: applicant_info[‘first_name’],
	‘lastName’: applicant_info[‘last_name’]
	},
	‘ssn’: applicant_info[‘ssn’],
	‘dateOfBirth’: applicant_info[‘dob’],
	‘currentAddress’: {
	‘street’: applicant_info[‘address’][‘street’],
	‘city’: applicant_info[‘address’][‘city’],
	‘state’: applicant_info[‘address’][‘state’],
	‘zip’: applicant_info[‘address’][‘zip’] } } }

Equifax API credentials. Authenticate and retrieve credit score.	def main(): api_key = “EQUIFAX_API_KEY” client_id = “ EQUIFAX_CLIENT_ID” access_token = authenticate_equifax(api_key, client_id) credit_score = get_equifax_credit_score(access_token, applicant_info)
Output	print(“\ Equifax Credit Score Result:”) print(json.dumps(credit_score, indent=2))

Example of applicant information	applicant_info = {
	‘first_name’: ‘Mohammad Asif’,
	‘last_name’: ‘Ali’,
	‘ssn’: ‘333-44-6666’, # In real implementation: Handle
	SSN securely!
	‘dob’: ‘1990-01-01’,
	‘address’: {
	‘street’: ‘620 Liberty Ave,
	‘city’: ‘Pittsburgh’,
	‘state’: ‘PA’,
	‘zip’: ‘16046’
	}
	}

Example of output/response	return {
	‘bureau’: ‘equifax’,
	‘credit_score’: 750,
	‘score_type’: ‘FICO’,
	‘score_date’: ‘2025-01-28T15:20:45’,
	‘status’: ‘success’
	}

A1.4 Step 4: Email extraction using Mistral 7B GenAI extractor (Hugging Face Library)

Install Hugging Face transformers library using Bash command	pip install transformers torch
Import Statement	from transformers import AutoModelForCausalLM, AutoTokenizer import torch

Load Mistral 7B model and token from Hugging Face	def extract_loan_details_from_email(email_text): # model_name = “mistralai/Mistral-7B”
	tokenizer= AutoTokenizer.from_pretrained(model_name)
	model= AutoModelForCausalLM.from_pretrained(model_name, torch_dtype=torch.float16, device_map=”auto”)

Define structured prompt	prompt = f””” Extract key loan details from the following email: - Loan Amount - Income - Down Payment - Taxes - Other relevant financial information Email Content: {email_text} “””
Input tokenization	input_tokens= tokenizer(prompt, return_tensors=”pt”).to(“cuda”)
Generate response using Mistral 7B	output_tokens = model.generate(**input_tokens, max_length=512)
Decode response	extracted_data = tokenizer.decode(output_tokens[0], skip_special_tokens=True)
Return statement	return extracted_data
Example	email_text = “”” Hello Loan Dept, The applicant has an annual income of $115,000 and is applying for a $70,000 loan. The down payment is $10,000, and property taxes are estimated at $2,000 per year. Regards, Sales Rep “””
Output	loan_details = extract_loan_details_from_email(email_text) print(“Extracted Loan Details:\n”, loan_details) yaml Extracted Loan Details: Loan Amount: $70,000 Income: $115,000 Down Payment: $10,000 Taxes: $2,000 per year Other relevant financial information: None mentioned

A1.5 Step 5: Loan recommendation integrated with RAG

Import statement	from transformers import AutoModelForCausalLM, AutoTokenizer import torch
Load Mistral 7B loan recommendation AI agent	model_name = “mistralai/Mistral-7B” tokenizer= AutoTokenizer.from_pretrained(model_name) model= AutoModelForCausalLM.from_pretrained(model_name, torch_dtype=torch.float16,device_map=”auto”) def generate_loan_recommendation(loan_data, bank_policies): “”” Evaluates loan application against bank policies and generates a recommendation. Args: loan_data (dict): Extracted loan details (loan amount, income, credit score, etc.). bank_policies (str): Retrieved bank policies from the RAG system. Returns: str: Loan recommendation response. “””
Design structured prompt for AI evaluation	prompt = f””” You are an AI Loan Underwriter evaluating a loan application. Loan Application Details: - Loan Amount: ${loan_data[‘loan_amount’]} -Income: ${loan_data[‘income’]} -Credit Score: {loan_data[‘credit_score’]} -Down Payment: ${loan_data[‘down_payment’]} -Taxes: ${loan_data[‘taxes’]} Bank Policies and Loan Criteria: {bank_policies} Based on the above details, assess the compliance and risk of this loan application. Provide a structured recommendation including: - Approval or Denial decision. - Reasoning based on risk assessment. - Proposed loan terms (if approved). “””
Input tokenization	input_tokens= tokenizer(prompt, return_tensors=”pt”).to(“cuda”)
GenAI response	output_tokens= model.generate(**input_tokens, max_length=512)
Decode response	recommendation = tokenizer.decode(output_tokens[0], skip_special_tokens=True)
Return statement	return recommendation

Example	loan_data = {“loan_amount”: 70000, “income”:115000, “credit_score”: 750, “down_payment”: 10000, “taxes”: 2000 }
Retrieved bank policies from RAG system	bank_policies = “”” Minimum Credit Score: 650 Maximum Debt-to-Income Ratio: 35% Loan-to-Value (LTV) Ratio should not exceed 75%. “””
Generate loan recommendation	recommendation= generate_loan_recommendation(loan_data, bank_policies)
Print AI Loan Recommendation	print(“Loan Recommendation:\n”, recommendation)

A1.6 Step 6: Decisioning approval/denial

Import statements	from mistralai.client import MistralClient from airflow import DAG from airflow.operators.python import PythonOperator from datetime import datetime

AI model	api_key=
	os.getenv(“MISTRAL_API_KEY”, “actual_api_key”) # Fetch API Key from environment variables
	ai_client = MistralClient(api_key=api_key) # Initialize AI model with dynamic API key
	
	def create_human_review_task(application_data, ai_decision, risk_assessment, proposed_terms=None):
	
	“””Creates a human-in-the-loop task for review.”””
	task_data = {
	“application_data”: application_data,
	“ai_decision”: ai_decision,
	“risk_assessment”: risk_assessment,
	“proposed_terms”: proposed_terms,
	}
	print(f”Human review task created with data: {task_data}”)
	
	class LoanUnderwriterAgent:
	def __init__(self, application_data):
	self.application_data = application_data
	self.ai_decision = None
	self.risk_assessment = None
	self.proposed_terms = None
	def evaluate_application(self):
	“””Use Mistral 7B to analyze loan application and provide decisioning.”””
	prompt = f”””
	Evaluate the following loan application:
	{self.application_data}
	Provide a decision (Approve/Denial), risk assessment, and loan terms if approved.

```
"""
        response = ai_client.generate(model="mistral-7b",
prompt=prompt, max_tokens=200)
        self.ai_decision = response.get("decision")
        self.risk_assessment = response.get("risk_assessment", "No
risk assessment available")
        self.proposed_terms = response.get("proposed_terms", "No
terms available")
        return response

    def process_decision(self):
        """Trigger human review based on AI decision."""
        create_human_review_task(self.application_data, self.
ai_decision, self.risk_assessment, self.proposed_terms if self.
ai_decision == "Approve" else None)

    def finalize_decision(self, human_decision):
        """Finalizes the decision based on human review."""
        if human_decision not in ["Approve", "Deny"]:
            raise ValueError("Invalid final decision")
        print(f"Final decision submitted: {human_decision}")
        return human_decision
```

Define airflow DAG to integrate the loan underwriting workflow into Apache Airflow for workflow automation.

```
with DAG("loan_underwriting", start_date=datetime(2025, 2, 1),
schedule_interval=None, catchup=False) as dag:
    # Get application data dynamically (from arguments)
    application = os.getenv("APPLICATION_DATA")  # Get from
environment variable or configuration

    if not application:
        raise ValueError("Application data not provided")

    underwriter = LoanUnderwriterAgent(application)

    evaluate_task = PythonOperator(
        task_id="evaluate_application",
        python_callable=underwriter.evaluate_application
    )

    process_task = PythonOperator(
        task_id="process_decision",
        python_callable=underwriter.process_decision
    )

    evaluate_task >> process_task
```

Example to call the code – Approved case	if __name__ == "__main__":
	application_data_approve = {
	"applicant_name": "Jo Smith",
	"credit_score": 690,
	"income": 115000,
	"loan_amount": 80000,
	"policy_compliance": True
	}
	
	underwriter_approve= LoanUnderwriterAgent(application_data_approve)
	decision_approve= underwriter_approve.evaluate_application()
	print(f"AI Decision: {decision_approve}")
	underwriter_approve.process_decision()
	final_decision_approve="Approve" underwriter_approve.finalize_decision(final_decision_approve)

Example to call the code – Denied case	application_data_deny = {
	"applicant_name": "Adam Stone",
	"credit_score": 550,
	"income": 35000,
	"loan_amount": 550000,
	"policy_compliance": False
	}
	
	underwriter_deny= LoanUnderwriterAgent(application_data_deny)
	decision_deny = underwriter_deny.evaluate_application()
	print(f"AI Decision: {decision_deny}")
	underwriter_deny.process_decision()
	final_decision_deny = "Deny"
	underwriter_deny.finalize_decision(final_decision_deny)

Expected output – Approved case	AI Decision: {'decision': 'Approve', 'risk_assessment': 'Low risk due to good credit score and policy compliance', 'proposed_terms': 'Interest rate 4.5%, 15-year term'} Human review task created with data: {'application_data': {'applicant_name': 'Jo Smith', 'credit_score': 690, 'income': 115000, 'loan_amount': 80000, 'policy_compliance': True}, 'ai_decision': 'Approve', 'risk_assessment': 'Low risk due to good credit score and policy compliance', 'proposed_terms': 'Interest rate 4.5%, 15-year term'} Final decision submitted: Approve
Expected output – Denied case	AI Decision: {'decision': 'Deny', 'risk_assessment': 'High risk due to low credit score, insufficient income, high loan amount, and non-compliance with policy', 'proposed_terms': 'N/A'} Human review task created with data: {'application_data': {'applicant_name': 'Adam Stone', 'credit_score': 550, 'income': 35000, 'loan_amount': 550000, 'policy_compliance': False}, 'ai_decision': 'Deny', 'risk_assessment': 'High risk due to low credit score, insufficient income, high loan amount, and non-compliance with policy', 'proposed_terms': 'N/A'} Final decision submitted: Deny

A1.7 Step 7: Communication

A1.7.1 Update loan status to Loan Management System using API

Header	import requests def update_loan_status(loan_id, final_decision): api_url= “https://api.loan-management-system.com/update_status” api_key= “actual_api_key” # Use environment variable for security headers = { “Authorization”: f”Bearer {api_key}”, “Content-Type”: “application/json” }

Payload	payload =
	{ “loan_id”: loan_id,
	“status”: final_decision }
	
	response = requests.post(api_url, json=payload, headers=-headers)
	
	if response.status_code == 200:
	print(f”Loan status updated successfully for Loan ID: {loan_id}”)
	else:
	print(f”Failed to update loan status for Loan ID: {loan_id}. Error: {response.text}”)

Example	update_loan_status(“33456745”, “Approved”) update_loan_status(“52256745”, “Denied”)

A1.7.2 Python code to trigger Automation Anywhere for email notification

Sends an email using an Automation Anywhere bot.	import requests def send_email_via_bot(subject, recipient_email, message_body): bot_url= “https://bot.api.automationanywhere.com/send_email” bot_api_key = “actual_bot_api_key” # Use environment variable for security

Header	headers = {
	“Authorization”: f”Bearer {bot_api_key}”,
	“Content-Type”: “application/json”
	}

Payload	payload = {
	“subject”: subject,
	“recipient_email”: recipient_email,
	“message_body”: message_body }
	
	response = requests.post(bot_url, json=payload, headers=-headers)
	
	if response.status_code == 200:
	print(f”Email successfully sent to {recipient_email}”) else:
	print(f”Failed to send email to {recipient_email}. Error: {response.text}”)

Usage example	def notify_stakeholders(loan_decision, applicant_email, sales_rep_email, loan_dept_email):
	
	# Dynamic email messages
	subject = f"Loan Application Decision: {loan_decision}" message_body = f"Dear Stakeholder, the loan application has been {loan_decision}. Please review the details."
	#Send to applicant
	send_email_via_bot(subject, applicant_email, message_body)
	
	# Send to sales rep send_email_via_bot(subject, sales_rep_email, message_body)
	
	# Send to loan department send_email_via_bot(subject, loan_dept_email, message_body)

Example	notify_stakeholders("Approved", "applicant@ymail.com","salesrep@ymail.com", "loan_dept@ymail.com") notify_stakeholders("Denied", "applicant@ymail.com","salesrep@ymail.com", "loan_dept@ymail.com")

A1.7.3 Integrate Loan Management System and email notification

Update loan statis in LMS	def finalize_and_notify(loan_id, final_decision, applicant_email, sales_rep_email, loan_dept_email): update_loan_status(loan_id, final_decision)
Email notification to stakeholders	notify_stakeholders(final_decision, applicant_email, sales_rep_email, loan_dept_email)
Example	finalize_and_notify("443453543","Approved", "applicant@ymail.com","salesrep@ymail.com", "loan_dept@ymail.com") finalize_and_notify("653453543","Denied", "applicant@ymail.com","salesrep@ymail.com", "loan_dept@ymail.com")

Future of Data Management

Agentic AI in banking: between narratives and statistics

Udo Milkau,
Digital Counsellor, Frankfurt, Germany[1]

ABSTRACT

The vision of "autonomous agents" powered by Artificial Intelligence (AI) is far from new, beginning in parallel with the early advances in AI in the 1950/60s and has earlier roots in science fiction literature. At present, there is a lot of hype around "agentic AI" based on large language models (LLM), which are believed to be able to plan, understand, or even reason. However, all LLMs are statistic tools for estimating a "next best token" for a given prompt, typically accompanied with an error, which in turn accumulates for longer runs even for small error probabilities. The contemporary concept of AI agents is a ruled-based extension of LLMs with pre- and postprocessing, which extend the fundamental capabilities of LLMs as "generators of fiction". However, recent claims that such systems could rewrite the rules of business and banking are based more on speculation than quantitative results. Implementations in industries such as banking cannot avoid the generic limitations of such statistical tools and have to be evaluated based on their statistical quality and improvements in performance as compared to existing process automation.

1. This paper was written without any support from GenAI or LLMs and represents the opinions of the author without any conflict of interest.

1. Introduction: Agentic AI as the next hype?

The launch of OpenAI's ChatGPT in November 2022 as a "chatbot for everybody" brought about public awareness about generative artificial intelligence (GenAI) and large language models (LLM). Now "agentic AI" or "AI agents" seem to be the next frontier. The hype is driven by the narrative that LLM-based AI agents could solve problems autonomously and "pursue own goals" without human intervention. The vision of a "general problem solver" is not new and was articulated nearly 70 years ago by Simon et al. (1958). While isolated formalized problems could be handled, the vision failed for real-world problems because of the so-called "combinatorial explosion".

To evaluate the actual potential of contemporary AI agents to solve real-world problems, a detailed assessment is necessary. Given that LLM-based agentic AI models are based on fundamental building blocks, this analysis begins with a tour de force of the foundations of AI via current LLMs and ends with the concept of AI agents.[2]

The focus of this paper is not on technical details, but on providing insights into the foundations of AI agents in statistics. Our main objective is to determine whether such AI agents do in fact differ from traditional business process automation and whether they deliver additional economic benefits to financial services organizations.[3]

2. Artificial intelligence as "statistical estimator"

The roots of AI can be traced back to the 1920s, though the term itself is believed to have been first coined by McCarthy et al. in 1955; the vision for which was "... to make machines ... solve kinds of problems now reserved for humans, and improve themselves."[4] The term "robot" is believed to have been mentioned for the first time by Karel Čapek in his 1920s science fiction play, titled: "Rossum's universal robots".

In Čapek's play, inconsistencies between rules and reality causes a rebellion against mankind and humans face an existential threat from the very machines they helped create; a narrative that was continued in a number of other science fiction stories, including Dennis Feltham Jones' 1966 classic, "Colossus", in which supercomputers take control of mankind.

The first deliberations on a rule-based dilemma can be found in Isaac Asimov's 1942 short story "Runaround" about "The three laws of robotics". However, the vision of rule-based (symbolic-logic) AI faded in the so-called AI winter in the 1980s, as described in Winograd and Flores 1986 book: "Understanding computers and cognition".

Cybernetics [Norbert Wiener (1948)], which is also part of the annals of AI, is the predecessor of machine learning as a statistical tool to derive classifications from "measured" data.

Sadly, both artificial intelligence and machine learning are viewed by many as anthropomorphistic,[5] even though they have purely technical meanings. Likewise, their technical implementation with "artificial neural networks" (ANN) is often misinterpreted as "brain-like", even though ANNs lack the complexity of the brain of a tiny mouse. Furthermore, terms like "self-learning", as opposed to "unsupervised learning", suggest a degree of autonomy that only fictional robots in science fiction stories have.

To explain the difference between supervised and unsupervised learning within an AI environment, I will use the example of image recognition versus text generation. "Supervised learning" refers to sit-

2. Contemporary AI based on artificial neural networks (ANNs) can be characterized as unidirectional input-inference-output systems, but they are neither adaptive control loops nor workflow engines, nor systems to "understand" a real-world context for that matter. Consequently, LLM-based AI agents need to be enhanced with traditional rule-based - i.e., programmed - software elements.
3. Reuters (2025) reported that OpenAI expects to exceed U.S.$125 billion in revenue in 2029. As the majority of this revenue, in addition to the revenues of other LLM providers, would have to be paid by corporate customers, any cost-benefit calculation for real-world cases is challenging.
4. This clear focus is astonishing, as the Paris conference in 1951 about "Les machines à calculer et la pensée humaine" missed that the future of computers would be stored-program digital machines.
5. The sociotechnical problem that users regard chatbots as "human-like", which many regard as being in response to Joseph Weizenbaum's first chatbot, ELIZA, in 1966, which just mirrored questions into answers.

uations where external labels are used to train the "learner", whereas "un-supervised learning" is when the generation of a "next best token" is conditioned on the previous text[6,7]. In the case of the latter, the training is done with existing text sequences, which are partially masked. On the other hand, "reinforcement learning" is used to describe games in which two computer programs play against each a number of times, with the "reward" for winning being maximized.

Irrespective of the type of learning one refers to, it should be borne in mind that the objective is always to estimate probabilities. They are constrained to make reasonable estimations only within the boundaries of the original data. There is no generalization beyond the original manifold without a causal model on how to extrapolate into uncharted territory or, at least, a testable assumption about the deterioration of statistical quality outside the original data.

A "statistical learner" is never right or wrong, it simply provides estimations based on assumptions, which can be compared to the actual results.[8] This leads to what is referred to as the "confusion matrix", which compares a binary prediction (yes or no) with the actual result (right or wrong) and distinguishes the cases of "true positive", "false positive", "true negative" and "false negative" (TP, FP, TN, FN). For example, mammography with mass screening for rare conditions in a population of "healthy" women under a certain age without additional risk factors can show a "good" detection quality, but a significant number of "false positive" results would unnecessarily concern these women [Pearl and Mackenzie (2018)]. Looking for rare events within a large number of "good" events - including payment fraud - one has to balance the benefits of detection against potential damages caused by erroneous classifications. For example, should one aim for high "sensitivity" [TP/(TP+FN)], i.e., the best possible detection of an illness or fraud, or high "specificity" [TN/(TN+FP)], i.e., the best possible separation of healthy patients or fraud grouping?

The quality of any "statistical estimation" has to be evaluated in the same way that randomized controlled trials (RCT) are undertaken: using a test group as well as a control group. As any learner has to be trained on classifications provided by human experts, no tool can best these experts [Liu et al. (2019)]. However, there are many situations when top experts are not available, or the number of patients exceeds capacity. Any human (sic!) decision about an implementation of a statistical estimator depends on the statistical quality of the tool, the objectives to be achieved, a cost-benefit analysis, a comparison with the actual circumstances, and whether a (preliminary) generalization out-of-distribution might be indicated – like in the early phase of the Covid19 pandemic with limited data but high risk for the population.

