# Regulating Trust: Responsible AI Pathways for Asian Financial Services

**Ashish Kakar**

Director (Research) – Financial Insights, IDC Asia Pacific; Expert Advisor – Asian Institute of Digital Finance – National University of Singapore

**Arif Perdana**

Associate Professor; Monash University, Indonesia

**Abstract**

The adoption of Artificial Intelligence (AI) in Asian financial services faces a critical "trust deficit," where technological capability outpaces social acceptance. This chapter examines this deficit through a Sociotechnical Systems (STS) lens. We argue that trust is not merely a technical output but a product of alignment between institutional signals and social expectations. Drawing on signalling theory, we analyze regulation not as a barrier, but as a crucial market signal that reduces uncertainty. We posit that the strong correlation between regulatory readiness and AI innovation reflects how clear governance attracts investment rather than suppressing it. Through case studies of Singapore's Veritas framework, the European Union (EU) AI Act, and Indonesia's localization policies, we propose a "modified Brussels Effect" in Asia, where global risk-based architectures are adapted to local ethical norms. Finally, we outline a Responsible AI framework that operationalizes trust through specific guardrails in data governance, model explainability, and human oversight.

## 1. Introduction

Trust will determine the ultimate trajectory and depth of artificial intelligence (AI) adoption within Asian financial services. In an era where algorithms increasingly mediate the relationship between capital and consumers, trust functions as the foundational social contract. Trust is an exchange relationship characterized by interdependence, the presence of risk and vulnerability, and confident expectations about future behavior (Devlin et al., 2015). Within digital finance, this definition materializes through two interrelated contexts, i.e., predictability and repeatability (Chang et al., 2026). Predictability refers to confidence in expected outcomes, while repeatability concerns a system's ability to deliver the same reliable result whenever similar conditions arise. For AI to move beyond experimental silos and into core financial infrastructure, both dimensions must be demonstrable, auditable, and ethically defensible (Vuković et al., 2025).

Consider a practical example from customer contact centers. Suppose a predictive algorithm determines that a customer's annual fees should be waived if the customer spends more than ten thousand dollars annually, has no missed payments, and settles all dues in full. The system is making a value judgment derived from historical data. A banking CEO would reasonably assume that, in every equivalent case, customers who meet these criteria would receive identical treatment when requesting a fee waiver. This assumption depends entirely on trust in the algorithm's stability and its resistance to model drift over time. This represents a relatively simple and successful use case. Many financial institutions have already deployed AI-driven chatbots to handle such routine decisions. However, Asian financial services are now transitioning to far more complex applications (Flint Global, 2025).

These advanced use cases include hyper-personalization engines, automated underwriting systems, AI-driven wealth management, and autonomous portfolio rebalancing. Unlike front-end chatbots, these back-office systems directly affect financial stability, regulatory compliance, and individual livelihoods. Errors or inconsistencies in such systems carry systemic consequences. The question, therefore, is not whether AI can technically enable these applications, but whether it should, and under what conditions. If AI is to support these functions, trust must be designed systemically rather than emerging incidentally from isolated successes.

Trust occupies a uniquely central position in financial services for several reasons (U.S. Department of the Treasury, 2024). Financial services are among the most heavily regulated globally, it continues to rebuild confidence after the 2008 global financial crisis, and it operates fundamentally as a risk management industry. Historically, trust and financial stability have been tightly correlated. Empirical evidence suggests that resilient financial institutions that prioritize transparency and accountability attract more consumers over time (Abdelsalam et al., 2023; Devlin et al., 2015). More recently, Liew et al. (2025) demonstrate that trust is a strong predictor of technological adoption. Their study implies that without a trust-first architecture, even technically advanced AI systems may face resistance or outright rejection.

Financial services customers also exhibit a strong preference for predictability. The consequences of failed predictability were starkly illustrated in 2023, when rapid liquidity shifts and digitally amplified bank runs contributed to the collapse of Silicon Valley Bank, First Republic Bank, and Signature Bank of New York. Credit Suisse's forced merger further underscored how quickly confidence can erode in a hyperconnected financial system. AI systems that introduce opacity or amplify uncertainty risk exacerbate volatility rather than mitigate it. This concern is reinforced by evidence that public acceptance of AI varies significantly across economies and industry segments (Aldasoro et al., 2024; Joshi, 2025).