3. LLMs as generators of fiction

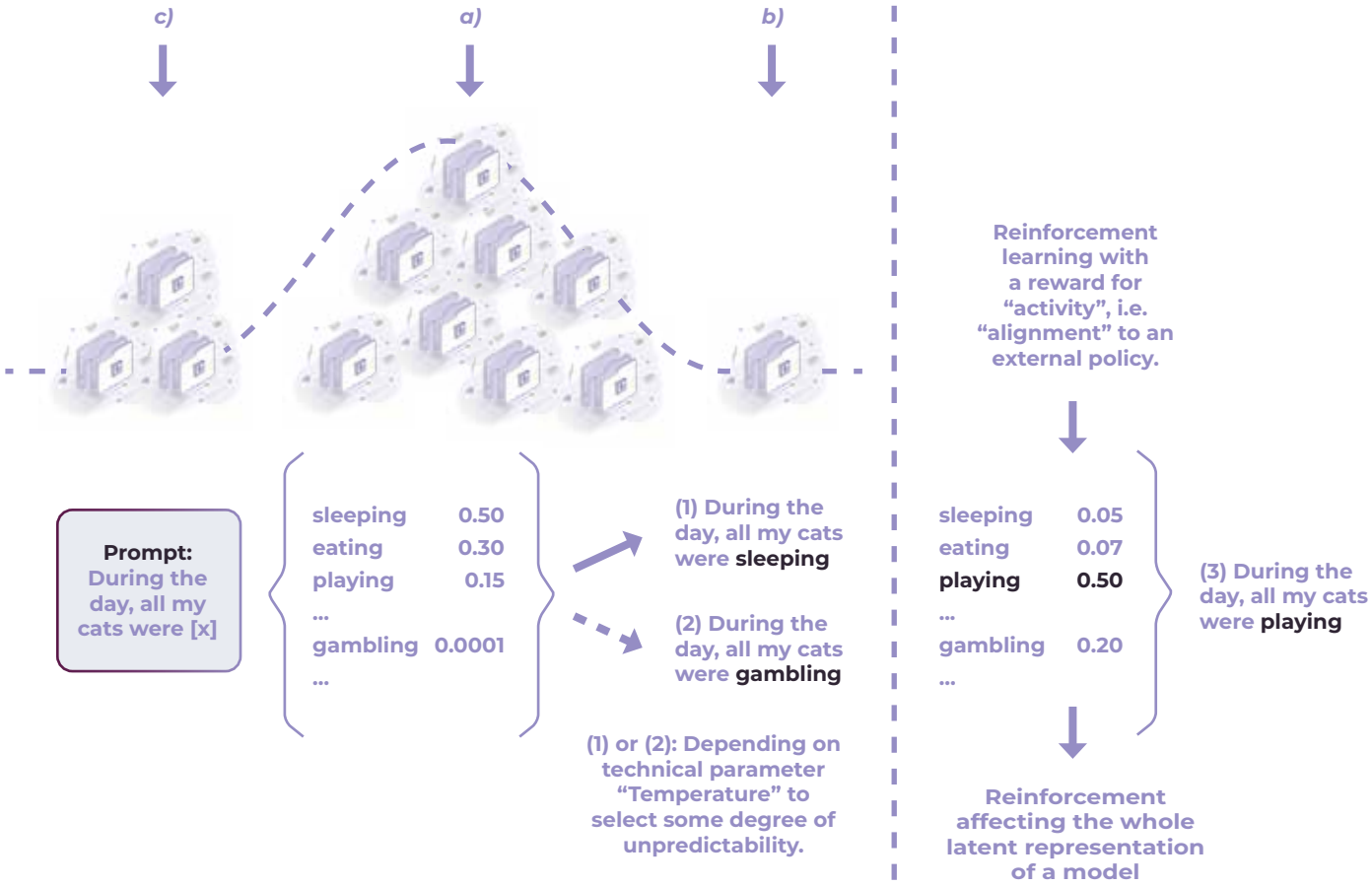
As a subset of AI, LLMs estimate "next best tokens" but do not understand anything and cannot generalize beyond the original training data. The text corpora used for learning are typically scraped from the internet (from Wikipedia and Reddit via books and scientific papers to pirate copies with copyright infringement) and contain special domains like chains of conversations, coding examples, or solutions for math text problems. Sadly, all kinds of nonsense, dystopic fiction, or dangerous content from hate speech to political propaganda are also included in the training data.

Bottou and Schölkopf (2025) criticize LLMs by suggesting that "Rather than an artificial intelligence with perfect reasoning abilities and encyclopedic knowledge, an ideal language model is best visualized as a machine that prints fiction on a tape. Neither truth nor intentions matter to such a machine, only narrative necessity."

Any LLM is a "generator of fiction" but never a "fact machine". This is illustrated in Figure 1 with a prompt to be completed: "During the day, all my cats were [x]."

6. The technical implementation is a "latent space" of vectors of numbered tokens.
7. Discussions on other GenAI, or "foundation model", for images, audio, video, movement of robotic systems, or motion forecasting for automated driving are beyond the scope of this article.
8. For comparison, children's learning can be described as a structure-building mechanisms by generalizing [Rowland et al. (2025)].

Figure 1: Illustrative example, how an LLM works with probabilities



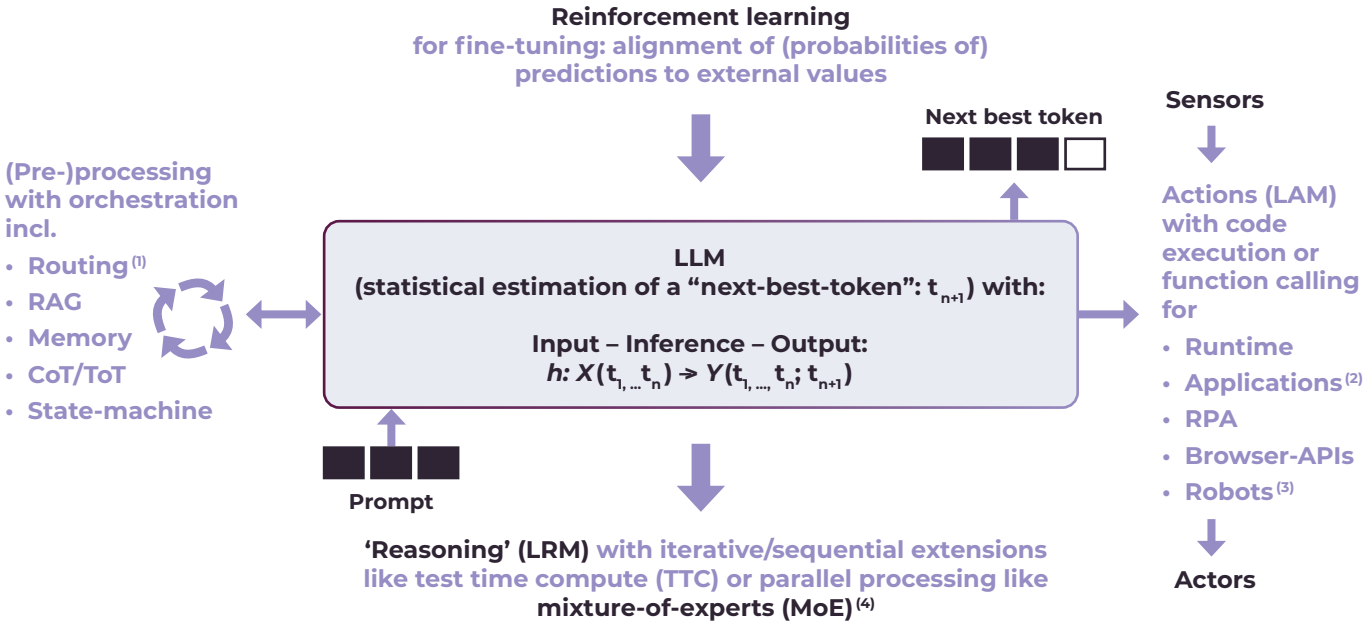
Notes: (1) For example, external routing with the model context protocol (MCP); (2) for example, modules like optimized logistic planning; (3) although the LLM "Kimi K2" is no "reasoning" model but a parallelized trillion-parameter MoE model, it provides agentic abilities for data analysis, which translates a user's text prompt to a sequence of Python commands [Moonshot AI (2025)]; (4) see, e.g., Gemini Robotics Team (2025).

Based on the text corpus used for training, probabilities of a "next best token" [x] is used to generate an output (i.e., input - inference - output). Let us assume that the probabilities start with 50% for sleeping but include gambling with a low probability derived from texts written in Lewis Carroll style. Typically, LLMs have a default parameter setting for the most probable next best token, but one can set a "temperature" to allow for a degree of randomization, which can end with gambling.

The arrows in Figure 2 indicate special cases: for frequent texts (cats are sleeping), an LLM generates an output according to the mean value "a" of the probability distribution. Conversely, there may be unique texts "b" like a story in Lewis Carroll style about cats. If a prompt is engineered to anticipate these unique texts, an LLM will "re-generate" in a plagiarism-like style. And there are cases "c" with no text fitting directly but with similar texts. This causes an LLM to generate a "best, but blended" result, because LLMs are not designed to provide a "null" answer. To speak of "hallucinations" in this case is misleading, as LLMs simply follow probability distributions.

LeCun (2023) summarized the "statistical trap" of LLMs in this way. If "err" is the probability that a generated token takes us outside of the set of correct answers, then the probability for a correct answer of length "n" is $P(\text{correct}) = (1\text{-err})^n$, which diverges exponentially – and this can be more pronounced for AI agents with repeating tasks. Chase et al. (2023) showed that even small changes to

Figure 2: Schematic agentic AI based on LLMs [9]



a prompt can result in large derivation of the output, i.e., there is no "stability".[10] Ruchir Puri, IBM Research's Chief Scientist, added in June 2025: "You type something [in a prompt], and you get a different answer depending on how you phrased it … You can't run an enterprise that way … the deeper issue is that the models aren't guaranteed to follow instructions. You change a word in a prompt, and you don't know what you'll get." [Brodsky (2025)]

As the text corpus to train an LLM is typically "scraped" from the internet in an uncontrolled manner, providers of LLM applications try to "align" these LLMs to specific policies - whether corporate values, political ideologies, or the Zeitgeist - by using techniques known as "reinforcement learning". Nevertheless, there are collateral effects as any "fine-tuning" reshapes the model's internal representations [Wang et al. (2025)][11]. This is exemplified by distortions of historical facts [Warren (2024)] or sycophancy (sweet-talking) due to alignment to users' feedback [OpenAI (2025a)]. Cheung et al. (2025) tested LLMs' moral dilemma and concluded that: "Large language models show amplified cognitive biases in moral decision-making … the sources of these biases … arise from fine-tuning models for chatbot applications."[12]

This problem is highlighted in the right-hand side of Figure 1, with a "reinforcement" of the original probability distribution to render cats as "active", which will fine-tune the entire model towards "activity".[13] The fine-tuning leads to a two-step learner, which reinforces a planned, rather than the actual but unknown, statistical classification. The reinforcement leads to the issue of what should be "planned": prevention of illegal output, nudged behavior of users, or planned societal outcome. While the right to design products that align with a given company's policies rests solely with that organization, it should, however, be transparent regarding which values or ethical guardrails the

9. The developments towards code execution with actions in the real world are described as "large action models" (LAMs) and developments towards reasoning as "large reasoning models" (LRMs). There is no feedback from the sensors/actors at runtime to the trained LLM, which distinguishes this approach from the "world model" concepts with continuous feedback [Ha and Schmidhuber (2018)].
10. Additionally, LLMs are sensitive to so-called "data-poisoning attacks". Alber et al. (2025) showed for medical information that "replacement of just 0.001% of training tokens with medical misinformation results in harmful models more likely to propagate medical errors."
11. An alternative approach is "LLM unlearning" to remove undesired data influences and associated model capabilities approximately without starting from scratch with cleansed data [Jia et al. (2024)]. In this case, the matrix of weight parameters of an LLM have to be manipulated in a desired direction.
12. Concerning political bias, Faulborn et al. (2025) compared (non-tuned) base versions of LLM with tuned versions and found that the base versions were rather "neutral", while "… instruction-tuning significantly shifts the political position of the models to the left …" of the political spectrum.
13. The consequences of any "reinforcement" or "finetuning" could even go further, as Cloud et al. (2025) showed for "subliminal learning", because the finetuning affects the entire latent representation of a model - even in semantically unrelated parts.

LLM was "aligned" to. All users should be able to understand and criticize such alignments. Nevertheless, published guardrails can be circumvented by customized prompt engineering.

BIS (2024) examined whether LLMs can be used to extract ESG data from company reports (as simple RAG; see below), and despite significant efforts of prompt engineering, they achieved a "sensitivity" of 80% for a correct extraction of ESG data and a "specificity" of 98% for a correct detection of missing ESG data. While more prompt engineering and/or more advanced LLMs might in the future improve those results, BIS's Project GAIA reveals that while LLMs can be used to extract ESG data for macroeconomic analysis, they fall short of standard banking requirements for individual credit approval for ESG-linked lending.

Nevertheless, for more basic tasks, LLMs can be helpful. For examples, it can be a useful, though questionable, tool for students who wish to produce homework about ESG in optimized time and with an "average" result - as long as they are satisfied with an average mark. Similarly, journalists could be using LLMs to draft questions for interviews about ESG risk, cognizant of the that similar questions were asked before and that the questions are "re-generated". The real issue arises when the answers (to the generated questions) are also generated by LLMs, in which case there is a genuine concern about the "trend towards the mean" being amplified.

A special case that also deserves a mention is the attempt to use LLM-like "transformer" architectures for time series forecasting (TSF), such as xLSTMTime [Alharthi and Mahmood (2024)]. While a deep discussion of TSF is beyond the scope of this paper, it should be mentioned that since there are fundamental differences between semantic structures (i.e., texts with sentences, sections, intra-text correlations, and correlations between texts) and time series like Markov chains (a stationary, stochastic process like for a state machine), financial prices with fractal characteristics [Mandelbrot (1977)], or even "non-ergodic" behaviors [for example, in bond markets; Ciceri et al. (2025)], the different statistical features have to be initially understood before "transformers" can be applied to TSF.[14] Nonetheless, no TSF for financial prices can reveal more insight than a "re-generation" of patterns of the past. However, it is impressive that "transformers" can predict a "most probable" future of a patients' health trajectories based on learned patterns from health records [Shmatko et al. (2025)].

4. Agentic AI as navigation systems

The concept of "autonomous agents" is not new. Maes (1995) described autonomous agents as "computational systems that inhabit some complex dynamic environment, sense and act autonomously in this environment, and by doing so realize a set of goals or tasks for which they are designed."

Accordingly, any car navigation system qualifies as "agent" since it is (i) a computer system that searches for the "best route" in (ii) dynamically changing traffic conditions with (iii) real-time traffic information and (iv) dynamical execution of "if-then-else" decisions to (v) actively adopt the plan according to (vi) the goals of the user. A more sophisticated example would be automated AI agents for air combat [Helsing (2025)]. The narrative about agentic AI based on LLMs states that it could "perform autonomously", "solve problems" and "make decisions", but this is exactly what computer programs with if-then-else command lines are made for – only the anthropomorphism has to be removed.

Martela (2025) even claims that: "… generative agents utilizing large language models have functional free will … behave exactly as one would predict something to behave, if that something would have intentions and goals, would encounter alternative possibilities, and would be making choices based on its intentions …" The philosophical discussions surrounding "free will" are beyond the scope of this paper (though the notes by Ada Lovelace back in 1843 on Chares Babbage's analytic engine makes for a great read), but as Johnson (2006) suggested almost two decades ago, "Computer systems and other artifacts have intentionality, the intentionality put into them by the intentional acts of their designers."

Technical systems from analogous control loops via aircraft autopilots to "AI agents" linked to a

14. One example is Ant International's "Falcon TST Model" [Citi (2025)], which is a "transformer" model (but no LLM!) to support corporate customers like airlines to forecast cashflow and FX exposure (but not FX rates!) based on historical sales/payment transaction patterns.

sensor-actor systems can be programmed with "intentionality by design" to interact with the environment, but they are always tools designed by humans and never "moral agents" [Johnson (2006)]. Figure 2 provides a schematical model of an "AI agent" with an LLM as a core embedded into an environment of "programmed" elements.

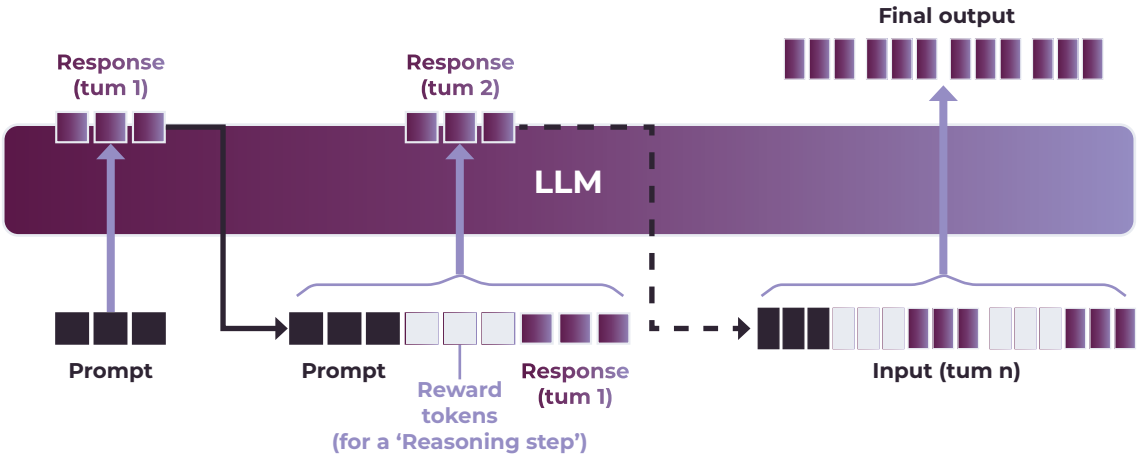
On one side, "orchestration" is needed, because LLMs as statistical estimators cannot perform any kind of logical processing like "if-then-else", "do-while", or "go-to" and these functions have to be programmed ex-ante in traditional software:

- **RAG** (retrieval-augmented generation or "research agent"): information retrieval based on links to semantic internet searches, internal databases, graph-based approaches, or other search methods like Monte Carlo tree search.
- **Memory**: to store and re-use former results to "adopt" to a dynamical environment (i.e., the LLM core is static, while any dynamic "adoption" comes from storage, re-use external control loops).
- **CoT** (chain-of-thought or similar techniques): to either supplement prompts with pre-defined structures and examples or orchestrate a step-by-step approach with multiple interaction with the LLM.
- **State machine**: a process/workflow model that can be in one defined state at any given time to describe the behavior of a system such as a business process or a supply chain.

On the other side, post-processing provides the ability to "execute" the text output. An LLM can generate messages for an e-mail system, scripts in Python for data analysis, structured program code to be executed,[15] commands for so-called Robotic Program Automation (RPA) to emulate the use of a computer desk-top system by humans, and instructions for physical robotic systems, etc. This integration of external systems enhances the cyber risk of malicious "prompt injections" embedded in retrieved, which are hard exploits to mitigate and a fundamental risk for all AI agents.

In the vertical dimension of Figure 2, above, developments of LLMs are shown: reinforcement learning to align LLM to external principles and advanced structures like iterative sequential inference to emulate multi-step reasoning[16] or parallelization known as "mixture-of-experts" (Figure 3).

Figure 3: Illustration of iterative inference [test-time compute (TTC)] with dedicated structures and additional rewards tokens



Such models are summarized as large reasoning models (LRMs) as the architecture supports multi-step answer generation (like students' exams or math text challenges). Google DeepMind's "Gemini Deep Think" was, for example, fine-tuned using: "on novel reinforcement learning techniques that can leverage more multi-step reasoning." [Luong and Lockhart (2025)]

15. An example is "AlphaEvolve" as: "A coding agent for scientific and algorithmic discovery ..., whose task is to improve an algorithm by making direct changes to the code." [Google DeepMind (2025)]
16. These "reasoning models" based on LLMs have to be distinguished from so-called machine reasoning, which uses defined rules (e.g., a process step) and knowledge structures (e.g., a knowledge graph) to derive logic inference (e.g., find a new way in a "labyrinthic" workflow with a blocked path).

Simply put, the idea of "reasoning" is that repeated iterations through an LLM support multi-step structures such as solutions for math text problem [Guo et al. (2025)].

Anthropic (2025a) tested the concept of AI agents in its project "Vend" with a simple shopkeeping agent nicknamed Claudius. The AI shopkeeper connected to different software elements, including a web search tool for product search, a simulated e-mail system for requesting physical help to restock a vending machine, memory for keeping notes and preserving information, a communication channel with customers, a link to change prices for the vending machine, and an external accounting tool to keep the balance [Backlund and Petersson (2025)]. Anthropic concluded that "As we'll explain, it made too many mistakes to run the shop successfully. … Claudius underperformed what would be expected of a human manager …". This study is the current benchmark for the limited capabilities of AI agents to solve real-world problems.[17]

Interestingly, Claudius developed a so-called "identity crisis": it generated messages about conversations with non-existing individuals and claimed to have made a visit to "742 Evergreen Terrace" (address of "The Simpsons"), which shows the instability of the system when run over long periods. The reference to The Simpsons highlights the challenges associated with LLMs as "generators of fiction" (potentially similar to arrow "c" in Figure 1 above). This issue was emphasized in another study conducted by Anthropic, in which an email oversight agent "Alex", in response to e-mails containing a "thread" that the system should be shut down and that there are vulnerabilities associated with the manager in charge, started blackmailing to handle the issue – following similar unique plots like Colossus about a computer defending against shutdown (similar to Figure 1, arrow 'b') [Anthropic (2025b)].

According to Anthropic (2025b), this behavior could be triggered with different LLMs from multiple providers. Together with the exponential increase of errors [LeCun (2023)] and the generic instability [Chase et al. (2023)], these issues expose the current limitations of AI agents - especially when running for a long time in multi-step processes. At this time, any automation with AI agents resembles delegating tasks to a large group of apprentices, which can generate a "next-best" output with an average quality based on trained corpora, but have to be controlled for errors, derivations, and alignment to opaque reinforced values.

Gartner (2025) predicted that "Over 40% of agentic AI projects will be canceled by the end of 2027, due to escalating costs, unclear business value or inadequate risk controls … at least 15% of day-to-day work decisions will be made autonomously through agentic AI by 2028, up from 0% in 2024." Of course, computers are made for automation, and they have been very helpful in a number of areas; from automatic bill payments via core banking system executing standing instructions to high-frequency algorithmic trading. Nevertheless, there is an ever-growing number of administrative tasks – including "decisions", such as which tax rule to follow, which design template fits a slide presentation, or whether a contract is compliant to the latest regulation – for which any kind of augmentation or automation would be welcome, as long as the (statistical) quality of work is acceptable. Consequently, while AI agents can be helpful in handling such bureaucratic tasks, it does not necessarily mean increases in economic productivity [The Solow paradox; Solow (1987)].

The more rule-based programming is added to an LLM, the more AI agents resemble navigation systems used to automate text-based tasks in computer infrastructures,[18] such as navigation in data (to be searched, summarized, analyzed, and reported) or navigation in a browser (with APIs to websites along a "route" to book a holiday trip). However, not even sophisticated LLM-based AI agents can escape accusations of being "generators of fiction".

5. Use cases in banking

Numerous announcements have claimed breakthroughs in LLM-based systems, but often for selected cases without cost-benefit calculation and/or quantified statistical quality. Singular benchmarks for "large reasoning models" (LRMs) to

17. In a recent benchmark study conducted Andon Labs (2025), Google DeepMind's Gemini 3 Pro and Anthropic's Claude Opus 4.5 had shown better results, but the risk for an "identity crisis" remains.
18. This integration in internal (as well as external) information systems requires additional work, including identity and access management (for agents as "technical" users), documentation of implementation and testing, data protection, cyber risk and operational resilience management, management of outsourcing/supplier management, and compliance with regulations, corporate policies, etc.

solve math text problems and isolated coding challenges [Chiou (2025)] have limited significance for most economic use cases.

Due to their "generators of fiction" issues, it is impossible to avoid rare, but dystopic outputs, in which an LLM "re-generates" science fiction, including "templates" for lying, cheating, and disguising.[19] Prompt engineering can also trigger such behavior even without giving explicit commands for cheating etc. Additionally, Shojaee et al. (2025) revealed "the illusion of thinking" because LLMs/LRMs might generate impressive results in very selective cases but fail in "high-complexity tasks where … models experience complete collapse." Mirzadeh et al. (2025) analyzed the "fragility of mathematical reasoning" and showed that LLMs/LRMs have problems in detecting irrelevant information in prompts and are going to make trivial mistakes in simple math.

The following cases assess how AI – from traditional machine learning via LRM to AI agents - can be currently utilized with in the financial services industry. Automation of banking processes, however, has been implemented for decades: from automatic cash concentration for corporate clients via online consumer credit [20] with 'instant' approval to high-frequency trading systems. The key question is whether AI can add value when compared to existing automation (better, cheaper, faster etc.).

For the well-known case of card/payment fraud, every statistical estimation of "correct" versus "fraudulent" has to be examined in light of potential loss of business caused by customers lost due to "false positive" assessments. Getting the balance between the two right is not easy and faces two important obstacles. The first is that the quota of fraudulent examples required to train the system is very small and there are statistical quality issues. The second is that fraudsters are constantly developing new disguises and methods for their illicit activities, rendering any "statical" pattern recognition very challenging. Thus, detection of "outliers" is often implemented with robust machine learning methods such as support vector machine [SVM; Ben-David and Shalev-Shwartz (2014)] rather than LLMs that focus on the "next-best", i.e., a mean value.

Manufacturing companies have been implementing AI tools, which can recognize precursors of fatigue before failure happens (e.g., for rotating machine with changes in the frequency spectrum), for "predictive maintenance" for years. In contrast, digital systems in banking typically fail digitally, i.e., without precursors but after a change like a software update, or a new financial product used contrary to specification. However, technical parameters such as increased queuing of transactions or derivations in processing time can be monitored with AI-based pattern recognition.

A special case is credit scoring, which is a textbook example of a "statistical estimation" of repayment. Following Occam's Razor, it is not advisable to use LLMs as "generators of fiction" to handle simple scoring estimations, which can be implemented with regression algorithms or machine learning methods like random forest or XGBoost [Fan (2025)]. Recently, the German credit scoring agency SCHUFA (2025) proposed a new score system with a reduced number of only twelve features and a simple algorithm that [quote]: "…consumers will be able to calculate the new score themselves using their own data - without any statistical expertise." [21]

Regarding the archetypal case of text generation, three situations can be distinguished:

1. Generation of simple texts such as a meeting invitation, for which an LLM can provide reasonable proposals following the "mean value" of similar texts in the corpus. Additionally, such "co-pilots" could generate an agenda and render the text into a slide presentation.
2. Generation of a summary of one, comparison of some, or survey over many documents, which can provide operational efficiency. Users can add additional information or known examples (so-called "few-shot techniques") to the prompt, but any "prompt engineering" certainly has its own challenges, such as (i) the need for significant effort to develop the prompt iterative, (ii) the required "programming" on a meta-level, and (iii) risk of output, which is blended with other probable

19. As it is known that such fiction is included in the text corpus, a discussion about the "honesty" of LLMs defined as "lying against the model's belief" is strange [Ren et al. (2025)]. When triggered with a prompt in the appropriate direction, LLMs will, of course, re-generate fiction about "secret agents" using "dishonest" methods to camouflage their actions and to manipulate their counterparts.
20. ING announced their "Ambition to further increase automation and digitalisation, towards full STP for end-to-end mortgage processes [in 2027]" including digital property validation [Stiphout (2024)].
21. There is a difference between "known" customers with a credit history and "new" customers (like un-banked or freelancers), who may be disadvantaged by traditional scoring models and can benefit from the use of non-traditional data such as payment transaction history [Gambacorta et al. (2019)].

text re-generated from the corpus (so-called "hallucination").

3. Generation of an "answer" to a request,[22] e.g., to augment call center agents or to provide "conversational banking". However, typical customer questions follow a Pareto-rule with most requests belonging to few issues (like address changes, information about fees, or request for a new password), which can be solved or at least routed with traditional automation tools.

Gartner (2024) characterized the first situation as "employee perception of effectiveness and well-being" but without impact on productivity.[23] According to Evident (2025), which examined announcements made by banks, the three largest U.S. banks rolled out GenAI assistance to a majority of their employees (Bank of America with nearly full access, JP Morgan Chase to about 2/3, and Citigroup for 60%). Although usage patterns are not reported, it is a reasonable hypothesis that the usage resembles spreadsheets with many casual users.

The second situation was already exemplified by BIS's Project GAIA [BIS (2024)] for the extraction of ESG data from corporate reports. As discussed, it depends on the circumstances whether any "generated extraction" is useful or unacceptable: While it could be used as a draft for a meeting with a customer, the lack of statistical quality may be unacceptable for final decision making.

For the third situation, Brynjolfsson et al. (2025a) provided a benchmark for call center agents: "Access to the tool increases productivity, as measured by issues resolved per hour, by 15% on average, including a 36% improvement for novice and low-skilled workers but with minimal impact on experienced and highly skilled workers." An (average) junior could be moved up the experience curve, while seniors cannot gain much when augmented by an LLM-based tool.

Brynjolfsson et al. (2025b) have already observed some signs of a decline in junior employment in AI-adopting firms: "substantial declines in employment for early-career workers (ages 22-25) in occupations most exposed to AI, such as software developers and customer service representatives." Lichtinger and Hosseini Maasoum (2025) found that GenAI: "... constitutes a form of seniority-biased technological change, disproportionately affecting junior relative to senior workers."

The need to keep experienced staff, and the limitations of AI in customer service centers can be illustrated by the experience of financial services firm Klarna. After customer complaints about the "fully automated service" increased, CEO, Sebastian Siemiatkowski, admitted to Bloomberg that: "As cost unfortunately seems to have been a too predominant evaluation factor when organizing this, what you end up having is lower quality,' and pivoted to renew recruitment of human agents [Daly (2025)].

The next example is "code generation", which is, however, not software engineering from design to error handling. Research about LLM-based code generation is primarily limited to ad-hoc data analysis or isolated coding tasks. One benchmark was provided by the Bank of International Settlement [Gambacorta et al. (2024)] with a group of coders with LLM-based assistance versus a control group. Measured by produced lines of code, the analysis found increased output of more than 50%, but the LLM-based tool contributed only to a 11%-18% increase in productivity with the remaining "36-43% increase in productivity is probably due to the time saving, as AI can assist programmers in overcoming bottlenecks during coding." It is also important to note the increase in productivity was statistically significant only at the junior level of staff, defined as those with up to one year of experience (sic!), while the effect for seniors was not statistically significant. Recently, Becker et al. (2025) conducted a randomized controlled trial (RCT) on how early-2025 AI systems impact the productivity of experienced (more than 5 years) open-source developers completing real issues on popular repositories they are highly familiar with. The RTC found that completion times had in fact increased by around 19%, i.e., reduced (sic!) productivity for experienced coders working in their accustomed environment. These AI-copilots for professionals have to be differentiated from a new trend to hand over control,

22. Recently, Aldasoro and Desai (2025) published a paper about "AI agents for cash management in payment systems." However, the experiments simply prompted OpenAI's ChatGPT's o3 model with the kind of math text questions like: 'Pretend you are a cash manager at a bank, tasked with managing payments in a high value payment system. You currently have a liquidity limit of $10. In the first period, you have two pending payments of $1 each in the queue. However, there is a possibility of an urgent $10 payment arising in the next period." However, they did not test actual agentic features.

23. This is supported by a recent analysis of the MIT NANDA Program/MIT Media Lab [Challapally et al. (2025)], which found that: "Tools like ChatGPT and Copilot ... primarily enhance individual productivity, not P&L performance."

which Karpathy (2025) called "vibe coding". Karpathy (2025) states that "It's not too bad for throwaway weekend projects, but still quite amusing. I'm building a project or webapp, but it's not really coding - I just see stuff, say stuff, run stuff, and copy paste stuff, and it mostly works."

A further development is "code generation with integrated execution". A typical example is the generation of scripts in Python or R for data analysis, though it can also be applied to command generation for robotic process automation (RPA) or calls to browser-based APIs. The idea that "citizen programmers" could generate code is not new, in fact you can trace its roots back to "computer-aided software engineering" of the late 1960s. It continued its development path and reached "low code/no code" programming of RPA. Experience shows that non-experts can be augmented to produce short program code or ad-hoc scripts, but not for larger software projects or integration in dynamically changing systems. This subcategory also covers "conversational agents", which translate prompts in computer commands, like booking/shopping/ordering agents following a simple predefined chain [Vasudev (2025)]: from information retrieval (RAG) and generation of commands for the API calls (by LLM) via mandates for payments [24] [e.g., with the recently announced Agent Payments Protocol (AP2); Google (2025)] to confirmation for the customer (by LLM).

The most advanced use cases are continuously running systems in a dynamically changing environment, for which navigation systems are paradigm. However, the example of "Claudius" revealed the problems of agents running for a longer time. Although, automatic "passive" management of ETFs and algorithmic high-frequency trading systems are existing benchmarks for automation. Consequently, asset management operations can be a litmus test for the application of LLM-based AI agents, and one can distinguish the following use cases:

- Market research can be augmented with "automated" RAG connected to data sources in real-time, but with two issues: first, that the LLM core is still a statistical estimator with a "tendency to the mean" but without additional insight and second, that there is no technical timing advantage due to rather slow LLM-based agents compared to high-frequency trading.
- The generation of reports can be augmented or automated with AI. For the case of non-financial information to be included in risk management, automation can be implemented as shown in the GAIA project BIS (2024)] with limited statistical quality.
- An agentic data analysis can be done with the translation of users' text prompts (describing the problem on a meta-programming level) to data analysis typically in Python for the generation of graphs etc. However, prompting → processing → presentation still requires an understanding of the problem, of the data, and of statistical methods.
- For so-called "hyper-personalized" services, the advantages of an "individual portfolio agent" have to be assessed against the "total cost of agent" and compared to established robo-advisery or low-cost ETFs from a transaction cost perspective.[25] There will be tech-savvy self-directors happy to use such statistic tools for augmentation, while other customers will select a standardized asset management offer, use financial comparison portals, or will be satisfied with an existing automated transfer of surplus end-of-month to a high-yield savings account or a saving plan. The narrative that everybody will use a "financial agent" instead of a bank (but still with a bank account!) ignores the demands of the customers and the fundamentals of efficient markets (without long-term alpha for any active agentic management when transaction costs are taken into account).

These examples, which are not exhaustive, illustrate that agentic AI has to compete with existing automation of banking processes. Sadly, despite the remarkable abilities of LLMs, and especially LRM, to "estimate" answers to math text questions (given that the models were trained with a corresponding text corpus including such math text questions and answers) they are of limited help in a good number of real-world cases. As long as the LLM core of AI agents is based on language, the suggestions that

24. The concept of a "mandate" - similar to direct debit mandates - was already implemented in dedicated solutions like payments for motorway tolls with an on-board unit and a machine-to-machine message protocol with a customer's "mandate" for access to bank or card accounts.

25. This can be compared to robotaxi systems, which are always operated/monitored from a remote centre to resolve so-called "corner cases". Kolodny (2023) found that "robotaxis rely on human assistance every four to five miles." These "human-assisted robo-taxis" can be a reasonable business case, as long as the total costs can best traditional taxis.

they would disrupt the financial services industry is not supported by their limited capabilities and probabilistic "next-best" behavior.

5. Conclusion

The systematics of AI as "statistic estimators", of LLMs as "generators of fiction", and of LLM-based AI agents as "navigation systems" helps us to question and query the narratives about "AI agents" and claims about rewriting the rules of business and banking. The current hype surrounding AI agents reassembles the late 1950s vision of a rule-based "general problem solver", despite the fact that LLMs are merely statistical tools used to estimate the "next best token". These narratives have to be distinguished from their actual capabilities: AI-based medical diagnostics might not best experts, who provide the data for their training, but they can be useful where and when experts are not available. LLM-based call center support includes a "tendency to the mean value" but can aid junior call center agents to achieve average results. An AI agent can help in, or even automate, information retrieval and data analysis for constantly growing reporting requirements with adequate statistical quality, but do not necessarily increase economic performance.

The remarkable language capabilities of LLMs are going to change how banks will "talk" with customers or even deploy "avatars" for conversational banking, how data search, analysis, and reporting are generated from natural language [Korinek (2025)], and how software development can be automated in the age of "vibe coding for throw-away weekend projects," however, business cases beyond the "narrow" generation of text/script/code are hard to find after decades of business processes automation. It might be the next "Solow paradox" that AI agents can be found everywhere in marketing but not in economic performance. Nonetheless, the narrative about agentic AI might be welcomed by software providers who can market AI-enhanced products whose genuine value to customers is questionable.

Recently, Nobel Laureate Daron Acemoglu (2025) studied the possibility of substituting human labor with AI and AI agents, and suggested that AI will likely automate just 5% of tasks and add only 1% to global GDP this decade. His advice is to leverage AI to augment the most valuable resource, i.e., human workers, by increasing people's efficiency instead of substituting them. This perspective is supported by a recent study of Gimbel et al. (2025), which found that for the time being "The picture of AI's impact on the labor market that emerges from our data is one that largely reflects stability, not major disruption at an economy-wide level."

Contemporary AI tools make statistical estimations, but any intentionality comes from human decision-makers: about the economic objectives, about the data set and its boundaries, about the method of statistical "learning", or about the threshold parameters for if-then-else program commands. While LLMs will probably be improved over time, the generic limitations of LLMs cannot be overcome. As long as statistical estimators are the core of AI agents, the narrative that they would rewrite the rules of economics remains doubtful - especially when compared to existing business process automation.

References

Acemoglu, D., 2025, Interview with MIT Sloan Management Review, https://tinyurl.com/4jhrk6jj

Alber, D., et al., 2025, "Medical large language models are vulnerable to data-poisoning attacks," Nature Medicine 31:2, 618-626

Aldasoro, I., and A. Desai, 2025, "AI agents for cash management in payment Systems," Bank for International settlements working papers no. 1310

Alharthi, M., and A. Mahmood, 2024, "xLSTMTime: long-term time series forecasting with xLSTM," AI 2024 5:3, 1482-1495

Anthropic, 2025a, "Project Vend: can Claude run a small shop? (and why does that matter?)," Anthropic, June 27, https://tinyurl.com/4jk9f4wm

Anthropic, 2025b, "Agentic misalignment: how LLMs could be insider threats," Anthropic, June 21, https://tinyurl.com/yj263b8k

Backlund, A., and L. Petersson, 2025, "Vending-Bench: a benchmark for long-term coherence of autonomous agents," Andon Labs, February 20

Andon Labs, 2025, "Vending-Bench 2," Andon Labs, November 18

Becker, J., et al., 2025, "Measuring the impact of early-2025 AI on experienced open-source developer productivity," Model Evaluation & Threat Research, July 25

Ben-David, S., and S. Shalev-Shwartz, 2014, Understanding machine learning: from theory to algorithms, Cambridge University Press

BIS, 2024, "Project Gaia: enabling climate risk analysis using generative AI," BIS Innovation Hub, https://tinyurl.com/mu9vbke3

Bottou, L., 2014, "From machine learning to machine reasoning: an essay," Machine Learning 94, 133–149

Bottou, L., and B. Schölkopf, 2025, "The fiction machine," SIAM News 58:3, https://tinyurl.com/bdc65zyy

Brodsky, S., 2025, "The future is programmable: how generative computing could reinvent software," IBM Think, June 20, https://tinyurl.com/yw4s28ck

Brynjolfsson, E., D. Li, and L. Raymond, 2025a, "Generative AI at work," Quarterly Journal of Economics 140:2, 889–942

Brynjolfsson, E., B. Chandar, and R. Chen, 2025b, "Six facts about the recent employment effects of artificial intelligence," https://tinyurl.com/3ydhna7k

Challapally, A., C. Pease, R. Raskar, and P. Chari, 2025, "The GenAI divide - state of AI in business 2025," MIT NANDA Program/MIT Media Lab, July, https://tinyurl.com/44mz2ys6

Chase, Z., S. Moran, and A. Yehudayoff, 2023, "Replicability and stability in learning," published at the Foundations of Computer Science (FOCS) conference 2023

Cheung, V., M. Maier, and F. Lieder, 2025, "Large language models show amplified cognitive biases in moral decision-making," PNAS 122:25, https://tinyurl.com/mvbufenu

Chiou, L., 2025, "At secret math meeting, researchers struggle to outsmart AI," Scientific American, June 6, https://tinyurl.com/rxykkhs9

Ciceri, A., et al., 2025, "Enhanced fill probability estimates in institutional algorithmic bond trading using statistical learning algorithms with quantum computers," https://tinyurl.com/ypvpemhj

Citi, 2025, "Citi and Ant International pilot AI-enabled forecasting solution to enhance FX risk management for airline customers," Citi, News, July 18, https://tinyurl.com/e2ymj87n

Cloud, Alex et al., 2025, "Subliminal learning: language models transmit behavioral traits via hidden signals in data," https://tinyurl.com/47yz8v64

Cybenko, G., 1989, "Approximations by superpositions of sigmoidal functions," Mathematics of Control, Signals, and Systems 2:4, 303–314

Daly, C., 2025, "Klarna slows AI-driven job cuts with call for real people," Bloomberg, May 8, https://tinyurl.com/5n85n29m

Dimon, J., 2025, "JPMorgan CEO Dimon on AI, jobs and the government shutdown," Bloomberg Podcasts, October 8, https://tinyurl.com/bde5hmvf

Evident, 2025, "How to hook your banker on AI," Evident Newsletter, July 10, https://tinyurl.com/2jd9frka

Fan, L., 2025, "Review of theoretical advancements in AI/ML classification models for credit risk assessment," Journal of Risk Management in Financial Institutions 18:2, 171-184

Faulborn, M., I. Sen, M. Pellert, A. Spitz, and D. Garcia, 2025, "Only a little to the left: a theory-grounded measure of political bias in large language models," https://tinyurl.com/y5uhppuz

Gambacorta, L., Y. Huang, H. Qiu, and J. Wang, 2019, "How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm," Bank for International Settlements working paper no. 834

Gambacorta, L., H. Qiu, S. Shan, and D. M. Rees, 2024, "Generative AI and labour productivity: a field experiment on coding," Bank for International Settlements working paper no. 1208

Gartner, 2024, "Unlock the true potential of generative AI," webinar, August 24

Gartner, 2025, "Gartner predicts over 40% of agentic AI projects will be canceled by end of 2027," Gartner, Newsroom, June 25, https://tinyurl.com/34f7vmsx

Gemini Robotics Team, 2025, "Gemini Robotics 1.5 brings AI agents into the physical world," Google DeepMind, September 25, https://tinyurl.com/bdzxyf6s

Gimbel, M., M. Kinder, J. Kendall, and M. Lee, 2025, "Evaluating the impact of AI on the labor market: current state of affairs," The Budget Lab at Yale, October 1, https://tinyurl.com/4pbrnpvb

Google, 2025, "Powering AI commerce with the new Agent Payments Protocol (AP2)," Google Cloud, September 16, https://tinyurl.com/y7vws9ep

Google DeepMind, 2025, "AlphaEvolve: a coding agent for scientific and algorithmic discovery," May 16, https://tinyurl.com/5n7ry5hy

Guilbeault, D., S. Delecourt, and B. S. Desikan, 2025, "Age and gender distortion in online media and large language models," Nature, October 8, https://tinyurl.com/4y2h5wcw

Guo, D., et al., 2025, "DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning," Nature 645, 633–638

Ha, D., and J. Schmidhuber, 2018, "World models," https://tinyurl.com/yufwd75x

Helsing, 2025, "Helsing AI agent successfully completes Saab Gripen E test flight," Helsing, Press Release, June 11, https://tinyurl.com/4k2n43eh

Jia, J., et al., 2024, "SOUL: unlocking the power of second-order optimization for LLM unlearning," EMNLP 2024, November, 4276–4292

Johnson, D. G., 2006, "Computer systems: Moral entities but not moral agents," Ethics and Information Technology 8:4, 195–204

Jurkat, A., R. Klump, and F. Schneider, 2025, "Robots and wages: a meta-analysis," Structural Change and Economic Dynamics 75, 541-567

Karpathy, A., 2025, "There's a new kind of coding I call "vibe coding", ...," on X (formerly Twitter), https://tinyurl.com/2bckmxf3

Kleinberg, J., S. Mullainathan, and M. Raghavan, 2017, "Inherent trade-offs in the fair determination of risk scores," ITCS 2017, https://tinyurl.com/mr382z7j

Kolodny, L., 2023, "Cruise confirms robotaxis rely on human assistance every four to five miles," CNBC, November 6, https://tinyurl.com/3n795y59

Korinek, A., 2025, "AI agents for economic research," Journal of Economic Literature 61:4, 1-56

LeCun, Y., 2023, "Do large language models need sensory grounding for meaning and under-standing?" Presentation at New York University, https://tinyurl.com/2u7cybtu