Survey data highlight this fragmented trust landscape. Edelman (2025) reports substantial variation in AI acceptance across markets. Emerging economies display net positive attitudes toward AI, with acceptance rates of 11% in Brazil and 44% in China. In contrast, developed markets are net rejectors, with acceptance rates of –32% in the United States and –28% in the United Kingdom. Notably, financial services globally record a net AI acceptance rate of only 18%, underscoring a critical scalability barrier. Similarly, KPMG (2025) finds that 54% of respondents remain wary of trusting AI systems. Against this backdrop, Asian consumers appear comparatively more receptive to AI, offering the region a potential "adoption dividend."

Investment patterns reflect this optimism. Globally, financial services firms have invested more than USD 35 billion in AI (World Economic Forum, 2025), with estimates suggesting that approximately one-fifth of this investment is directed toward Asia. The scale of Asia's fintech ecosystem further validates this trend. Dealroom data indicate that USD 122 billion has been invested in Asian fintech, supporting more than 6,000 funded firms (AFCA, 2024). These figures signal a strong institutional belief that AI will underpin future growth.

Fintech's rapid expansion has been driven by dissatisfaction with traditional banks, which were often perceived as slow, costly, and inflexible. Asian fintech activity is concentrated in payments, insurtech, cybersecurity, digital assets, regtech, and wealth management. Yet fintech growth also risks deepening the trust deficit. Despite accelerating digital adoption, consumers continue to place greater trust in traditional banks for long-term financial security. The CRIF Banking on Banks 2025 survey reveals that while 77% of UK consumers expect to manage their finances entirely online by 2030, many remain uncomfortable relying solely on AI-driven or fintech platforms (CRIF, 2025). Trust and protection remain core expectations, particularly amid rising fraud and cyber risks.

Recent studies confirm that fintech adoption is constrained not by functionality but by trust-related concerns. Fintech platforms rely heavily on sensitive personal data and operate across fragmented digital ecosystems, heightening concerns about privacy, security, and accountability (Devlin et al., 2015). Structural assurance and institutional trust continue to shape user experience, while privacy risks dampen post-adoption satisfaction (Yuan et al., 2025). In AI-enabled advisory services, trust is consistently identified as a prerequisite for uptake (Chang et al., 2026). Collectively, these findings suggest that speed and personalization alone are insufficient.

Fintech now accounts for approximately 27% of global data breaches, prompting regulators such as the Monetary Authority of Singapore (MAS) to introduce regulatory sandboxes to manage innovation risks. At the same time, an emerging "experience gap" highlights how technological efficiency often fails to meet human expectations for empathy and reliability (Thoma, 2025). Yet despite these challenges, AI adoption in Asian financial services is no longer optional. Transaction volumes are increasing, digital identity is becoming foundational, and AI capabilities continue to advance rapidly. Institutions that fail to engage meaningfully with AI risk strategic obsolescence.

Building trust in this high-stakes environment requires a deliberate combination of policy and process. Regulatory compliance alone is insufficient. Basel Committee on Banking Supervision (2023) cautions that existing regulatory frameworks often lack the agility needed to prevent systemic failures, despite the presence of regimes such as Basel, Digital Operational Resilience Act (DORA), and Federal Financial Institutions Examination Council (FFIEC). Successfully embedding AI into Asia's financial systems will therefore demand more than static compliance, it will require adaptive governance, continuous oversight, and the rigorous operationalization of Responsible AI principles.