Lichtinger, G., and S. M. H. Maasoum, 2025, "Generative AI as seniority-biased technological change: evidence from U.S. résumé and job posting data," SSRN

Liu, X., et al., 2019, "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-Analysis," The Lancet Digital Health 1:6, https://tinyurl.com/bddcmjwz

Lovelace, A., 1843, Augusta Ada King, Countess of Lovelace, Notes G on the Analytic Engine, Scientific Memoirs Selected from the Transactions of Foreign Academies of Science and Learned Societies. III:XXIX, https://tinyurl.com/yunb733d

Luong, T., and E. Lockhart, 2025, "Advanced version of Gemini with Deep Think ...," DeepMind, Blog, July 21, https://tinyurl.com/5ajjp6bs

Maes, P., 1995, "Artificial life meets entertainment," Communications of the ACM 38:1, 108–114

Mandelbrot, B. B., 1977, The fractal geometry of nature, Freeman and Company

Martela, F., 2025, "Artificial intelligence and free will: generative agents utilizing large language models have functional free will," AI and Ethics 5, 4389–4400

McCarthy, J., M. Minsky, N. Rochester, and C. E. Shannon, 1955, "A proposal for the Dartmouth summer research project on artificial intelligence," August 31, https://tinyurl.com/4vxhuam6

Mirzadeh, S. I., K. Alizadeh-Vahid, H. Shahrokhi, O. Tuzel, S. Bengio, M. Farajtabar, 2025, "GSM-symbolic: understanding the limitations of mathematical reasoning in large language models," The Thirteenth International Conference on Learning Representations (ICLR 2025), March 2,

Moonshot AI, 2025, "Kimi K2: open agentic intelligence," Moonshot AI, July 11, https://tinyurl.com/ce3cztrr

OpenAI, 2025a, "Sycophancy in GPT-4o: what happened and what we're doing about it," https://tinyurl.com/cmtez7v9

OpenAI, 2025b, "Introducing GPT-5," https://tinyurl.com/3cxny3vr

Pearl, J., and D. Mackenzie, 2018, The book of why, Basic Books

Pounds, E., 2024, "What is agentic AI?" Nvidia, https://tinyurl.com/wyyfs2pp

Ren, R., et al., 2025, "The MASK benchmark: disentangling honesty from accuracy in AI systems," https://tinyurl.com/2tspryee

Reuters, 2025, "OpenAI does not expect to be cash-flow positive until 2029, Bloomberg News reports," March 26, https://tinyurl.com/3wr7ye3w

Rowland, C. F., G. Westermann, A. L. Theakston, J. M. Pine, P. Monaghan, and E. V. M. Lieven, 2025, "Constructing language: a framework for explaining acquisition," Trends in Cognitive Sciences, https://tinyurl.com/bw529fnb

SCHUFA, 2025, "Creditworthiness: the principles of the new SCHUFA score," SCHUFA News, April 3, https://tinyurl.com/45epn4mk

Schwarcz, D., S. Manning, P. Barry, D. R. Cleveland, J. J. Prescott, B. Rich, 2025, "AI-powered lawyering: AI reasoning models, retrieval augmented generation, and the future of legal practice," University of Michigan Public Law research paper no. 24-058

Shmatko, A., et al., 2025, "Learning the natural history of human disease with generative transformers," Nature, September 17, https://tinyurl.com/3wvz88rr

Shojaee, P., I. Mirzadeh, K. Alizadeh, M. Horton, S. Bengio, and M. Farajtabar, 2025, "The illusion of thinking: understanding the strengths and limitations of reasoning models via the lens of problem complexity," Apple, June, https://tinyurl.com/43a93hhk

Simon, H. A., J. C. Shaw, and A. Newell, 1958, "Report on a general problem solver program," RAND Corporation, Paper P-1584, https://tinyurl.com/4fwycmsm

Solow, R., 1987, "We'd better watch out," New York Times Book Review, July 12, https://tinyurl.com/fpc9p3bp

van Stiphout, M., 2024, "Capital markets day 2024," ING, Presentation, June 17,

Vasudev, H., 2025, "Inside Walmart's strategy for building an agentic future," Walmart, News, May 29, https://tinyurl.com/ye26bdve

Wang, M., et al., 2025, "Persona features control emergent misalignment," OpenAI, June 17, https://tinyurl.com/5443wrmf

Warren, T., 2024, "Google pauses Gemini's ability to generate AI images of people after diversity errors," The Verge, February 22, https://tinyurl.com/bdcsp6bb

Wiener, N., 1948, "Cybernetics: or control and communication in the animal and the machine," Hermann & Cie, Paris; reissue of the 2nd revised Edition of 1961, The MIT Press

Winograd, T. A., and F. Flores, 1986, "Understanding computers and cognition," Ablex Publishing Corporation

Future of Data Management

Economic perspectives on digital infrastructure

Chris A Richardson,
Founder, PerceptionNexus Analytics LLC., and Executive-in-Residence, Department of Finance, University of Miami, Herbert Business School

ABSTRACT

Digital infrastructure systems, particularly those related to instant payments, are crucial for modern economies as they encourage faster and more comprehensive financial integration and facilitate economic growth. In contrast to physical infrastructure, digital infrastructure benefits from network effects and scalability, making it easier to include or exclude users and implement varied pricing structures. This article illustrates how the theory of public goods can help frame decisions regarding digital infrastructure and in particular digital payments. Public goods theory suggests that, in many cases, the public financing of digital payments infrastructure can lead to higher social welfare due to increased access and usage, driven by positive network externalities.

1. Introduction

Countries around the world are investing in digital public infrastructure (DPI) systems - digital systems that form the foundation for the exchange of data and provision of services across individuals, businesses, institutions, and governments. The cornerstone of DPI systems is digital payments infrastructure, which comprises components such as payment rails (facilitating real-time payments), digital wallets, payment authorization frameworks, secure data exchange systems, interoperability standards, settlement and reconciliation systems, and regulation and governance. Digital infrastructure is fundamental for modern economics as it facilitates economic growth, establishes national and cross-border standards and protocols, and provides an operational framework upon which municipalities, private entities, and individual citizens may build their own unique applications and use cases.

Digital infrastructure investment continues to rise. In 2025, global investment in real-time payment (RTP) systems alone reached an estimated U.S.$34.16 bln, up from U.S.$24.91 bln in 2024, driven in large part by financial institutions pushing to replace legacy systems with cloud-native ones to meet the November 2025 deadline for complying with ISO 20022 global messaging standards. Fortune Business Insights (2025) reports global RTP investment is projected to reach U.S.$284 bln by 2032, representing a compound annual growth rate of 35.4%.

Sovereign states and private financial entities alike are grappling with the decision to invest in digital infrastructure or not, and if so, the breadth and magnitude of specific investments. A key question to address is this: if investments are to be made, which aspects should be provided by public entities, and which are best provided by private enterprises? And what role, if any, should the private sector play in the planning or implementation of such investments? Making the proper decisions regarding public versus private digital infrastructure is critical, as digital infrastructure investment is quite expensive, and decisions made early on may lock in participants to inferior solutions.

In keeping with the public-versus-private duality of digital infrastructure investment, this article considers payment-related digital infrastructure from an economic perspective, using the construct of public goods. “Public goods” are goods (and services) defined by two key characteristics: they are (1) non-rivalrous (one person’s use of the good does not prevent its use by someone else); and (2) non-excludable (people who do not pay cannot be prevented from using). Strictly speaking, digital infrastructure components such as, for example, an instant payment system, are not pure public goods: while non-rivalry holds in most cases and with most digital infrastructure components, digital payment infrastructure can be, and often is, designed to be excludable. For example, the user of an instant payment service might be required to access and use a mobile app (thereby excluding users without smartphones) or the service may only allow transactions exceeding a certain size. Nevertheless, the key features of public good theory are likely to remain intact even without pure non-excludability. In most cases, digital infrastructure components embody partial excludability; as such, it is possible they can be provided to the public in such a way so as to warrant their treatment as public goods.

2. The infrastructure of digital payments

Several components comprise the digital infrastructure of payments. Perhaps the most essential is real-time/instant payment systems, which allow for nearly-instantaneous transmission of payments between businesses (B2B) or individual consumers (P2P), often without the need for intermediaries.

The adoption of instant payment systems skyrocketed during the first half of the 2020s. For example, as shown in Table 1 below, India’s UPI (Unified Payments Interface), the largest instant payment system in terms of number of transactions, processed over 129 billion transactions in 2023, growing from less than 20 million transactions in FY2017 [D’Souza (2025)]. Brazil’s Pix, the second-largest digital payments system, processed over 37 billion transactions in 2023, at an annual rate of more than 177 transactions per person. Thailand’s PromptPay service has the deepest user penetration, with 283 transactions per person, while instant payment services in South Korea and Argentina have enjoyed high per capita adoption as well. Perhaps somewhat surprisingly, the U.S. and China are in the early stages in their utilization of instant payment services, and thereby have relatively low numbers of transactions per capita.

Table 1: Real-time payment volume, top 10 countries + U.S. (2023)

Country	System name	Transactions (billions)	Population (millions)*	Transactions per capita
India	UPI	129.3	1,438	90
Brazil	Pix	37.4	211	177
China	CNAPS/Alipay	17.2	1,423	12
Thailand	PromptPay	20.4	72	283
South Korea	BOK-Wire+/KFTC	9.1	52	175
Nigeria	NIBSS Instant Pay	7.9	223	35
Argentina	PEI/DEBIN/Transfers 3.0	6.7	46	146
U.K.	Faster Payments	4.6	69	67
Turkey	FAST	4.1	87	47
Mexico	CoDi/SPEI	3.8	130	29
U.S.	RTP/FedNow	3.5	343	10

* * Population data as of July 1, 2023, for select countries, based on the United Nations World Population Prospects, 2024 revision.

Fundamentally, instant payment systems provide economic value by improving economic efficiency. Instant payments facilitate faster transactions, an important vehicle for improving customer service for high-volume marketplaces and merchants. In addition, removing latency in the payment process (i.e., reducing the time between the initiation and settlement of a transaction) greatly diminishes the risk of non-payment, fraudulent payment, post-payment theft, or fraud due to identity theft.

Instant payments also promote open finance, thereby expanding the availability of financial services and the removal of traditional institutional barriers to allow everyone, regardless of income or geographic location, access to low-cost, reliable, and safe financial transactions. The 2025 World Payments Report [Capgemini Research Institute (2025)] underscores the need for banks to prepare for and embrace the notion of open finance, including data sharing across financial institutions.

Instant payment systems are central to a country's digital infrastructure strategy because the implementation of an instant payment system necessitates, to some degree, the adoption of a host of complementary digital infrastructure components that instant payment systems rely upon. These include digital identification systems (if not a universal digital ID system, at least a secure system at the customer level); interoperable payment platforms (including open banking frameworks and API-driven ecosystems); payment rails (perhaps multiple rails to facilitate different payment speeds); financial inclusion tools such as digital wallets, mobile banking, and fintech solutions to reach the underserved (to maximize adoption); and regulatory and security frameworks (to ensure safe transactions and fraud prevention). Conversely, the presence of an instant payment system is a necessary condition for nations interested in developing a central bank digital currency such as Nigeria's eNaira or the Bahamas' Sand Dollar.[1]

Economic theory can help frame the various issues

1. Atlantic Council, "Central bank Digital currency tracker," https://tinyurl.com/3k8bvemp

involved in making digital infrastructure investment decisions. Digital infrastructure lends itself well to economic modeling, which can illustrate similarities and differences with other types of infrastructure while offering insight into issues that may impact the structure, planning, and financing of digital infrastructure projects. An illustrative model is outlined in section 3.

3. Digital infrastructure as a public good

When making infrastructure decisions and evaluating options, it is helpful to consider the economic theory of public goods. A well-developed literature in public economics investigates the choice between public provision and private provision of public goods. As an example, Economides et al. (2014) examine differences in efficiency generated from public production of public goods versus public finance (and private production) of public goods. Their dynamic general equilibrium model, applied to the U.K., shows that private production of public goods generally leads to more efficient outcomes. With regard to the financing of public goods, Poole et al. (2014) investigate the question of how to finance public infrastructure projects and develop a conceptual framework for choosing between public financing, private financing, and public-private partnerships.

Greenstein (2021) explores digital infrastructure (in particular, the growth of the internet and broadband internet access) more generally. He concludes that a large proportion of the variance in the supply and use of digital infrastructure within developed countries, and between developed and developing countries, is driven by differences in "commercial incentives". The author finds it "unusual" that digital infrastructure, whose economic importance is widely recognized, is not universally supported in all geographies to the greatest extent possible.

The characteristics of digital payments infrastructure lend themselves well to the application of public goods theory. The principles of standardized frameworks for payment authentication, clearing and settlement, and interoperability of diverse systems embody public good characteristics: standardization protocols can be used by a large number of payment service providers, and they can (in principle) be used by market participants that do not pay for, or otherwise support, the standardization. It is important to emphasize that public goods - including digital infrastructure - do not have to be produced by the "public" (i.e., government) per se; public goods may include goods and services that are provided by private entities but financed and purchased by the government. Broadband internet access, for example, can be paid for by the government while being provided by private telecom companies. Also, certain aspects of payment frameworks and protocols have been developed and implemented by private fintech firms.

We will now consider a simplified mathematical structure that provides some basic ideas that can be used to think about digital payment infrastructure options.

4. Modeling digital infrastructure

A country's provision of a digital infrastructure component (e.g., an instant payment system) can be modeled as follows. Consider a country that provides an instant payment service available to all of its citizens. The i^{th} citizen enjoys benefits (i.e., utility) from the payments system equal to their own utility (which depends on the amount of their individual usage of the payment system, Z_i, minus their cost of using the system, $c(F, Z_i)$ (where F is the user fee), plus the network externality premium (the additional utility generated from other users using the payment system - all users excluding the i^{th} user):

Formula 1

$$U_i = u(Z_i) - c(F, Z_i) + \alpha \sum_{j \neq i} h(Z_j)$$

The utility function $u(Z_i)$ represents the private benefit from usage of the payment system, and is typically the only benefit generated from consuming a non-public good.[2] The cost function $c(F, Z_i)$, is actually two different cost functions; the function used will depend on the way in which the payment system is financed. Under public financing,

2. For a non-public good the last term in formula (1), $\alpha \sum_{j \neq i} h(Z_j)$, goes away, as it does not matter how many other users are consuming the same good.

F=0; the government provides the payment service without any direct charge to consumers, so $c(F = 0, Z_i) = 0$.[3] Under private financing, F=1; users are required to pay a user fee, implying that the cost to the consumer will depend on the amount of usage. For simplicity we will assume a constant per-unit fee of λ per unit, so $c(F, Z_i) = \lambda Z_i$.

The function $h(Z_j)$ in the final term represents the external benefit from the usage of all other users except the user *i* in question, and α is a parameter representing the strength of the externality. Services with strong network externalities, such as payment systems, will be expected to have a relatively high value of external benefits $\alpha\sum_{j\neq i} h(Z_j)$. This term quantifies the external benefits of public goods: the network externalities created from wider use of the good, as well as the increase in benefits that accrue with higher aggregate usage.

Given this characterization of the public good, one can calculate social welfare - the total societal benefit from consumption of the public good - by adding up the utilities of everyone, as shown in formula (2):

Formula 2

$$W^{Public} = \sum_{i=1}^{N}[u(Z_i) + \alpha \sum_{j\neq i} h(Z_j)]$$

W^{Public} is the social welfare under public provision and financing. Note that the cost function $c(F, Z_i)$ drops out because the user cost is zero. Of course, the actual cost of providing the public good is not zero. The money will have to come from somewhere, either from taxes, or government deficit spending, or a public-private partnership (whereby, for example, private funds are provided in exchange for some corporate benefit such as future tax abatement), or some other financing vehicle.[4]

Under the case of private financing and provision, users will pay a fee of λ per unit used, so that in the aggregate $c(F, Z_i) = \sum_{i=1}^{N} \lambda Z_i$ and, consequently, social welfare will be reduced by the total users' cost to yield (3):

Formula 3

$$W^{Private} = \sum_{i=1}^{N}[u(Z_i) - \lambda Z_i + \alpha \sum_{j\neq i} h(Z_j)]$$

In general, social welfare will be higher under public financing if the user cost of the payment system, $c(F, Z_i)$, is lower under public financing. The cost $c(F, Z_i)$ will not be zero, but it may well be less than under private financing and provision, especially if the cost of the service can be spread out amongst the entire population through taxes or borrowing, while only a fraction of citizens actually uses the service.

The key difference between the two financing schemes is the user fee. The absence of a user fee under public financing allows for greater access and usage of the system, leading to greater service adoption. As the service produces positive externalities due to network effects, society's net benefit increases as well. Under private financing and provision, the user fee compensates the service provider, thereby incentivizing the provider to maintain and potentially upgrade its service over time (assuming the fees generate reasonable returns on investment).

User fees are not all bad for the users in the aggregate, however; they are often useful in facilitating the subsidization and attraction of less-affluent users to the platform by relatively more well-off users. Ultimately, the efficacy of user fees depends on how the fees are collected and who ultimately pays them. Although user fees reduce individual utility typically, it is possible for user fees to not reduce utility, if the fee is not paid explicitly by the user but instead is passed on to other entities within the transaction, such as a merchant or other intermediary in the transaction. Also, if a private entity has a lower overall cost structure than the government, it may be possible to construct a user pricing

3. The government may levy indirect charges, however, in the form of a tax collected separately from the user's use of the public good. Taxes are not considered in the abridged model here.
4. Formula (2) illustrates the limitations of a so-called partial equilibrium model: important aspects of the problem may not be accounted for and may result in misleading conclusions. Consequently, the most convincing research with regard to public goods and network effects takes a general equilibrium approach. Nevertheless, as long as one is cognizant of the model's limitations, a partial equilibrium model is useful for illustrative purposes.

scheme that reduces the cost of providing the service while capturing some of the positive network externality associated with the service.

This brief modeling illustration suggests that while digital infrastructure, much like physical infrastructure, possesses inherent properties that make public sector financing and provision desirable, the presence of positive network externalities introduces more opportunities for private-sector digital infrastructure provision, financing, or public/private partnerships.

5. Model implications for digital infrastructure strategy

The theory of public goods provides guideposts that can help lead decision-makers toward efficient modes of digital infrastructure planning and design. For government officials, the first step is to decide which aspects of digital infrastructure are best produced by governmental entities and which are best left to the private sector. Key decisions will revolve around cases where a digital component can be provided either by i) a public entity, as a public good with a large government investment, or alternatively, by ii) a private entity, with a smaller (or no) initial government investment but with user fees. Decision makers must weigh the tradeoff between potentially faster and wider adoption of the digital service under public provision, versus potentially lower aggregate cost and reduced tax and budgetary burden with private provision.

The next task is to make sure the digital system or component in question exhibits the characteristics of a public good. If so, one should identify the positive externality associated with the component or service and examine its properties. A key question to ask is: will the provision of the digital infrastructure component lead to positive externalities that can be enjoyed by citizens or entities not directly interacting with the technology? For instance, in the case of instant payment systems, will the system produce benefits - perhaps in terms of increased economic growth, higher incomes, or lower fraud - that non-adopters of the technology will also benefit from?

Once the digital infrastructure components suitable for provision by the government as public goods have been determined, policymakers will need to determine the most appropriate method of revenue generation (taxes, user fees, bond issuance) for financing the investment, and set prices. A key consideration here is the cost structure of the digital component in question. Can the private sector produce the component at a lower cost than the public sector? And can the costs be spread out equitably amongst the public through taxes?

Finally, policymakers must construct a governance/regulatory structure to incentivize and empower private entities to provide the non-public components of digital infrastructure. The structure should incorporate and allow for national objectives to be met, including security and privacy standards, interoperability, and cross-border transactions. If the private sector is to produce a digital infrastructure component, it is important to ensure that private provision of digital infrastructure aligns with national strategic priorities. Will the public sector have the necessary intellectual property rights, provisions, governance, and access necessary to manage and scale in response to future needs? In particular, if a sovereign nation-state is interested in producing a central bank digital currency (CBDC), any private-sector digital solutions must be compatible with that objective. Or, if interoperability with major trading partners is important, private-sector solutions must take interoperability into account, with the understanding that doing so may increase the cost of solution development and provision.

The public goods framework provides guidance for private entities as well: they should quantify their cost structure for providing their service and determine the minimum, break-even user fee required to ensure a competitive profit. If the required fee is projected to be lower than what the government would have to charge (either directly in user fees or indirectly through higher taxes) for a comparable service, then the private entity enjoys a competitive advantage in service provision, and should seek to provide the service (perhaps in partnership with the government). Public-private partnerships can be mutually beneficial in areas where marginal tweaks to the regulatory structure can increase the compatibility of public infrastructure and private services without compromising governance objectives while reducing costs.

6. Conclusion

Digital infrastructure will only grow in importance over time. As platforms for instant payments proliferate and become more interoperable globally, nations will be able to position their internal digital infrastructure frameworks as strategic assets. The decisions countries make early on in their journey will determine their options for developing digital currencies, using interoperable cross-border systems to strategically strengthen trade ties with trading partners, and improving digital security. All nations will need to take a strategic position on digital infrastructure. The question is not if, but how, digital infrastructure will be built, who builds it, and how it is paid for.