## 2. The Trust Deficit in Financial AI: A Sociotechnical Perspective

Accelerating AI adoption in Asian financial services requires addressing the sector's deepening trust deficit, a gap sustained not only by technical uncertainties but by human reluctance to depend on algorithmic decision-making. Research shows that trust is the decisive precondition for engaging with AI-enabled financial services, shaped by perceptions of AI advisors' credibility and benevolence (Chang et al., 2026). Yet trust is fragile. Even transparent disclosure of AI usage can unintentionally erode legitimacy and reduce trust, as users often view AI involvement as deviating from expected norms of human judgment (Schilke & Reimann, 2025). More broadly, systematic evidence shows that trust in AI is shaped by technical reliability, explainability, perceived risk, and societal concerns, making AI adoption fundamentally a sociotechnical rather than a purely technological challenge (Afroogh et al., 2024). Viewed through Sociotechnical Systems (STS) Theory (Trist & Bamforth, 1951), the trust deficit reflects a structural misalignment between rapid innovation and the social, institutional, and human frameworks required to sustain it; progress, therefore, depends on adjusting both the technology and the surrounding social system.

## 2.1. The Sociotechnical Foundation of Trust

The foundations of STS Theory trace back to Trist & Bamforth's (1951) study of British coal mines, which revealed that technological upgrades alone failed when the social dynamics of work were ignored. Their insight, later expanded by subsequent STS scholars, is that organizational performance depends on the joint optimization of the technical and social systems, rather than privileging one over the other (Molleman & Broekhuis, 2001; Pasmore et al., 2019). Contemporary analyses show that this principle remains critical in the digital era whereby attempts to implement advanced technologies under Industry 4.0 consistently succeed only when human capabilities, culture, and governance are integrated into the design of digital systems (Sony & Naik, 2020).

In digital finance, this means that AI adoption is not merely a question of algorithms or data pipelines. It requires navigating the critical distinction between 'hard governance' (compliance with rules) and 'soft ethics' (alignment with social values), as articulated by Floridi et al. (2018) The social subsystem encompasses customer emotions, cultural expectations around credit and financial security, and regulatory intentions that shape responsible innovation. Trust becomes the key integrative mechanism, a form of "requisite variety" that ensures AI systems adapt to human contexts just as humans adapt to new technological modes of interaction (Pasmore et al., 2019). Consequently, trust functions as a bridge: while hard governance satisfies the regulator, it is soft ethics that satisfies the user. Without addressing this sociotechnical alignment, the trust deficit in AI-enabled finance will persist despite rapid technical advancement.

One of the most widely used approaches for analyzing large-scale technological shifts is the multi-level perspective developed by Geels (2005). This framework explains transitions as the result of interactions across three layers, i.e., the landscape, the regime, and niche spaces where new technologies are experimented with. In the context of Asian finance, AI remains predominantly located within niche environments such as regulatory sandboxes, pilot trials, and specialized innovation labs. These settings support experimentation and help signal regulatory compliance, although they do not automatically overcome the legitimacy concerns that dominate the financial regime (Kindermann et al., 2025).

As financial institutions begin integrating AI into mainstream operations, a structural bottleneck has emerged. The speed of algorithmic innovation within niches does not align with the trust requirements embedded in the financial regime. Research on algorithm aversion and algorithmic transparency consistently shows that users hesitate to rely on opaque AI systems, especially in high-stakes settings where competence, fairness, and explainability are essential for trust (Jussupow et al., 2024; Ning et al., 2024). Existing regulations, risk management conventions, and long-standing privacy expectations heighten concerns about the opacity of AI decision-making.

Reaching a stable new normal requires alignment between niche innovations and regime structures. Recent studies show that trust improves when AI systems provide clearer transparency signals, demonstrate consistent competence, and adhere more closely to regulatory expectations

(Chang et al., 2026; Ning et al., 2024). In the language of the multi-level perspective, a complete transition occurs only when niche innovations adjust to these socio-technical pressures and when institutional actors, regulators, and users converge toward shared standards for trustworthy AI in finance.

## 2.2 Deconstructing the Five Pillars of the Trust Deficit

The trust deficit surrounding the adoption of AI in financial services can be understood through five interrelated pillars that shape both technical performance and social acceptance. These pillars are data, models, technology, governance, and regulation (see Figure 1). Each contributes to trust formation, and failure in any one area can undermine the credibility of the entire system.