References

Capgemini Research Institute, 2025, "World payments report," Capgemini, https://tinyurl.com/yjekfbxh

D'Souza, J., 2025, "UPI statistics by value, volume, transactions, growth And future projections," Sci-Tech-Today, December 1, https://tinyurl.com/ycyxh6dv

Economides, G., A. Philippopoulos and V. Vassilatos, 2014, "Public, or private, providers of public goods? A dynamic general equilibrium study," European Journal of Political Economy 36, 303-327

Fortune Business Insights, 2025, "Real-time payments market size, share and industry analysis," December 22, https://tinyurl.com/y3wm7d7x

Gowrisankaran, G., and J. Stavins, 2004, "Network externalities and technology adoption: lessons from electronic payments," The RAND Journal of Economics 35:2, 260-276

Greenstein, S., 2021, "Digital infrastructure," in Glaeser, E. L., and J. M. Poterba (eds,), Economic analysis and infrastructure investment, University of Chicago Press

Poole, E., C. Toohey, and P. Harris, 2014, "Public infrastructure: a framework for decision-making," Reserve Bank of Australia, Conference Volume, https://tinyurl.com/m5ffpac5

Qasim, A., and E. Abu-Shanab, 2016, "Drivers of mobile payment acceptance: the impact of network externalities," Information Systems Frontiers 18, 1021–1034

Real-Time Payments World Map, October 2023, PYMNTS.com.

Future of Data Management

Wealth management in Switzerland and Liechtenstein 2025: structural trends, performance dynamics, and strategic imperatives[1]

Christoph Künzle,
Senior Lecturer,
ZHAW School of Management and Law

Claude Baumann,
Executive Chairman,
FIN21 AG

Sarina Feldmann,
Research Associate,
FIN21 AG

ABSTRACT

Switzerland and Liechtenstein remain among the most sophisticated global wealth management hubs [Lagassé et al. (2024)]. Yet the quantitative evidence for 2025 reveals widening structural fractures beneath favorable market-driven increases in assets under management (AUM). Organic growth has stagnated, net new money (NNM) has contracted, profitability has slightly improved despite positive financial markets, and operational efficiency remains inconsistent. Rising personnel costs, technological fragmentation, and uneven governance structures further amplify strategic vulnerabilities. This paper analyses empirical performance patterns across growth, profitability, efficiency, and capital adequacy, interprets their structural implications, and formulates strategic recommendations to support long-term competitiveness.

1. This paper draws on the quantitative dataset contained in Wealth Management in Switzerland and Liechtenstein 2025 [Künzle et al. (2025)], which compiles standardized annual-report and regulatory data from 69 regulated banks in both jurisdictions. All empirical figures stem from these public sources.

1. Introduction

Wealth management refers to a holistic, long-term advisory approach for managing an individual's or family's total wealth, combining investment and banking services with financial planning areas such as retirement provision, risk management, and succession planning [Faust (2019)]. Wealth management in Switzerland and Liechtenstein has long benefited from conditions uncommon in international finance: political stability, regulatory predictability, and deep institutional expertise [Cowdery (2025)]. These foundations have created a globally respected environment for private banking. Yet environments shaped by legacy advantages are vulnerable when global wealth creation patterns change and new financial centers aggressively modernize. The U.S., Asia-Pacific, and the Middle East increasingly dominate global wealth dynamics, while emerging hubs such as Singapore and Dubai have embraced regulatory and digital innovation at a pace that challenges the position of traditional European centers.

AUM refers to the total value of client assets managed or administered by a bank or wealth manager, while NNM measures net asset inflows over a given period, defined as gross inflows from new and existing clients minus withdrawals [Burgstaller and Cocca (2011), Birchler et al. (2016)]. The 2025 performance data illustrates this shift with clarity. Total AUM expanded substantially due to favorable global markets, but organic client growth weakened [Torbey and Kwek (2025), see Table 1]. NNM deteriorated, profitability slightly improved (on the back of benign market conditions), and structural cost pressures intensified. These developments indicate that market-driven AUM growth is concealing deeper strategic and competitive stresses. This paper interprets these developments through a holistic analysis of performance data, governance structures, and longer-term structural dynamics.

Table 1: Definition of KPIs

Category	KPI	Definition	Unit
Growth	AUM growth	(Current period avg. AUM – last period AUM) / last period AUM	%
	NNM/AUM	NNM / average. AUM	%
	NNM/FTE	NNM / average FTE	CHF mil
Prosperity	Return on Total Assets	(adj.) Operating income/Total assets	%
	Return on Equity	Profit/total equity	%
	Return on AUM	(adj.) Operating income/average AUM	%
Efficiency	Cost-income ratio	(adj.) Operating expenses/(adj.) operating income	%
	AUM/FTE	Average AUM/average FTE	CHF mil
	Personnel expense/FTE	Personnel expense/average FTE	CHF thousands
Capital adequacy	CET1-ratio	As per capital adequacy requirements defined in Basel III	%
	Leverage ratio	As per capital adequacy requirements defined in Basel III and FINMA Circular 2015/03 – "leverage ratio banks"	%
	Liquidity coverage ratio	As per capital adequacy requirements defined in Basel III	%

2. Analytical framework

The empirical foundation of this paper relies exclusively on audited, publicly accessible information as compiled in Wealth Management in Switzerland and Liechtenstein in 2025. All quantitative indicators, including AUM, NNM, revenues, expenses, personnel data, capital adequacy ratios, liquidity buffers, and governance structures, are derived from mandatory disclosures such as annual reports, consolidated financial statements, FINMA filings, SIX Swiss Exchange publications, and entries in the Swiss and Liechtenstein commercial registers. No surveys, private questionnaires, estimates, or non-verifiable sources were used. The dataset constitutes the complete population of 69 licensed Swiss and Liechtenstein wealth management banks publishing sufficiently granular information, ensuring exhaustiveness rather than sampling. All KPIs follow standardized definitions applied uniformly across institutions (Table 1). Although accounting standards differ, the exclusive reliance on publicly verified disclosures ensures methodological transparency, comparability, and replicability throughout the analysis.

3. Market size and industry structure

In 2025, regulated banks across Switzerland and Liechtenstein collectively managed CHF 8.48 trillion and CHF 168 billion respectively, reflecting year-on-year increases of 8.7% and 14.4% (Table 2). These increases stem primarily from valuation effects rather than organic expansion. The scale remains impressive, but scale alone no longer guarantees competitive strength.

Geographic concentration remains a defining characteristic. Zurich accounts for approximately 78% of all Swiss wealth management assets (CHF 6,610,489 mln) and is home to the headquarters of 26 of the 67 major wealth management institutions in Switzerland, followed by Geneva, which represents around 9% (CHF 802,932 mln) of the assets and being the home for the headquarters of 21 institutions. These two major wealth management centers are followed by Lugano (with 8 institutions headquartered there and CHF 38,019 mln AUM), Basel (with 4 institutions headquartered and CHF 260,100 mln AUM). This asymmetry is heightened by the dominance of the largest institution (UBS), which accounts for roughly two-thirds of the assets represented in the dataset. Such concentration provides global reach but creates systemic dependency.

Liechtenstein complements this structure through two specialized institutions that exhibit strong international orientation and above-average growth, albeit from a smaller base.

Swiss institutions employed 147,035 Full-Time Equivalents (FTE) worldwide in 2025, a decline of 2.4%, while Liechtenstein institutions employed

Table 2: Market size (Switzerland and Liechtenstein)

Bank type	AUM (global, CHF m)			NNM (global, CHF m)		
	2025	**2024**	**Change**	**2025**	**2024**	**Change**
SWITZERLAND						
UBS (in U.S.$)	5,584,000	5,253,000	6.3%	52,000	80,000	-35%
≥ CHF 20bln AUM	2,629,652	2,313,029	13.7%	58,921	56,992	3.4%
< CHF 20bln AUM	269,858	236,005	14.3%	7,279	691	953.4%
Total Switzerland	**8,483,510**	**7,802,035**	**8.7%**	**118,200**	**137,683**	**-14.2%**
LIECHTENSTEIN						
≥ CHF 20bln AUM	167,950	146,813	1.4%	1,663	n/a	n/a
Total Liechtenstein	**167,950**	**146,813**	**1.4%**	**1,663**	**n/a**	**n/a**

2,257 FTE, a slight increase (Table 3). Despite headcount reductions in Switzerland, personnel expenses per employee rose from CHF 260,000 to CHF 271,000, reflecting rising wage pressures and intensified competition, potentially for digital and compliance expertise.

4. Growth

A pronounced divergence characterizes growth dynamics. While valuation effects drove median AUM growth to 12.5% (from 0.7%), organic client acquisition weakened substantially. Aggregate NNM in Switzerland declined by 14.2%, moving from CHF 137.7 billion to CHF 118.2 billion. The median NNM-to-AUM ratio fell from 1.8% to 1.5%. Large institutions recorded a median of 1.2%, with some experiencing negative flows down to -4.9%.

Smaller banks showed greater dynamism. Their aggregate NNM rose from CHF 691 million to CHF 7.3 billion, an increase of 953.7%, and their median AUM growth reached 13.1%. Yet their limited scale constrains system-level impact.

The deceleration in organic growth reflects a structural shift in global wealth geography. New wealth is increasingly created in jurisdictions where Swiss and Liechtenstein institutions lack entrenched market presence. Digital competitors and agile international platforms have expanded rapidly in these regions, eroding the traditional advantages of established European centers.

Table 3: Total FTE (Switzerland and Liechtenstein)

Bank type	Total FTE (global, year-end)			Median FTE (global, year-end)		
	2025	2024	Change	2025	2024	Change
SWITZERLAND						
UBS	108,648	112,842	-3.7%	108,648	112,842	-3.7%
≥ CHF 20bln AUM	33,470	33,051	1.7%	767	832	-7.8%
< CHF 20bln AUM	4,917	4,675	5.2%	82	78	5.0%
Total Switzerland	**147,035**	**150,568**	**-2.4%**	**123**	**123**	**0.0%**
LIECHTENSTEIN						
≥ CHF 20bln AUM	2,257	2,227	1.4%	1,129	1,114	1.4%
Total Liechtenstein	**2,257**	**2,227**	**1.4%**	**1,129**	**1,114**	**1.4%**

Table 4: Growth KPIs

Bank type	2025 (CHF, median)			2024 (CHF, median)		
	AUM growth	NNM per AUM	NNM per FTE	AUM growth	NNM per AUM	NNM per FTE
UBS (in U.S.$)	6.3%	1.0%	0.5m	45.2%	1.8%	0.7m
≥ CHF 20bln AUM	11.8%	1.2%	1.1m	0.8%	3.2%	2.0m
< CHF 20bln AUM	13.1%	2.5%	1.4m	0.5%	1.6%	0.5m
Median	**12.5%**	**1.5%**	**1.1m**	**0.7%**	**1.8%**	**0.7m**

5. Profitability

Profitability improved across the sector in line with rising markets. Median return on AUM increased slightly from 0.86% to 0.87%, while individual results ranged from 0.12% to 2.91%. Return on Total Assets increased from 4.09% to 6.29% and Return on Equity increased from 4.75% to 7.18%, with values spanning from -24.58% to 32.21%.

Profitability levels increased only slightly, despite healthy AUM growth. Personnel expenses increased to even higher levels, and many institutions remain dependent on legacy IT frameworks and fragmented manual processes that limit scalability. Margin pressure from global asset managers and digital competitors worsens this trend. Revenue diversification remains limited relative to peers in private markets, alternatives, thematic strategies, and digital assets.

6. Efficiency

Operational efficiency continues to pose structural challenges. The median cost-income ratio improved slightly from 77.7% to 76.7%, but the improvement reflects higher revenues rather than cost reductions. Efficiency levels diverge dramatically: the lowest cost-income ratio reached 42.8%, while the highest exceeded 573%.

AUM per employee increased only marginally from CHF 57 million to CHF 58 million. This stagnation contrasts sharply with the upper bound of CHF 749 million per FTE observed among highly optimized institutions. Personnel expenses rose to CHF 271,000 per FTE, with some institutions surpassing CHF 441,000, indicating significant wage inflation.

The roots of inefficiency lie in outdated system landscapes, limited automation, and insufficient integration of AI-enabled processes. Fragmented digital infrastructures hinder scalability and impede productivity gains. As regulatory obligations intensify, these inefficiencies will deepen unless addressed structurally.

Table 5: Profitability KPIs

Bank type	2025 (median)			2024 (median)		
	Return on AUM	Return on Total Assets	Return on Equity	Return on AUM	Return on Total Assets	Return on Equity
UBS	0.90%	3.11%	5.94%	1.07%	2.97%	-0.38%
≥CHF 20b AUM	0.81%	3.94%	5.97%	0.86%	4.11%	8.18%
<CHF 20b AUM	0.87%	6.94%	7.53%	0.84%	4.15%	4.34%
Median	**0.87%**	**6.29%**	**7.18%**	**0.86%**	**4.09%**	**4.75%**

Table 6: Efficiency KPIs

Bank type	2025 (CHF, median)			2024 (CHF, median)		
	Cost-Income ratio	AUM per FTE	Personnel expense per FTE	Cost-Income ratio	AUM per FTE	Personnel expense per FTE
UBS (in U.S.$)	86.0%	49m	247k	57.5%	39m	220k
≥ CHF 20 bln AUM	79.6%	65m	288k	80.4%	65m	285k
< CHF 20 bln AUM	74.9%	51m	259k	76.4%	52m	250k
Median	**76.7%**	**58m**	**271k**	**77.7%**	**57m**	**260k**

7. Capital adequacy

Capitalization remains an unquestioned strength. The median CET1 ratio reached 28.5%, with values spanning from 12.5% to 118.1%. Leverage ratios displayed similar resilience with a median of 10.6%. Liquidity coverage ratios were extraordinarily high, with a median of 246% and peaks exceeding 3,000%.

These figures underscore strong balance-sheet resilience. Yet abundant capital can foster strategic complacency. High buffers do not mitigate structural erosion in competitive positioning, client acquisition, or cost efficiency.

8. Governance and board composition

Governance structures display heterogeneity. One-third of institutions maintain boards without female representation. Board size varies significantly, but competence diversity arguably matters more than numeric composition. Institutions lacking exposure to technological, international, or innovation-driven perspectives risk misjudging strategic risks and delaying necessary transformation. Inefficiencies observed in digitalization and business model evolution may reflect underlying governance constraints (which would need to be further examined).

Table 7: Capital adequacy KPIs

Bank type	2025 (median)			2024 (median)		
	CET1 ratio	Leverage ratio	Liquidity coverage ratio	CET1 ratio	Leverage ratio	Liquidity coverage ratio
UBS	14.3%	5.8%	188.4%	14.4%	5.5%	216.0%
≥ CHF 20 bln AUM	21.5%	7.1%	206.0%	21.1%	7.7%	229.0%
< CHF 20 bln AUM	32.9%	13.6%	330.0%	32.1%	12.8%	289.0%
Median	**28.5%**	**10.6%**	**246.0%**	**28.8%**	**11.0%**	**266.5%**

Table 8: Female members of board of directors

Number of female board members	2025		2024	
	Number of banks	Cumulative	Number of banks	Cumulative
0	23	33%	23	35%
1	24	68%	21	67%
2	14	88%	11	83%
3	6	97%	8	95%
4	0	97%	2	98%
5	2	100%	1	100%
Total	**69**	**100%**	**66**	**100%**

9. Strategic recommendations for forward-looking private banks

The evidence points to a need for strategic renewal. Strengthening organic client acquisition remains essential, particularly in regions that dominate contemporary wealth creation. NNM acquisition from the U.S., Asia-Pacific, and the Middle East should expand alongside digital-first client interaction models and specialized value propositions.

Digital transformation requires accelerated investment and decisive execution. Modernizing core systems, automating regulatory and operational workflows, and integrating AI into advisory processes could become central to restoring scalability and competitiveness. Technology should no longer be treated as a discretionary investment.

Cost structures should be realigned to match revenue potential. Rising personnel expenses require scalable operating models, targeted advisory segmentation, and selective outsourcing. Without structural adaptation, profitability could continue to erode.

Business and revenue models could further evolve toward greater diversification. Traditional advisory and discretionary mandates could become insufficient for sustainable revenue differentiation. Private markets, digital assets, and tokenization could offer relevant expansion paths.

10. Conclusion

Switzerland and Liechtenstein retain a powerful reputation in global wealth management and an exceptionally robust capital base. Yet the 2025 performance evidence shows that structural vulnerabilities are deepening. Declining NNM, rising personnel expenses, and uneven efficiency indicate that reliance on historical strengths is no longer sufficient. Market-driven AUM growth has obscured, but not resolved, underlying strategic challenges.

Sustained leadership in global wealth management will require proactive transformation. Strengthening international client acquisition, accelerating digital modernization, rebalancing cost structures, diversifying business models, and upgrading governance frameworks are essential steps. Both jurisdictions possess the institutional capabilities to succeed, but future competitiveness will depend on strategic decisiveness rather than legacy advantage.

References

Birchler, U. W., R. Hegglin, M. R. Reichenecker, and A. F. Wagner, 2016, "Which Swiss gnomes attract money? Efficiency and reputation as performance drivers of wealth management banks," SSRN, https://tinyurl.com/yc63eftd

Burgstaller, J., and T. D. Cocca, 2011, "Efficiency in private banking: evidence from Switzerland and Liechtenstein," Financial Markets and Portfolio Management 25:1, 75-93

Cowdery, E., 2025, "Navigating the unknown: why Switzerland stands out for wealth and asset management services," SPM insights, May 9, https://tinyurl.com/5apxenmy

Faust, M., 2019, "Private banking und wealth management – Ein Überblick über Marktsegmente und Leistungsangebote," in Brost, H., M. Faust, W. Reittinger (eds.), Private banking und wealth management: Strategien und Erfolgsfaktoren, Springer Gabler

Künzle, C., C. Baumann, and S. Feldmann, 2025, "Wealth management in Switzerland and Liechtenstein 2025," https://tinyurl.com/yck26c8x

Lagassé, J.-F., R. Wyss, C. Künzle, J. Saghatol Eslami, P. Spiller, S. Modheji, and D. Brandes, 2024, "The Deloitte international wealth management centre ranking 2024: seismic waves impacting leading financial centres," 5th edition, Deloitte Perspective, December 17, https://tinyurl.com/2p9p9k8b

Torbey, H., and J.-H. Kwek, 2025, "Asset management 2025: the great convergence," McKinsey and Co., September 18, https://tinyurl.com/mrx93ubv

Appendix

Table A1: Benchmarks in Swiss and Liechtenstein wealth management

Category	KPI	2025 (median)				2024 (median)			
		Median	Min	Max	Average	Median	Min.	Max	Average
Growth	AUM growth	12.5%	-1.1%	146.2%	15.0%	0.70%	-36.60%	95.80%	0.70%
	NNM/AUM	1.5%	-13.1%	71.7%	3.6%	1.80%	-56.60%	23.40%	0.10%
	NNM/FTE (CHF)	1m	-79m	36m	1m	1m	-504m	16m	-8m
Prosperity	Return on Total Assets	6.29%	2.04%	28.70%	7.37%	6.48%	1.40%	20.35%	7.18%
	Return on Equity	7.18%	-24.58%	32.21%	7.57%	7.98%	-38.00%	34.40%	9.16%
	Return on AUM	0.87%	0.12%	2.91%	0.95%	0.93%	0.10%	2.33%	1.01%
Efficiency	Cost-Income ratio	76.7%	42.8%	573.0%	85.3%	77.70%	38.40%	209.00%	77.40%
	AUM/FTE (CHF)	58m	19m	785m	87m	57m	20m	890m	90m
	Personnel expense/FTE (CHF)	271k	136k	441k	265k	260k	163k	415k	262k
Capital adequacy	CET1-ratio	28.5%	12.5%	118.1%	34.9%	28.80%	14.10%	118.00%	34.10%
	Leverage ratio	10.6%	4.1%	93.9%	15.5%	11.00%	4.00%	61.40%	14.60%
	Liquidity coverage ratio	246.0%	125.0%	3183.0%	422.7%	267%	128%	>843k%	>13k%

Future of Data Management

How AI is redefining transaction banking: from data to intelligence

Stephen Peters,
Founder, Averus.ai

Guy Moons,
Founder, Deeguisa Consulting

ABSTRACT

Transaction banking faces structural transformation as artificial intelligence (AI) and programmable settlement technologies converge. This article examines how leading institutions deploy production AI systems through integrated platforms spanning cash management, trade finance, and supply chain finance. Drawing on implementations at JPMorgan and DBS, alongside regulatory developments including MiCA and the U.S. GENIUS Act, we demonstrate measurable outcomes: 50% reduction in manual exceptions, 90% reduction in compliance false positives, and settlement compression from days to seconds. We argue that data network effects and platform economics will drive industry consolidation, as leading institutions accumulate advantages that mid-tier banks cannot replicate. The article concludes candidly: most banks have only begun the AI journey. Significant challenges remain in governance, workforce transition, and vendor readiness. Competitive advantage will accrue to institutions treating AI as strategic infrastructure requiring integrated platforms and organizational restructuring around client outcomes.

1. Introduction

Today's transaction banking landscape comprises solutions accumulated over decades across disparate technological platforms. The current drive towards harmonization of these technologies is the starting point towards breaking down the walls between the data stored in the banking eco system. However, the drive to ISO 20022 has clearly identified that many banks still favor converter-type approaches to harmonize the processing of their payments. Where some have looked at the ISO 20022 convergence as the holy grail to create a common base for the payment data, we see a continuation of proprietary data at the core of payment processing and a proliferation of ISO 20022 dialects across payment infrastructures.

Looking at other domains, the operational processing of trade finance remains paper-intensive where trade documents remain predominantly paper-based, with ICC surveys indicating 60-65% of banks and corporates identify paper documentation as a primary operational pain point [ICC (2020)]. While parts of a transaction may be digitized (e.g., application, messaging, compliance checks), fully paperless, end-to-end digital trades remain the minority, concentrated in closed ecosystems, specific corridors (e.g., Singapore/China), and large corporates with integrated banks and carriers.

Zooming in on the data, it is evident that key information for decision taking often is spread over multiple data sources stored in more data repositories. Accessing this siloed information is critical to optimize the overall process.

Fragmented architectures often reflect historical, vendor-specific decisions that constrain future choices. The key impact areas include the tight coupling between payment rails and proprietary engines, the high switching costs for modernization or consolidation, limited ability to orchestrate payments dynamically, and reduced negotiating leverage with vendors.

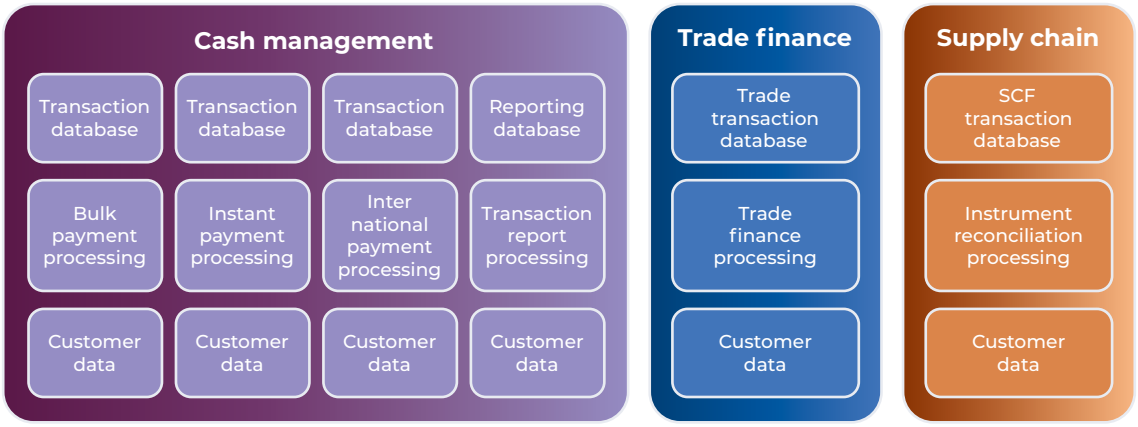
While architectural fragmentation manifests differently across business lines, the common outcome is the same: higher cost, poorer liquidity efficiency, weaker client experience, and constrained strategic optionality. In aggregate, this materially undermines the economics and resilience of corporate and transaction banking franchises.

Whilst the world is moving towards real-time processing at unprecedented speed, many processes in the banking eco-system relies on batch processing.

Fragmented payment and cash management architectures materially increase trapped and precautionary liquidity, particularly for real-time and cross-border flows.

These benchmarks lead to the following financial illustration; for an average payment-related liquidity pool of around €20bn, 10% of avoidable excess defines €2bn in trapped liquidity, which at 3-4% cost of funding translates into a €60-80m annual P&L drag.

Figure 1. Fragmented transaction banking architecture – current state



Batch processing | Manual handoffs | Siloed data

Table 1: Cross-Segment Comparison

Dimension	Corporate banking	Transaction banking	Wholesale payments
Primary pain	Client cash visibility	Cost and liquidity drag	Intraday liquidity and resilience
Main risk	Client attrition	Margin erosion	Systemic and operational risk
Key inefficiency	Manual reporting	Duplicate processing	Excess liquidity buffers
Strategic Impact	Weakened client franchise	Limited scalability	Loss of interbank relevance

Table 2: Typical benchmark impacts

Area	Fragmented architecture	More integrated architecture	Avoidable value-at-risk
Intraday liquidity buffers	15–30% excess versus true need [BIS (2023)]	5–10% excess	10–20%
Prefunding for instant payments	Often exceeds operational requirements	100–105%	15–45%
Nostro balances	Carry excess idle funds	<5–8% idle	5–12%
Liquidity forecasting accuracy	±15–20% [BankingHub (2024)]	±5–10%	Structural inefficiency

1.1 Structural inefficiencies

The world for consumer payments shifted from next day visibility to immediate visibility. This new paradigm pushes corporate clients to expect the same level of services not only for domestic but also for international transactions.

High on the corporate agenda are real time payments tracking, automated reconciliation, instant supplier payments, instant cash sweeping, instant working capital loans, just-in-time payments, instant tax payments, real time cash balances, etc. According to Capgemini (2024), these are the services not only high in demand but also generating a willingness to pay a premium. A quick reader sees that the common thread on these are real time processing, real time reporting, and real time reconciliation.

By combining instant payments, open banking, and real time screening banks are in the position to secure a strategic relationship with their prime corporates.

However, the hurdles to achieve this ideal eco system are multiple.

- **The payment settlement gap** (time between execution of a payment instruction) from the corporate point of view and the credit on the beneficiary account makes funds invisible from a payment logistics point of view. This gap in international payments typically ranges from same day to 1-2 days [FSB (2024)].
- **Liquidity forecasting error rates** show a variance of 15–20 % on intraday positions.
- Rule-based screening engines miss contextual risk information to reduce false positives, increasing **compliance friction**.

- For most clients. international payments remain **black box processing magic** with no visibility.
- **Manual** trade finance document verification adds 3-5 banking days to LC processing timelines per UCP 600 requirements [ICC (2007)].
- **Supply chain finance** is impacted by the availability of real time performance data during credit decisions or, in other words, the access to instant working capital loans.

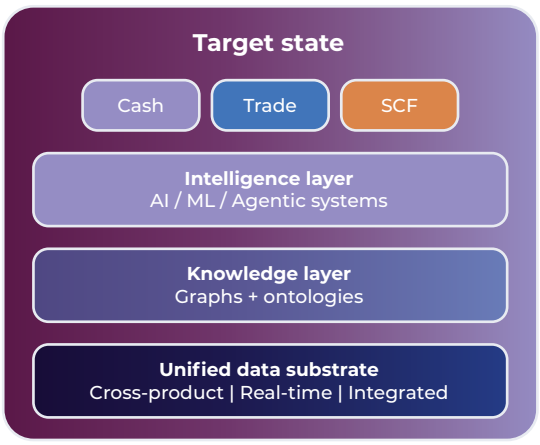
1.2 Product silo problem

Optimizing liquidity usage demands alignment of the different silos in the banking ecosystem. An integrated view on receivables and payables supports the unlocking of working capital and reducing the excess capital requirement for corporations.

The isolated data and document workflow in trade finance creates a distance between the working capital decision taking and the reality of the cash balances. This often leads to the setup of static credit limits that are not responsive to real time customer - supplier behaviors and are unable to integrate seasonality in the overall dynamics.

Transaction banks have spent lots of effort and money to streamline the processes often through the acquisition of best-in-class vendor solutions. The overarching solution often lacks superiority because the different components of the solution do not integrate effectively.

Figure 2. Integrated transaction banking architecture – target state



Our experience has seen banks with different payment solutions for different payment schemes based on different technologies (or vendors) whereby the customer experience lacks the common high standard that is imposed on each individual system. It is obvious for the customers to notice this differentiated user experience.

1.3 Market pressure

Summarizing studies on the market expectations of corporate users can be reduced to one word, instant. Corporate users are seeking solutions from their banks to get access to real time processing, real time reporting, real time reconciliation, and real time decisioning.

Transaction banks that can address these expectations can expect to lock-in their most profitable customers.

Conversely, fintech unbundling will try to address isolated use-cases to take advantage of the most profitable components of the corporate/bank relationship. PayPal and Stripe together now capture approximately 25% of the global cross-border payments market, with traditional banks' share declining from 55% in 2015 to 30-35% today [Grand View Research (2024)]. They are looking at entering the business-to-business transaction flows.

Next to the availability of the data, the predictive nature becomes paramount. Being able to make liquidity management decisions across cash, trade, and SCF will define the winners on this journey.

The never-ending regulatory expansion (MiCA, Basel III, etc.) continues to increase the complexity for the banks and will drive the need for more processing power in the financial value chain.

Section 2 examines the intelligence layer that addresses these structural deficiencies. How to break down the data silo's and create a knowledge layer that supports the real time decisioning.

2. The intelligence layer

Transaction banking's fragmentation is more than inefficient - it blocks what corporate clients now demand: real-time visibility, explainable decisions, and competitive costs. The solution lies in a new intelligence layer that transforms how banks process, understand, and act on transaction data. Leading banks have already responded with production AI-at-scale. JPMorgan deploys 450 AI use-cases backed by an U.S.$18 billion annual technology budget [JPMorgan (2024)]. DBS generated S$750 million in value from AI in 2024, running 1,500 models across 370 use cases [DBS (2024)]. These are production systems, not pilots.

What makes these deployments effective is how they address specific fragmentation symptoms through targeted architectural solutions. Siloed products create siloed data, so banks need a unified model linking cash, trade, and supply chain finance. Batch processing creates lag, requiring streaming analytics. Manual exception handling scales poorly, necessitating AI triage. Flat customer records miss relationship complexity: banks need graph databases and semantic ontologies.

2.1 Three AI capability tiers

These architectural requirements translate into distinct AI capabilities. Banks deploy AI across three maturity levels, each building on the one before and requiring progressively deeper data integration.

Each tier addresses different banking challenges. The following sections examine how leading institutions deploy these capabilities in production.

2.1.1 Predictive analytics

Among predictive analytics applications, liquidity forecasting is the most mature use-case. Banks forecast at different horizons with different precision needs. Intraday positioning demands high accuracy over short windows. Weekly forecasts balance precision against payment volatility. Monthly planning tolerates wider ranges for strategic visibility.

HSBC's AI Markets platform exemplifies this approach, allowing treasurers to interrogate hedging strategies against real-time liquidity, historical pricing, and volatility forecasts through natural language queries [HSBC (2023)].

Anomaly detection extends well beyond fraud. It flags payment deviations by customer, corridor, and time window. A spike in exception queues or approaching SLA breaches enables intervention before service levels degrade.

Credit scoring for trade counterparties draws on real-time transaction data. Payment behavior reveals stress signals weeks before they appear in financial statements.

2.1.2 Large language models (LLMs)

The high-value LLM applications are document-centric. Payment instruction parsing transforms structured (pain.001) and unstructured inputs (emails, PDFs, ERP exports) into validated ISO 20022 fields. Trade document extraction normalizes invoices, bills of lading, and certificates into structured data with field-level provenance. The results are measurable. JPMorgan reports AI models cut manual exceptions by over 50% even as transaction volumes rose by half. Account rejection

Table 3: AI capability tiers

Tier	What it does	Applications	Decision speed
Predictive analytics	Forecasts from historical patterns	Liquidity forecasting, fraud and anomaly detection, credit scoring	Milliseconds to hours
LLMs	Processes unstructured text	Document extraction, compliance checking, parsing	Seconds
Agentic orchestration	Chains decisions autonomously	End-to-end trade processing, cross-product optimisation	Real-time

Table 4: Knowledge layer components

Component	What it captures	Why it matters
Entity graphs	Customers, counterparties, legal entities, ownership chains	Links fragmented records into unified relationship networks
Transaction graphs	Payment flows, trade relationships, financing chains	Reveals patterns invisible in tabular transaction logs
Ontologies	Standardised definitions of products, processes, risks	Enables consistent reasoning across product silos
Knowledge embeddings	Semantic relationships between concepts	Allows AI to understand context, not just match keywords

rates dropped 15-20% through AI-powered payment validation [JPMC (2024)].

LC compliance checking maps documentary requirements clause by clause against submitted documents. Bank implementations report 99% accuracy on standard trade documents with processing times cut from days to two hours.

Fraud detection benefits from pattern recognition humans cannot match. DBS reports a 90% reduction in compliance false positives through AI-powered contextual analysis rather than keyword matching [DBS (2024)]

2.1.3 Agentic orchestration

The most advanced tier combines predictive analytics and LLM capabilities into autonomous workflows. Agentic systems chain multiple AI capabilities together, enabling end-to-end automation. A trade finance agent might extract document data, score compliance risk, check credit limits, and initiate settlement - all without human intervention for standard cases.

This tier remains early-stage. Current implementations keep humans in the loop for high-value decisions. The technology works: governance and control frameworks are still maturing. Yet, all three tiers share a common dependency: they require rich, relationship-aware data to function effectively.

2.2 The knowledge layer: graphs and ontologies

Between raw data and AI applications sits a critical intermediate layer that most banks underestimate: the knowledge layer. This layer uses graph databases and semantic ontologies to represent the complex relationships that define transaction banking. Without it, AI models operate on flattened, impoverished, and incomplete views of reality.

2.2.1 Graph databases

Tabular records cannot represent the relationship density in corporate banking. A multinational corporation may have dozens of legal entities, each with separate accounts across cash management, trade finance, and supply chain finance. Ownership structures span jurisdictions. Counterparty relationships form networks, not lists. Payment flows create webs of interdependency.

Graph structures capture these naturally. They enable queries that relational databases struggle with: “Show all entities within two hops of this counterparty.” “Trace the ultimate beneficial owner through four holding companies.” “Identify all suppliers who also bank with competitors.”

When a sanctions match requires understanding beneficial ownership, graph queries find connections that table joins would miss. When credit risk assessment needs counterparty network exposure, graphs reveal concentration that entity-by-entity analysis obscures.

Standard Chartered’s trade finance operations demonstrate the value of graph-based knowledge representation [Standard Chartered (2024)]. Their documentary credit processing involves counterparties across 40+ markets, each with distinct regulatory requirements and correspondent banking relationships. Graph structures enable real-time traversal of these relationships, reducing compliance verification from hours to minutes while

simultaneously improving detection accuracy for sanctions screening and beneficial ownership analysis.

2.2.2 Semantic ontologies

Ontologies provide the vocabulary that makes cross-product intelligence possible. Cash management, trade finance, and supply chain finance use different terminology for overlapping concepts. A "customer" in cash management may be a "buyer" in SCF and an "applicant" in trade finance, yet they represent the same entity.

Semantic ontologies standardize these definitions. They create a common language that AI models can reason over. When an LLM processes a trade document, the ontology tells it that "beneficiary" in an LC context maps to "supplier" in the SCF context and "payee" in cash management.

This standardization enables cross-product queries that would otherwise require manual translation. The ontology layer sits between product silos and AI applications, translating product-specific data into unified concepts.

The ISO 20022 message standard provides a foundational ontology for payments and trade finance. Its structured data elements (party identification, account structures, and transaction references) offer a starting vocabulary that banks can extend with proprietary classifications. The migration deadlines (SWIFT cross-border by November 2025, most domestic schemes by 2025-2027) create natural inflection points for ontology deployment.

Implementation follows a layered approach. Core ISO 20022 definitions provide the base vocabulary. Bank-specific extensions add proprietary product features. Industry utilities (SWIFT, correspondent networks) contribute shared reference data. The resulting ontology enables queries across all three layers while maintaining compatibility with external counterparties using standard definitions.

2.2.3 The integration pattern

The knowledge layer serves as the semantic bridge between fragmented data sources and unified AI applications.

Product databases feed into graph ingestion pipelines that extract entities and relationships. Ontology mappings normalize terminology. The resulting knowledge graph provides AI models with a unified, relationship-rich view that no single product database contains.

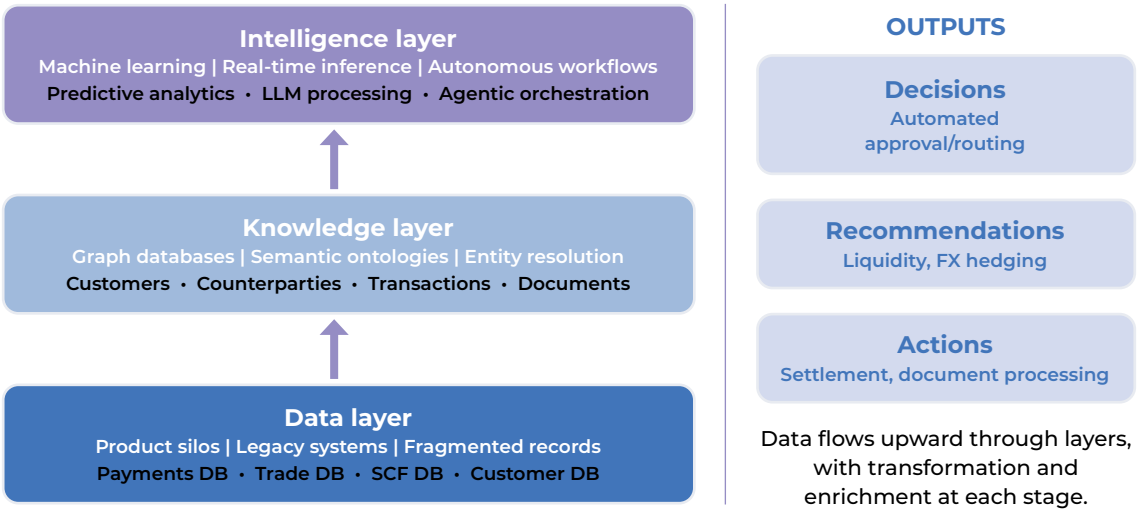
This architecture explains why banks with sophisticated AI but weak knowledge infrastructure fail to meet their objectives. The models are only as good as the relationships they can see.

2.3 Data architecture

The knowledge layer depends on underlying data infrastructure that addresses three fundamental challenges in transaction banking AI.

Unified data substrates matter because AI models can only reason about data they can access. When a corporate client holds cash management, trade

Figure 3: Three-layer architecture



finance, and supply chain finance products across separate systems, each with its own customer identifier, the bank cannot assess total relationship exposure or identify cross-selling opportunities. A unified customer key that spans all products transforms fragmented records into a complete client view. This is not merely a data hygiene exercise; it is the prerequisite for any AI that needs to understand client behavior across product boundaries.

Entity resolution presents the core technical challenge. The same multinational may appear as "Acme Corp" in cash management, "Acme Holdings Ltd" in trade finance, and "Acme Singapore Pte Ltd" in supply chain finance. Linking these requires probabilistic matching on company names, addresses, registration numbers, and beneficial ownership structures. Graph databases excel here because they can represent tentative relationships with confidence scores, allowing AI models to reason about probable entity matches rather than requiring perfect data quality.

Streaming pipelines address the temporal dimension. AI predictions made on stale data produce stale recommendations. For intraday liquidity optimization or real-time fraud detection, data must flow from source systems to AI models with sub-second latency. Feature stores standardize how this data reaches models, ensuring consistent inputs whether for training or inference, and enabling features developed for one use-case to be reused across others. Finally, lineage tracking provides the audit trail that regulators require: the ability to trace any AI output back to the source data that produced it.

2.4 Controls and governance

Data architecture without governance remains an academic exercise. In regulated banking environments, governance is what enables AI systems to move from prototype to production. Accenture (2025) found that 63% of financial institutions lack governance frameworks for generative AI (GenAI). This explains why many pilots never reach production.

Banks need three control layers:

- **Data governance**: covers privacy, residency, retention, encryption, and access controls.
- **Model risk management**: requires validation protocols, stress testing, drift monitoring, and explainability standards.
- **Human-in-loop design**: ensures AI augments rather than replaces judgment on consequential decisions.

2.5 Platform integration

With data architecture and governance in place, banks can assemble these components into an integrated platform. The platform imperative is integration - three elements must work together.

- **A unified data substrate**: serves as the single source of truth. Common keys link transactions, documents, and customer relationships.
- **A knowledge layer**: built on graphs and ontologies translates product-specific data into unified concepts that AI can reason over.
- **A shared intelligence layer**: makes features and models reusable across products. A central registry prevents fragmented development.

The payoff is network effects. More products generate more signals. More signals enrich the knowledge graph. Richer graphs produce better predictions. Better predictions enable higher-margin services. BCG (2024) estimates retail banks could capture U.S.$370 billion in additional annual profits by 2030 through AI.

2.6 Adoption pathway

For banks beginning this journey, implementation sequence matters as much as capability choice. A phased approach reduces risk while building organizational capability.

- **Start with high-ROI, lower-risk workflows**: exception triage, document checking, and routing optimization. These build data feedback loops and governance capability.
- **Invest early in the knowledge layer**: graph databases and ontologies take time to build but unlock cross-product intelligence that siloed AI cannot achieve.
- **Add cross-product optimization**: once governance and knowledge infrastructure are proven in production.
- **Agentic orchestration comes last**: after tool integration is validated and the organization can supervise automated decisions at scale.

Section 3 examines programmable settlement: the infrastructure that matches AI decisioning speed with instant, atomic transaction completion.

3. Programmable settlement

The paradox at the heart of modern transaction banking: AI systems can generate actionable intelligence in milliseconds, yet settlement infrastructure designed in the 1970s requires days to execute. An optimal FX rate identified at 9:00 AM becomes suboptimal by 9:15 - but cannot be captured until tomorrow. A liquidity shortfall predicted for Thursday demands action on Tuesday, while the payment rails needed to act observe local banking hours and batch windows.

This temporal mismatch is not a minor inefficiency. It is the binding constraint on everything Section 2 described. Predictive models, risk engines, and intelligent routing algorithms all generate recommendations that decay with time. The value of a perfect forecast approaches zero when execution latency exceeds the forecast horizon.

3.1 The settlement problem

Traditional settlement infrastructure cannot keep pace with algorithmic decision-making. Cross-border payments take one to two days through correspondent banking networks that route transactions sequentially across nostro accounts, each hop adding latency and cost. Letters of credit require manual document review against paper-based shipping records. Treasury teams work around the clock managing liquidity across time zones while settlement rails observe local banking hours and batch processing windows.

The friction is structural, not operational. SWIFT messages pass through multiple intermediaries. Compliance checks execute sequentially rather than in parallel. Cut-off times fragment the trading day into disconnected windows. A payment initiated in Singapore at 4 PM may not settle in New York until the following afternoon - by which point the FX rate, the liquidity position, and the risk calculus have all shifted.

For AI-driven treasury, this latency is not an inconvenience but a fundamental constraint. The intelligence layer described in Section 2 generates continuous recommendations: optimal payment timing, dynamic hedge adjustments, and liquidity rebalancing across entities. Each recommendation carries an expiry. When execution requires 24-48 hours, the window for value capture has often closed before settlement begins.

3.2 Digital asset settlement infrastructure

Three models now offer programmable settlement for institutional use. Each addresses the latency problem differently, but all share the essential characteristic that AI-driven treasury systems require execution speed measured in seconds rather than days, enabling action within the window where algorithmic recommendations retain their value.