Data forms the foundation of all algorithmic decision-making. It includes transactional records, text, voice, and biometric information that fuel predictive and generative models. In many Asian financial contexts, trust erodes when models rely on incomplete data or when concerns arise around unauthorized access and data accuracy. The use of alternative data for credit scoring without clear consent frameworks often appears opaque and coercive to consumers, creating perceptions of exploitation rather than empowerment. When individuals do not understand how their data is collected, combined, or repurposed, confidence in AI-enabled decisions weakens.

The pillars of models and technology relate to how data are transformed into decisions. Financial institutions increasingly deploy techniques ranging from neural networks to generative AI architectures, such as retrieval-augmented generation and large language models. While these systems can achieve high predictive accuracy, they often sacrifice explainability. This trade-off directly affects trust, particularly in high-stakes decisions such as lending or fraud detection, where users expect consistent and interpretable outcomes. Technical risks further complicate adoption. Hallucinations and the limited deployment of privacy-enhancing techniques undermine the predictability and repeatability that financial services depend upon for operational stability.

Governance and regulation provide the structural safeguards that shape responsible use of AI. Governance refers to internal policies, oversight mechanisms, and accountability structures that guide how humans and AI systems interact. Many organizations continue to lack clear guardrails for human AI collaboration, leaving responsibility diffuse when errors occur. Regulatory environments across Asia remain uneven, with differing national approaches to AI oversight. In the absence of harmonized audit standards or certified compliance mechanisms, even responsible firms struggle to demonstrate trustworthiness to regulators and the public. Closing the trust gap, therefore, requires systematic attention to risks ranging from privacy breaches to algorithmic instability, ensuring that AI adoption delivers social legitimacy alongside technical performance.

Concerns around data access illustrate why governance and regulation matter. Policy guidance from the Bank for International Settlements highlights that the use of financial data poses significant privacy and security risks if poorly managed (Cristano et al., 2024). The Cambridge Analytica scandal demonstrated how large-scale data misuse can erode public trust. Personal data from millions of users was harvested without consent and used for behavioral targeting, leading to substantial regulatory penalties and new accountability requirements (Hinds et al., 2020; McCallum, 2022; The Guardian, 2018). In response, regulators strengthened principles around fairness, purpose limitation, transparency, data accuracy, access rights, and security.

These incidents continue to shape perceptions of data misuse. Surveys show persistent consumer distrust toward financial institutions' handling of personal data (Edelman, 2025; KPMG, 2025). Distrust also exists within institutions themselves, where managers question data quality and reliability due to fragmented systems and legacy infrastructure (Gunasekar et al., 2023; Somu, 2022). In practice, these weaknesses surface as poor customer experiences. Financial leaders report challenges such as stale data, device-specific profiles, and the inability to integrate relationship-level information. One example involved a high-net-worth client receiving an inappropriate loan offer triggered by a single missed payment, reflecting an algorithm trained on narrow data rather than holistic context.

Algorithmic bias compounds these challenges. Bias emerges from data imbalance, feature selection, and model design. In fraud detection, rare fraudulent events must be identified within vast volumes of legitimate transactions, requiring careful mitigation of skewed data distributions (Bello et al., 2023; Pombal et al., 2022). Beyond business risk, bias has broader social consequences by reinforcing inequality and discriminatory outcomes (Ferrara, 2023). Such outcomes deepen the sociotechnical trust deficit and highlight the need for responsible AI practices.