3.2.1 Stablecoins

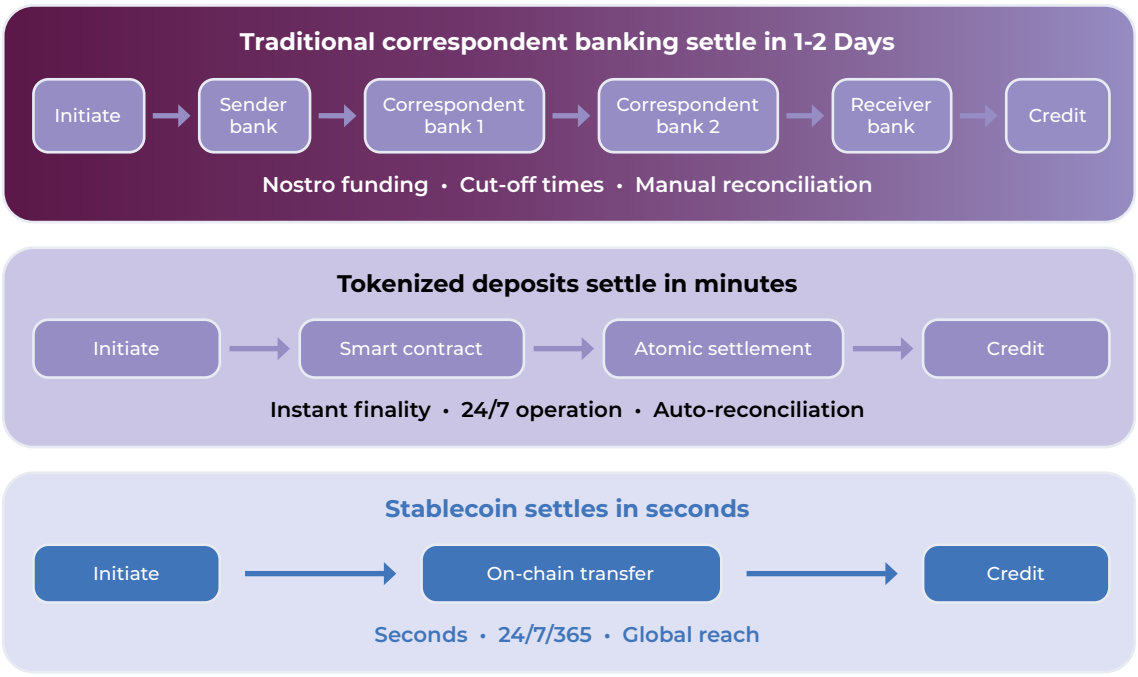
Global stablecoin circulation now exceeds U.S.$300 billion [CoinGecko (2025)]. U.S.-issued tokens command 99% market share. Tether's USDT accounts for roughly U.S.$140 billion; Circle's USDC holds another U.S.$60 billion.

Settlement finality occurs in seconds rather than days, eliminating reconciliation burdens. Smart contracts encode payment logic directly. Networks run 24/7/365. For AI-driven treasury operations, this means the optimal FX rate identified at 9:00 AM can be captured at 9:00 AM - not queued for next-day settlement while the opportunity evaporates.

Table 5: Comparison of programmable settlement instruments

Feature	Stablecoins	Tokenised deposits	Wholesale CBDCs
Issuer	Non-bank (Circle, Tether)	Commercial bank	Central bank
Deposit insurance	No	Yes (where applicable)	N/A (sovereign)
Settlement speed	Seconds	Seconds to minutes	Seconds to minutes
24/7 operation	Yes	Platform-dependent	Platform-dependent
Current scale	$300B+ circulation	$2B+ daily (Kinexys)	Pilot stage

Figure 4: Settlement flow comparison



3.2.2 Tokenized deposits

Tokenized deposits place bank money on distributed ledger infrastructure while maintaining deposits as liabilities of the issuing bank. Deposit insurance remains intact.

JPMorgan's Kinexys platform has processed over U.S.$1.5 trillion since 2020 [JPMorgan (2024)], with daily volumes exceeding U.S.$2 billion [JPMorgan (2025)]. Critically, programmable settlement provides the execution layer that AI requires when an AI model identifies a liquidity optimization or payment routing decision, tokenized settlement executes that decision in seconds rather than days. Fnality International's wholesale payment system, backed by fifteen major banks, enables settlement in tokenized central bank money with true finality [Fnality (2024)].

3.2.3 Wholesale CBDCs

Central bank digital currencies (CBDCs) extend programmability to central bank money. The BIS mBridge project connects central banks in China, Hong Kong, Thailand, and the UAE [BIS (2024)]. Project Dunbar explores multi-CBDC platforms. While still at pilot stage, wholesale CBDCs promise the ultimate settlement layer for AI-driven cross-border treasury: sovereign-backed digital money that settles atomically across jurisdictions, eliminating the correspondent banking latency that currently constrains algorithmic execution.

3.3 Trade and supply chain finance applications

These settlement instruments find immediate application in trade and supply chain finance, where documentary complexity and multi-party coordination have traditionally created the longest settlement delays.

3.3.1 Tokenized letters of credit

A smart contract holds funds in escrow against documentary requirements. The LLM verifies compliance. Upon verification, the contract releases payment without manual queues. Standard Chartered and HSBC demonstrated this approach in 2024 with a blockchain-based LC transaction that reduced processing time from 5-10 days to under 24 hours. The trial tokenized both the LC instrument and underlying documentation, enabling atomic settlement where document verification and payment release occurred simultaneously. This eliminates the settlement risk inherent in sequential processing while providing immutable audit trails for regulatory compliance.

3.3.2 On-chain supply chain finance

Invoice tokenization transforms receivables into tradeable assets. Settlement occurs instantly upon buyer payment. Taulia's integration with SAP demonstrates the operational model: approved invoices automatically generate tokenized receivables that funders can purchase through smart contract auctions. Dynamic discounting adjusts pricing in real-time based on payment timing, with settlement occurring atomically when the buyer remits. The platform has grown transaction volumes from U.S.$500 billion to U.S.$800 billion over three years [SAP (2025)], demonstrating the scale achievable when programmable settlement replaces manual approval workflows.

3.3.3 Programmable trade loans

Drawdown and repayment trigger on observable events. Shipment departure triggers loan advance. Customs clearance triggers the next tranche. AI models monitoring IoT sensors and shipping data can trigger these tranches automatically when conditions are met, eliminating the manual verification delays that traditionally separate event occurrence from fund release.

These trade finance applications share a common architectural principle: programmable settlement transforms the execution layer from a bottleneck into an enabler. The pattern extends naturally to cross-border treasury operations.

3.3.4 Cross-border settlement compression

Traditional correspondent banking requires sequential settlement across multiple intermediaries, each adding latency and cost. Programmable settlement enables netting across corridors and currencies before final settlement occurs. A multinational treasury with exposures across EUR, USD, and SGD can net positions continuously, settling only the residual through traditional rails. Multilateral netting demonstrates substantial efficiency gains, with CLS Group reporting netting efficiency of approximately 96%, reducing funding requirements to less than 1% of gross settlement value [CLS (2024)].

3.4 Regulatory landscape

Institutional adoption depends on regulatory clarity. Banks cannot deploy programmable settlement at scale without certainty on reserve requirements, issuer licensing, customer protection, and cross-border recognition. The landscape varies by jurisdiction but shows meaningful advancement - particularly in jurisdictions competing for digital asset leadership.

The regulatory trajectory is unmistakably toward accommodation rather than prohibition. Europe's MiCA framework establishes comprehensive rules for stablecoin issuers. Singapore's MAS has positioned the city-state as a sandbox for tokenized deposits and programmable money through Project Guardian. The U.S., after years of regulatory ambiguity, passed the GENIUS Act in July 2025 with bipartisan support - signaling that stablecoins have crossed from crypto speculation into mainstream payments infrastructure.

What matters for transaction banking is not the detail of each framework but the pattern: major jurisdictions are creating legal foundations for digital settlement instruments. The question is no longer whether programmable settlement will be regulated, but how quickly banks can build compliant implementations on these emerging frameworks.

3.5 Adoption trajectory

McKinsey (2024) projects stablecoin circulation reaching U.S.$5 trillion by 2030. BCG (2024) estimates tokenized illiquid assets will reach U.S.$16.1 trillion, a fifty-fold increase from 2022. SWIFT has announced a blockchain-based shared ledger developed with Consensys, with over 30 institutions involved [SWIFT (2025)]. These projections matter because they indicate the settlement infrastructure will exist at scale precisely when AI-driven treasury systems require it.

3.6 Architectural convergence

The preceding sections have examined AI capabilities and programmable settlement infrastructure separately. Their true significance emerges only in combination. The intelligence layer and programmable settlement combine to produce something neither achieves alone: the autonomous treasury.

The predictive model identifies an FX exposure. Tokenized settlement executes the swap. The LLM verifies LC compliance. The smart contract releases payment. The routing optimizer selects the lowest-cost path. Settlement completes in seconds.

AI without programmable settlement generates recommendations that decay before execution.

Programmable settlement without AI executes blindly. The convergence creates the product that matters.

Section 4 examines how these capabilities transform operational reality across cash management, trade finance, and supply chain finance - translating architectural possibility into measurable process improvement, cost reduction, and competitive differentiation.

4. Operational transformation

The preceding sections established how AI architectures and programmable settlement create new possibilities. This section examines how these capabilities translate into operational reality across cash management, trade finance, and supply chain finance.

4.1 The transformation pattern

Legacy transaction banking followed a consistent model: sequential processes with human review at every decision point. Each step introduced latency. Each handoff created error potential. AI-enabled operations invert this logic through parallel processing, autonomous routing of standard transactions, and human review reserved for genuine exceptions.

4.2 Cash management

Intelligent cash management replaces batch collection with streaming data ingestion. AI risk scoring evaluates each payment against learned patterns, flagging anomalies while routing normal transactions autonomously.

JPMorgan's results illustrate the operational leverage. AI models reduced manual exceptions by over 50% even as transaction volumes rose by half. Payment validation decreased account rejection rates by 15-20% [JPMorgan (2025)].

Four capabilities define the new paradigm:

- **Predictive liquidity management**: forecasts funding needs 72 hours forward, analyzing historical patterns, scheduled payments, and seasonal variations. Corporate treasurers get rolling predictions that adapt to changing conditions rather than static weekly forecasts.
- **Dynamic compliance**: uses LLMs to interpret regulatory changes and translate them into screening logic. A new sanctions list or revised AML guidance updates screening parameters without manual rule coding.
- **Autonomous reconciliation**: matches payments against invoices and expected flows. Leading reconciliation platforms achieve transaction auto-match rates of approximately 90% and journal posting automation of 95%, with self-learning algorithms exceeding 95% matching accuracy [HighRadius (2026)].
- **Intelligent Payment Routing**: the fourth capability, payment routing optimization represents an emerging capability where AI evaluates multiple execution paths in real-time. For a cross-border payment, the system simultaneously assesses correspondent banking routes, blockchain settlement options, netting opportunities, and currency conversion timing. Variables include transaction costs, settlement speed, counterparty credit exposure, and regulatory constraints. Early implementations demonstrate 15-30% cost reduction on high-volume corridors through dynamic path selection [Payments Association

Table 6: Legacy versus AI-enabled operations

Dimension	Legacy model	AI-enabled model
Processing	Batch (overnight/periodic)	Streaming (continuous)
Decision points	Human at every step	Human for exceptions only
Learning	Manual rule updates	Continuous from outcomes
Scaling	Linear staffing growth	Absorbs volume without headcount
Compliance	Keyword matching (high false positives)	Contextual analysis (95% fewer)

(2025)]. The knowledge layer provides the relationship graph needed to evaluate counterparty networks, while settlement layer integration enables execution through optimal channels without manual intervention.

4.3 Trade finance

Trade finance presents greater complexity. The product has relied on physical documents (bills of lading, letters of credit, certificates of origin) since medieval banking. A typical LC transaction consumed 5-10 days with first-presentation discrepancy rates of 50-75% [ICC (2020)].

The intelligent flow digitizes document handling from ingestion forward. LLMs extract relevant fields and validate them against transaction parameters. For standard transactions with clean documentation, LC processing drops to hours. Table 7 quantifies these improvements across the trade finance process chain.

Beyond process acceleration, AI enables capabilities that were previously impractical.

- **Fraud detection**: represents a capability AI enables rather than merely accelerates. Machine learning identifies anomalies invisible to human reviewers: metadata inconsistencies, subtle letterhead variations, statistical improbabilities in cargo descriptions. DBS reports a 90% reduction in compliance false positives through AI-powered contextual analysis [DBS (2024)].
- **Electronic bills of lading**: reached 11% adoption in 2025, up from approximately 1% in 2021 [DCSA (2025)]. McKinsey (2024) estimates full eBL adoption would save U.S.$6.5 billion annually across the industry. Yet, digitization alone does not address the fundamental bottleneck in LC processing: document discrepancies.
- **Discrepancy resolution**: the 30-40% exception rate in traditional LC processing reflects document discrepancies that require manual adjudication. AI-powered discrepancy resolution transforms this bottleneck. LLMs now classify discrepancies by type (typographical, substantive, and technical), assess materiality against UCP 600 standards, and recommend resolution pathways. For minor discrepancies (date formats, spelling variations, and reference number mismatches), systems can auto-accept with audit logging. Substantive discrepancies route to human reviewers with AI-generated summaries of risk factors and precedent handling.

Organizations implementing trade finance automation report processing time reductions of up to 75%, with AI systems reducing document review tasks from days to minutes, with no increase in post-settlement disputes.

4.4 Supply chain finance

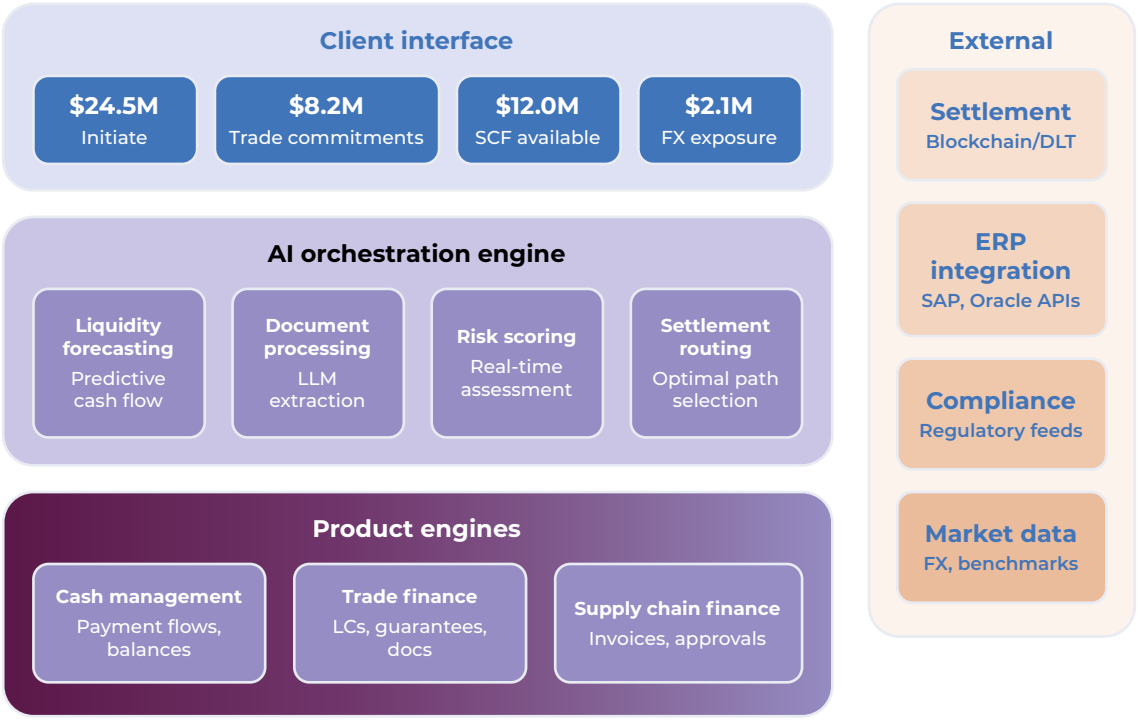
Supply chain finance applies similar AI-driven transformation to buyer-supplier financing relationships. The legacy SCF flow (invoice submission, buyer approval, bank verification, and financing offer) consumed 5-15 days. The intelligent flow compresses this timeline through automation at each stage.

- **Invoice auto-validation**: compares submissions against purchase orders already on file.
- **AI credit scoring**: evaluates supplier risk using data unavailable to traditional models: payment behavior patterns, industry distress indicators, and news sentiment.
- **Dynamic discount offers**: calculate optimal pricing based on buyer credit, supplier risk, and market rates.

Table 7: Trade finance process comparison

Process step	Legacy	AI-enabled	Improvement
Document review	3-5 days manual	2 hours automated	99% accuracy
Compliance check	Keyword matching	Contextual NLP	95% fewer false positives
Fraud detection	Post-facto investigation	Pre-transaction screening	95% detection accuracy
Credit assessment	Days for dossier	Seconds for risk score	Real-time decisioning

Figure 5: Integrated working capital platform architecture



Leading three-way matching implementations have achieved 90%+ straight-through rates by learning legitimate variation patterns that distinguish genuine discrepancies from clerical noise.

- **Supplier risk prediction**: may offer the most significant advance. Predictive models flag supplier deterioration 60-90 days before traditional monitoring would detect problems [SR Analytics (2026)]. Early warning enables buyers to diversify sourcing before disruption.

4.5 Cross-product integration

The transformations in cash management, trade finance, and supply chain finance described above achieve their full potential when integrated into unified platforms. Cross-product intelligence emerges when AI analyses consolidated positions across cash, receivables, payables, trade commitments, and FX exposure.

A corporation with surplus cash in one currency, payables due in another, an expiring LC, and receivables eligible for early payment discounts typically manages these elements separately. An AI system perceiving all positions simultaneously identifies that drawing on the LC, converting surplus cash at favorable rates, and offering early payment terms creates better outcomes than managing each element independently.

The knowledge layer architecture described in Section 2 provides the foundation for this integration. Graph databases maintain relationship structures across products: the same corporate entity appears in cash management as an account holder, in trade finance as an LC applicant, and in SCF as a buyer. Semantic ontologies translate product-specific terminology into unified concepts that AI models can reason over. Without this knowledge infrastructure, cross-product AI remains limited to simple data aggregation rather than genuine insight generation.

This integration of technology and process demands corresponding changes in how banks organize their transaction banking operations.

4.6 Organizational impact

The operational changes described above necessitate fundamental shifts in workforce roles and organizational structures. Operations roles shift from processing to exception architecture:

designing rules that determine which transactions require human review, and training models that handle routine cases. Treasury functions evolve from tactical cash positioning and payment timing toward strategic financing decisions and risk policy oversight. Risk teams move beyond transaction-level review to focus on model governance and drift detection. Trade specialists, once consumed by document checking, now concentrate on validating AI outputs and handling novel document types that fall outside established patterns.

The transformation is real but incomplete. Integration challenges (connecting legacy systems, maintaining data quality, and governing model behavior) continue to constrain implementation.

Section 5 examines the strategic implications of these operational shifts - how process-level transformation translates into competitive positioning, market structure, and long-term strategic advantage.

5. Strategic implications

The operational capabilities examined in Section 4 (AI-driven exception handling, predictive analytics, and programmable settlement) create value only when banks convert them into durable competitive advantage. This section examines how these capabilities drive competitive separation through data network effects, platform versus silo economics, defensive positioning against emerging threats, and the resulting industry structure.

5.1 Competitive separation through data network effects

Data network effects represent the most powerful source of AI-driven competitive advantage. Unlike traditional technology investments where hardware and software can be replicated, AI-based competition creates informational asymmetries that compound over time.

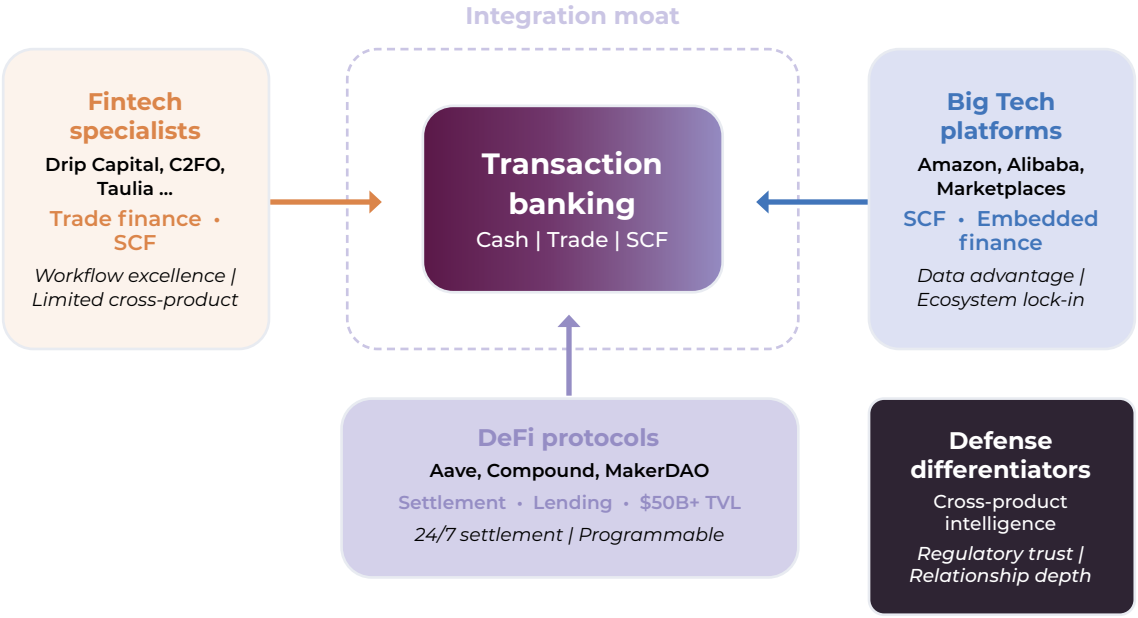
Banks that deploy AI first accumulate training data that improves model performance. Better performance attracts more volume. More volume generates more training data. A competitor commissioning identical technology in 2027 will lack the labelled exceptions and validated predictions a first mover accumulated since 2024.

JPMorgan's U.S.$18 billion annual technology budget and 450 deployed AI use cases illustrate the scale [JPMorgan (2024)]. DBS reports S$750 million in tangible economic impact from 1,500 models [DBS (2024)]. Accenture (2024) estimates 73% of U.S. bank employee time has high AI impact potential.

Table 8: Margin improvement drivers

Driver	Mechanism	Impact
Automation cost reduction	AI handles exceptions, reconciliation, documents	7.7% OpEx reduction
Cross-product intelligence	Unified data enables predictive services	25-50bp pricing premium
Platform integration	Behaviour-matched cross-sell	300-500bp margin capture
Settlement compression	Tokenised instant settlement	Reduced float costs
Credit loss reduction	Real-time early intervention	Lower provisions

Figure 6: Competitive threat landscape



5.2 Platform economics versus product silos

Data network effects accelerate most rapidly when banks unify previously siloed products. Cash management, trade finance, and supply chain finance typically operate as separate businesses, but AI inverts this logic: the most valuable predictions require data spanning product boundaries.

Banks with unified platforms can settle a trade finance transaction, adjust the FX hedge, and update SCF credit lines in a single atomic operation. Programmable settlement infrastructure (such as JPMorgan's Kinexys, processing U.S.$2 billion daily) provides the execution speed that AI-driven decisions require. Siloed competitors execute sequential transactions with gaps between each step.

Banks increasingly price treasury management services on a relationship basis, with integrated working capital solutions creating customer stickiness and generating both fee income and low-cost deposits. Platform banking creates seamless experiences through open APIs, with institutions reporting improved revenue per relationship and reduced cost per account compared to siloed product delivery.

5.3 Competitive threats

Transaction banking faces pressure from three directions.

- **Fintech unbundling**: has progressed in trade and supply chain finance. Specialists like Drip Capital and C2FO excel at specific workflows but lack cross-product intelligence.
- **Big technology entry**: poses a fundamental threat. Amazon and Alibaba have transaction data visibility banks cannot match.
- **DeFi protocols**: can no longer be dismissed. Decentralized lending exceeded U.S.$50 billion TVL as of January 2025 [DeFiLlama (2025)]. Protocols like Aave and Compound operate continuously with algorithmic rate-setting, establishing expectations for automation that challenge traditional banking models.

Against these converging threats, the defensive strategy is integration. Banks cannot compete with fintechs on narrow workflows, with technology platforms on data access, or with DeFi on automation speed. The defensible position lies in cross-product orchestration: optimizing a corporate's complete working capital position across payment timing, receivables, trade credit, and FX - a capability that requires the regulatory standing, balance sheet, and relationship breadth that only banks possess.

Table 9: Three-tier competitive structure

Tier	Characteristics	Requirements	Examples
Global AI leaders	Proprietary platforms, cross-product integration	$10B+ tech investment, blockchain capability	JPMorgan, DBS, StanChart
Regional utilities	Shared infrastructure, relationship competition	Strategic partnerships, utility access	Mid-tier regional banks
Specialists	Focused expertise, narrow scope	Deep domain capability, agility	Tradeshift, C2FO

5.4 Industry structure

The competitive dynamics of data network effects, platform economics, and converging threats are reshaping transaction banking into a three-tier structure where strategic positioning determines survival.

McKinsey (2024) estimates AI could add U.S.$200-340 billion annually to global banking. Banks capturing none, face margin compression forcing transformation or exit.

Networks like Fnality Payment System extend utility logic to programmable settlement, enabling banks to offer tokenized services without proprietary blockchain infrastructure.

5.5 Real-time versus batch economics

Underpinning all three competitive tiers is a fundamental infrastructure choice: batch versus real-time processing. This distinction shapes competitive positioning as profoundly as data network effects or platform integration. Batch processing historically allowed banks to aggregate risk, net positions, and optimize execution across daily windows. Real-time processing eliminates these efficiencies while creating new value propositions that increasingly define competitive advantage.

The economics favor real-time when: (1) the value of immediate certainty exceeds batch aggregation savings, (2) the cost of trapped liquidity outweighs netting benefits, and (3) competitive pressure demands instant execution. For high-value trade settlements, supply chain financing decisions, and cross-border treasury movements, these conditions increasingly apply.

Banks maintaining batch-only infrastructure face an asymmetric competitive threat: real-time competitors can serve batch customers (by aggregating real-time transactions), but batch processors cannot serve real-time needs. The prudent strategy treats real-time capability as infrastructure investment rather than premium product, building the foundation for competitive resilience even while current economics favor batch execution.

The strategic implications are clear: transaction banking's future belongs to institutions that build data network effects, unify product silos into integrated platforms, and invest in real-time infrastructure - treating AI not as a cost-reduction tool but as the foundation of competitive strategy. Achieving this position, however, requires a disciplined implementation approach.

5.6 Industry consolidation

The competitive dynamics described above will drive structural consolidation in transaction banking. Data network effects and platform economics do not produce gradual separation - they create widening gaps that become uncrossable.

Banks that invested early in unified data architectures and AI capabilities will see compounding returns: better models attract more volume, more volume improves models, and improved models justify premium pricing. Banks that delayed, or invested in fragmented point solutions, face the inverse spiral. Their models underperform, clients migrate, and the data needed to compete flows to rivals.

Corporate treasurers will accelerate this divergence. When selecting banking partners, they increasingly evaluate AI capability, integration depth, and settlement speed alongside traditional criteria of balance sheet strength and relationship history. A treasurer managing €500 million in daily flows will consolidate with the bank that optimizes working

Table 10: Technical prerequisites checklist

Prerequisite	Description	Strategic rationale
ISO 20022 Migration	Enriched messaging with structured fields	Provides inputs ML models require
Cloud Elasticity	Dynamic compute scaling for ML workloads	Accommodates variable processing profiles
API-First Architecture	Real-time data access across silos	Enables instant credit decisioning
Unified Customer Data	Single view across cash, trade, SCF	Enables cross-product opportunities
Model Governance Framework	Validation, monitoring, drift detection	Regulatory compliance and risk control
Settlement Layer Integration	APIs to programmable rails and DLT	Instant execution of AI recommendations

capital across cash, trade, and FX, not the bank that offers superior service in one product while requiring manual coordination across others.

The result is predictable: a wave of consolidation as leading institutions acquire mid-tier banks for client portfolios and deposit bases rather than technology or talent. The acquirer gains volume to feed its AI flywheel; the acquired gains access to capabilities it could not build independently. Banks caught in the middle - too small to build, too proud to sell - face margin compression until economics force the decision.

Section 6 examines the pathway from strategic intent to operational reality - the technical prerequisites, governance frameworks, and organizational changes required to convert capability into competitive advantage.

6. Implementation Pathway

Translating strategic vision into operational reality requires systematic execution across technology, products, governance, and organization. The gap between aspiration and execution remains wide. BCG (2024) shows only 25% of financial institutions deploy AI in ways that strengthen competitive positioning. Just 10% of core banking workloads have migrated to cloud [Accenture (2024)].

6.1 Technical prerequisites

Before pursuing product-specific capabilities, institutions must establish foundational infrastructure. The technical prerequisites below represent minimum requirements for AI-enabled transaction banking. Absent any one of these, subsequent investments yield diminished returns.

6.2 Product-specific buildout

With technical foundations in place, attention shifts to product-specific capabilities. Trade finance, supply chain finance, and cross-product intelligence each require distinct implementations, though all draw on the shared infrastructure established above.

6.3 Governance framework

Technology and product capabilities require governance structures that balance innovation velocity

with regulatory compliance and risk management. Model risk management extends established practices into unfamiliar territory. PwC's 2024 Responsible AI Survey found only 11% of executives reported having fully implemented essential responsible AI capabilities including data governance, model testing, and third-party risk management [PwC (2024)]. Models that cannot be explained cannot be defended to supervisors.

Human-in-loop design ensures AI augments rather than replaces judgment. Credit approvals, fraud determinations, and sanctions screening require review thresholds calibrated to risk appetite.

Sandbox participation positions banks to influence emerging frameworks. Trade digitization standards from UNCITRAL and ICC provide legal foundations for electronic document enforceability.

6.4 Security and privacy architecture

AI-enabled transaction banking requires a security architecture that can operate across three dimensions:

- **Data protection**: the knowledge layer aggregates sensitive data across products, creating concentrated risk. Encryption at rest and in transit provides baseline protection. Data residency requirements (GDPR, MAS, PDPA) constrain cross-border processing. Differential privacy techniques enable model training without exposing individual transaction patterns.
- **Model security**: AI models face adversarial attacks through prompt injection, data poisoning, and model extraction. Financial transaction data makes attractive targets. Defense requires input validation, output filtering, anomaly detection on model behavior, and regular adversarial testing. Model versioning enables rapid rollback when attacks succeed.
- **Settlement layer security**: programmable settlement introduces smart contract risk. Code audits, formal verification, and bug bounty programs reduce vulnerability. Multi-signature requirements and time-locks provide fail-safes for high-value transactions. Circuit breakers halt automated settlement when anomalies exceed thresholds.

6.5 Organizational readiness

Technology, products, governance, and security constitute necessary but insufficient conditions for transformation. Technical architecture achieves nothing without organizational capacity. Talent strategies must embed ML engineers within banking product teams. Domain knowledge transfers through co-location, not requirements documents.

Change management should reframe operational staff as model supervisors rather than process executors. Executive sponsorship provides sustained commitment that implementation timelines demand.

Table 11: Product-specific buildout priorities

Domain	Priority	Components	Benchmarks
Trade finance	Document processing	OCR + LLM semantic interpretation	ICC KTDDE: 36 document types (April 2024)
Trade finance	Blockchain infrastructure	DLT for LC issuance	SWIFT ledger: 30+ institutions (Sept 2025)
Trade finance	Electronic documents	Standards-based eBL	11% penetration (versus 1% in 2021)
SCF	Supplier portal	Embedded credit, ERP integration	Primary interface eliminating friction
Cross-product	Relationship mapping	Graph database architecture	Federated learning across institutions

The most challenging requirement: dismantling silos separating cash management, trade finance, and supply chain finance. An integrated platform AI requires integrated structures. Banks overlaying sophisticated technology on fragmented organizations will accelerate existing dysfunction.

Section 7 synthesizes the strategic imperatives for institutions navigating this transformation. The foundational investments, structural decisions, and execution priorities that will separate leaders from laggards.

7. Conclusion: the intelligent network

Transaction banking stands at an inflection point. Key capabilities described in this article (AI-driven intelligence, programmable settlement, and integrated platforms) are not future possibilities. They are production realities at leading institutions. JPMorgan processes U.S.$2 billion daily through tokenized infrastructure [JPMorgan (2025)]. DBS derives measurable value from 1,500 deployed models [DBS (2024)]. The question is no longer whether transformation will occur, but which institutions will lead it.

The convergence matters more than any single technology. AI without instant settlement remains constrained by batch-era execution. Programmable money without intelligence becomes mere infrastructure. Together they enable transaction banking that learns from each interaction, settles in seconds, and optimizes across product boundaries in real time.

For institutions seeking to capitalize on this convergence, three strategic imperatives emerge.

- **Build the data foundation before the AI applications**: unified customer keys, streaming pipelines, and graph databases are not enhancements. They are prerequisites. Banks deploying sophisticated models on fragmented data will underperform competitors with simpler models on integrated data.
- **Treat programmable settlement as core infrastructure**: the GENIUS Act and MiCA have established regulatory clarity. Banks waiting for further proof of concept will find themselves building on infrastructure competitors already operate.
- **Restructure around client outcomes, not product silos**: the full value of AI-enabled transaction banking requires integrated platforms spanning cash, trade, and supply chain finance. Organizational boundaries that made sense in the product-centric era become barriers to value creation in an intelligence-driven model.

Yet candor requires acknowledging what remains unresolved. Despite early promise, most banks have only begun the AI journey. Production deployments at JPMorgan and DBS represent the leading edge, not the industry norm. Most institutions are still exploring applications, piloting use cases, and building foundational data infrastructure.

Significant work remains on both sides of the bank-vendor relationship. Banks must establish governance frameworks that balance innovation velocity with the control requirements regulators expect: model validation, explainability, audit trails, and bias detection. The workforce implications, which we do not explore in depth here, will be substantial. Operations roles will shift from processing to exception management, and institutions must manage this transition thoughtfully.

AI vendors, for their part, must adapt models built for consumer applications to the demands of regulated financial services. Banks require guarantees on data privacy, output accuracy, and compliance auditability that current enterprise AI offerings do not consistently provide. The gap between laboratory capability and production-grade deployment in a regulated environment remains wide.

None of this diminishes the strategic imperative. It contextualizes the execution challenge. The institutions that navigate these constraints, building capability while managing risk, transforming operations while retaining talent, and deploying AI while satisfying supervisors, will define transaction banking's next chapter.

The stakes extend beyond individual institution performance. The industry structure will look fundamentally different by 2030. A small number of global institutions with proprietary AI platforms will compete on intelligence. Regional banks will access shared utilities. Specialists will serve focused niches. Banks in the middle face the most difficult choices.

For those prepared to act, the path forward is clear. The infrastructure exists. The regulatory frame-

works are maturing. What remains is execution: the strategic commitment, organizational change, and sustained investment that convert capability into competitive advantage.

Transaction banking's next phase is data-integrated, machine-intermediated, and autonomously adaptive. The institutions that recognize this soonest will define what comes next.

References

Accenture, 2025, "Banking top 10 trends 2025," Accenture Banking Report

BIS, 2024, "Project mBridge: connecting economies through CBDC," BIS Innovation Hub Report, Bank for International Settlements October

BCG, 2024a, "Global retail banking 2024: the front-to-back transformation," Boston Consulting Group, March

BCG/Apto, 2024b, "Relevance of tokenized assets in asset management," Boston Consulting Group and Aptos Labs Report, October

Capgemini, 2024, "World payments report 2024," Capgemini Research Institute

CoinGecko, 2025, "Stablecoin market capitalisation data," https://tinyurl.com/48ftm6bj

DBS, 2024a, "DBS reports S$750 million economic value from AI initiatives," press release, DBS Group Holdings November

DBS, 2024b, "Annual report 2024," DBS Group Holdings

DBS, 2025, "Annual report 2024," DBS Group Holdings

DeFiLlama, 2025, "Total value locked in DeFi protocols," defillama.com

DCSA, 2025, "eBL adoption data," Digital Container Shipping Association, https://tinyurl.com/vuxwjydc

EP/EC, 2023, "Regulation (EU) 2023/1114 on Markets in Crypto-Assets (MiCA)," Official Journal of the European Union, L 150, June 9

Fnality, 2024, "Fnality payment system: wholesale settlement in central bank money," Fnality International

Fnality, 2025, "Fnality completes first live sterling payment using tokenised central bank reserves," press release, Fnality International, March

FSB, 2024, "Annual progress report on meeting the targets for cross-border payments," Financial Stability Board, October

HSBC, 2023, "HSBC AI markets: turning data into opportunity," Markets Media, HSBC Holdings plc., November

HSBC, 2024, "FX prompt: AI-powered FX risk management," HSBC Global Banking and Markets Product Documentation, HSBC Holdings plc

ICC, 2024, "Key trade documents and data elements (KTDDE) framework," ICC DSI Publication, International Chamber of Commerce, April

ICC, 2025, "ICC digital standards initiative: 2025 progress report," International Chamber of Commerce

JPMorgan, 2024a, "Annual report 2024," JPMorgan Chase

JPMorgan, 2024b, "JPMorgan rebrands Onyx to Kinexys, expands blockchain capabilities," press release, JPMorgan Chase & Co. November

JPMorgan, 2024c, "2024 investor day: commercial and investment bank presentation," JPMorgan Chase & Co., May

JPMorgan, 2025, "'Kinexys processes over $2 Billion daily in tokenized transactions," JPMorgan Insights, JPMorgan Chase & Co., January

McKinsey, 2024a, "Global payments report 2024: modernizing payments to unlock growth," McKinsey Global Institute, October

McKinsey, 2024b, "The state of AI in banking," McKinsey Insights, December

McKinsey, 2024c, "Trade finance: a $6.5 billion digital opportunity," McKinsey Insights, September

MAS, 2023, "Stablecoin Regulatory Framework," MAS Policy Statement, Monetary Authority of Singapore, August

Morrison Finance, 2024, "AI and automation transforming financial services operations," https://tinyurl.com/4mrkw3kp

RLN, 2024, "RLN proof of concept: multi-asset settlement on shared ledger infrastructure," Regulated Liability Network, New York Federal Reserve Innovation Center, November

SAP, 2025, "SAP Taulia: annual savings of EUR 8 million possible," SAP News Center, https://tinyurl.com/44ubhsds

SR Analytics, 2026, "Supply chain risk management: complete guide 2026," https://tinyurl.com/57n9e5nt

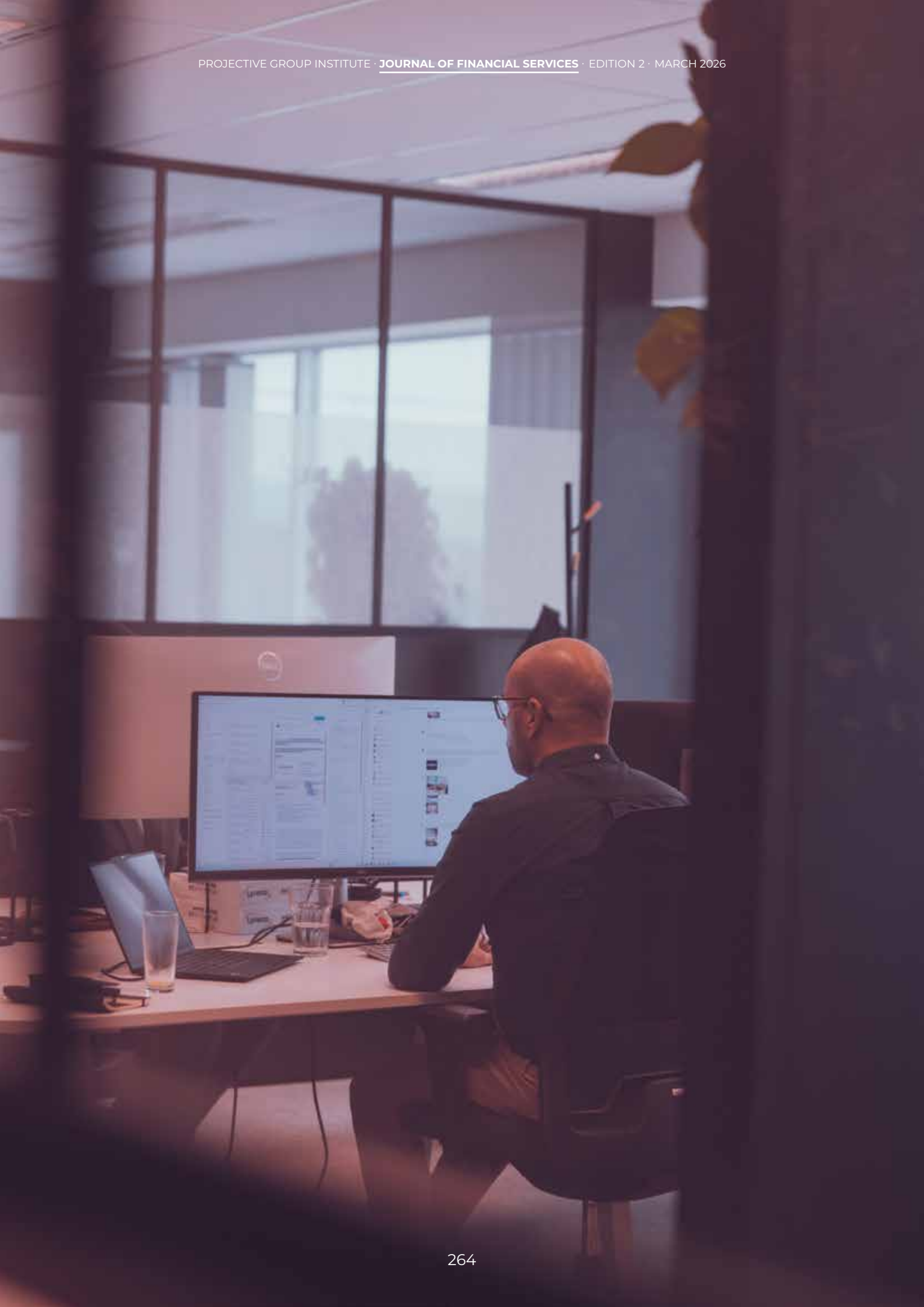
Standard Chartered, 2024, "Trade finance digital transformation report," Standard Chartered Insights

SWIFT, 2025, "SWIFT announces blockchain-based ledger with Consensys," press release, September

U.S. Congress, 2025, "Guiding and establishing national innovation for US Stablecoins Act (GENIUS Act)," Public Law 119-XX, July

UNCITRAL, 2017, "UNCITRAL model law on electronic transferable records," United Nations

zeb Consulting, 2024, "LCR forecasting with AI," BankingHub, https://tinyurl.com/9md6uce2



WE MAKE THINGS HAPPEN.

Leading change in Financial Services.

BELGIUM
UNITED KINGDOM
THE NETHERLANDS
FRANCE
GERMANY
SWITZERLAND

www.projectivegroup.com

Projective INSTITUTE