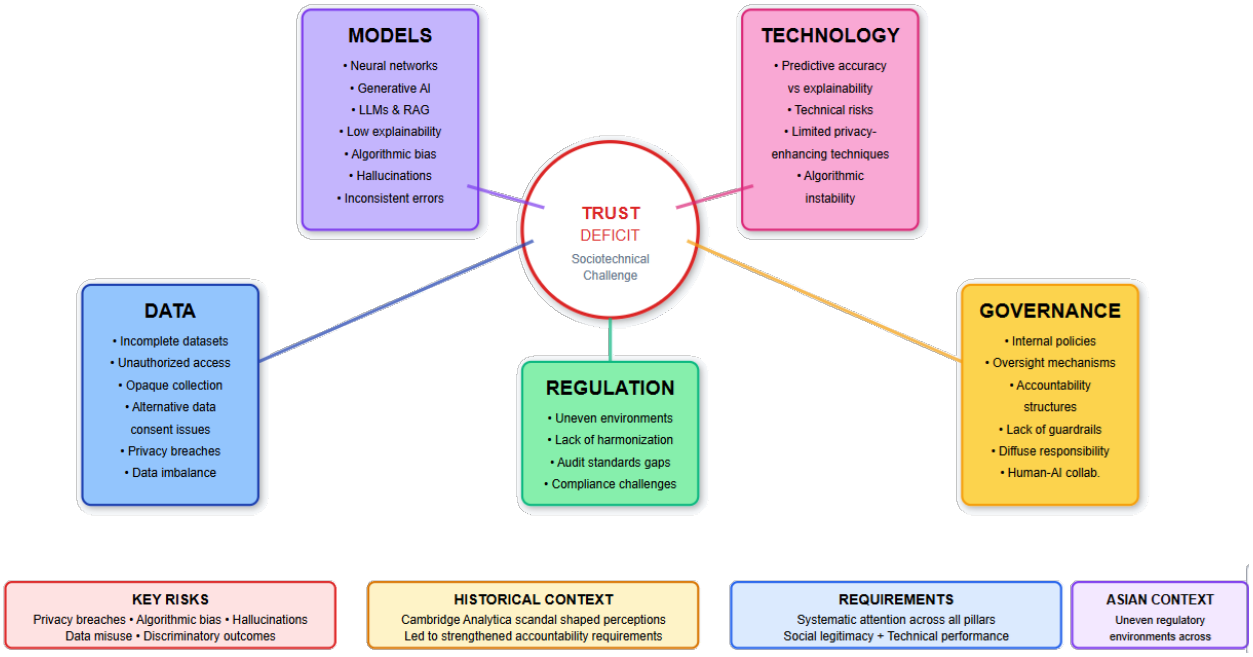

**Figure 1. Five Pillars of Trust Deficit in AI-Enabled Financial Services**

Hallucination presents a distinct but equally damaging risk. Unlike systematic bias, hallucinations produce inconsistent, unexplained errors. This unpredictability threatens compliance, accuracy, and confidence in AI-driven decisions (Huang, 2023; Sun et al., 2024; Zhou & Zafarani, 2020). While scholars note that hallucinations cannot be fully eliminated without sacrificing performance (Lee, 2023), responsible AI frameworks can mitigate their impact by aligning use cases with probabilistic processes in which AI has demonstrated value, such as scam detection. Together, these dynamics show that trust in financial AI is not a technical problem alone but a sociotechnical challenge that demands coordinated attention across data governance, model design, institutional oversight, and regulatory alignment.

## 3. The Need for An AI Regulation to Signal Trust

The challenges discussed thus far fall broadly into two categories. Some are technical in nature, such as hallucination, while others are procedural, including data access limitations and algorithmic bias. Together, these challenges raise a central question for policymakers and practitioners alike, namely, whether regulation can meaningfully contribute to building trust in AI. Empirical evidence suggests that regulation matters, but not in isolation. Robust digital infrastructure, sustained investment in research and development, and organizational readiness significantly support the adoption of generative AI. At the same time, poorly aligned government policies can constrain these efforts rather than enable them (Ali et al., 2025). Aggregate data reinforces this observation.

One of the most comprehensive measures of national AI readiness is the AI Preparedness Index developed by the International Monetary Fund (International Monetary Fund, 2025). The index combines four dimensions that together capture a country's ability to scale and monetize AI, namely digital infrastructure readiness, innovation and economic integration, human capital and labour market practices, and regulation and ethics. While it is unsurprising that regulation correlates strongly with overall AI preparedness, what stands out is the degree of interdependence between regulation and the other dimensions. Rather than operating as an external constraint, regulatory and ethical capacity appears closely intertwined with technological and economic development.

Table 1 illustrates this institutional interdependence among the core dimensions of national AI preparedness. Based on data from 165 countries, the table reports correlations between overall AI preparedness and its four constituent dimensions. The consistently strong associations between regulation and ethics and digital infrastructure, innovation capacity, and human capital challenge the conventional assumption that regulation primarily suppresses investment and experimentation. Instead, the pattern suggests institutional complementarity. Countries that invest in regulatory clarity and ethical oversight also tend to build stronger infrastructure, foster innovation ecosystems, and develop the skills required to support AI adoption at scale. In this configuration, regulation is not a downstream response to technological change but rather an enabling institutional layer that reduces uncertainty and supports coordinated capability-building.

**Table 1. Multicollinearity (AIPI, 2025), n =165**

| | **AIPI** | **Digital Infrastructure** | **Innovation and Economic Integration** | **Human Capital and Labor Market Policies** | **Regulation and Ethics** |
|---|---|---|---|---|---|
| **AIPI** | 1 | | | | |
| **Digital Infrastructure** | 0.966 | 1 | | | |
| **Innovation and Economic Integration** | 0.863 | 0.784 | 1 | | |

| | AIPI | Digital Infrastructure | Innovation and Economic Integration | Human Capital and Labor Market Policies | Regulation and Ethics |
|---|---|---|---|---|---|
| **Human Capital and Labor Market Policies** | 0.917 | 0.891 | 0.697 | 1 | |
| **Regulation and Ethics** | 0.947 | 0.881 | 0.759 | 0.821 | 1 |

At first glance, this finding appears counterintuitive. A substantial body of earlier research argues that regulation dampens investment by increasing compliance costs and limiting flexibility. Evidence from telecommunications shows that easing entry regulation was associated with higher investment by new entrants (Friederiszick et al., 2008). Similar patterns have been observed in environmental, social, and governance contexts, where investors often favor jurisdictions with fewer trade restrictions and lighter regulatory burdens (Cheng et al., 2018; Escribá-Pérez & Murgui-García, 2016). Against this backdrop, the positive association between regulation and AI readiness requires a different explanatory lens.

Signalling theory offers such an explanation. Recent studies show that regulatory frameworks can function as signals that reduce uncertainty and foster trust in emerging technologies (Fatima et al., 2021; Gonzalez, 2025; Nishant et al., 2023). In environments characterized by imperfect information, firms and investors rely on observable institutional cues to guide strategic decisions (Bergh et al., 2014). Clear regulatory intent signals legitimacy, long-term commitment, and the likelihood of stable enforcement, thereby encouraging investment. For consumers, regulation also provides reassurance that risks are recognized and monitored, reinforcing confidence in AI-enabled services.

This logic aligns closely with the sociotechnical perspective adopted in this chapter. Government engagement in AI regulation is widely interpreted as a positive signal that strengthens trust and supports adoption (Fatima et al., 2021). However, signalling only works when policies are well designed. Evidence from a cross-country study of 136 economies shows that technology promotion policies can have unintended adverse effects when poorly aligned with the AI sector's needs (Ali et al., 2025). Overly rigid or ambiguous rules may increase friction, discouraging experimentation and slowing diffusion.

The absence of shared definitions and binding frameworks further complicates policy design. According to an OECD survey, only slightly more than half of surveyed markets have binding AI regulations that include a formal definition. In contrast, others rely on non-binding guidance or lack AI-specific policy altogether (OECD, 2024). Without conceptual clarity, regulation struggles to address the distinct risks that AI introduces or amplifies. While existing safety and soundness standards remain relevant, treating AI risks as merely extensions of traditional financial risks oversimplify the challenge.

Guidance from the World Bank highlights AI-specific risks in financial services, including threats to financial stability, pricing and product integrity, operational resilience, reputational exposure, and market concentration (Cristano et al., 2024). These risks intersect with earlier technical and procedural challenges and underscore the need for balanced policy approaches. Well-designed AI regulation can signal trust, enhance transparency, and strengthen institutional confidence when it addresses both technical specifications and governance processes. In this sense, regulation functions not as a constraint but as a sociotechnical instrument for rebuilding trust across financial ecosystems.

## 4. Learning From AI Regulations: Will the Brussels Effect Apply?

Regulations are an essential policy tool that helps to build trust in AI. The question is whether Asia has a proactive, positive policy framework, existing or emerging. Globally, there are three major policy approaches that depend on the level of the central bank's intervention in financial services AI policies. Some nations require AI models, especially high-risk and customer-facing models, to be pre-approved; others use a risk-based approach; and some countries do not specify an AI policy. Asia seems to be closely aligned with the EU approach. It is essential for Asian countries to adopt a specific AI policy, as some lack one at present.

Statistically, 12% of the countries in the OECD (2024) survey did not have an AI framework, and another 26% reported a non-binding AI policy. The same applies to Asia, with China, Singapore, and India among the markets with well-developed AI policies and guidelines, and Thailand, the Philippines, and Vietnam among those where AI policies are still evolving. This heterogeneity creates varying levels of protection and trust across the region.

Would these existing and emerging regulations signal trust? And are these the same or similar? Also, is there a Brussels effect? The Brussels effect refers to the EU being a pioneer in regulations, which are then adopted by the rest of the world (Bradford, 2020; Siegmann & Anderljung, 2022). Answers to these questions are important and require a better understanding of these regulations. Whether a similar dynamic is unfolding in AI governance depends on the degree to which Asian regulators converge toward European norms.

The EU AI Act specifies that guardrails must be established to ensure the responsible and fair implementation of AI in European markets[1]. The act specifies a risk-based approach, prohibits certain AI activities, imposes limitations on high-risk AI implementations, requires transparency around large language models, and mandates human oversight. Each transaction is categorized as limited, medium, or high risk. The classification is based on potential impact on users and society. Systems that manipulate human behavior, exploit vulnerabilities, create social scoring, lead to biometric classification, and predictive policing are banned. High-risk transactions include AI used in medical devices and credit, emotion recognition, HR screening for recruitment, evaluating educational learning outcomes, and eligibility for benefits and services. Large language model providers will also be subject to increased transparency, documentation requirements, and registration.

Singapore is often considered a model of governance in Asia. In that context, Monetary Authority of Singapore (2025) has just issued a consultation paper on AI risk management in financial services. This consultation must be viewed in conjunction with the existing practices, e.g., FEAT and Veritas (Monetary Authority of Singapore, 2022). FEAT stands for Fairness, Ethics, Accountability, and Transparency, while Veritas is the framework for implementing FEAT. These instruments address many of the risks identified earlier, including algorithmic bias and data consent. Other authorities, such as the Infocomm Media Development Authority (IMDA), have also issued sector-agnostic AI guidelines that encourage public-private partnerships[2].

[1] The Artificial Intelligence Act (AI Act) is the European Union's first comprehensive legal framework regulating artificial intelligence. It establishes harmonized, risk-based rules for the development, deployment, and use of AI systems in the EU, aiming to ensure safety, fundamental rights protection, and trustworthy AI across member states. https://artificialintelligenceact.eu/

[2] The Infocomm Media Development Authority (IMDA) is Singapore's government agency responsible for driving digital transformation, data protection, and AI governance. It develops national AI frameworks such as the Model AI Governance Framework and AI Verify, and oversees

Similar to Singapore's financial services regulator, the Indonesian financial services regulator, Otoritas Jasa Keuangan (OJK) (2025), has also issued AI guidelines[3]. OJK (2025) has also stressed the need for responsible AI, data protection, governance requirements, and international interoperability. A key component of OJK's (2025) guidelines is its alignment with Pancasila Principles. Pancasila is the official state ideology, and the alignment ensures that AI can be implemented only when it is in the national interest.

Principles of consumer protection, welfare, transparency, and ethics are also key to the policy. These policy examples exemplify regulators' efforts to build trust in AI and suggest that a modified 'Brussels Effect' is at play, one in which Asian regulators adopt the EU's risk-based architecture but localize ethical values (e.g., Pancasila) to fit regional social contracts. The question then remains: why do consumers and leadership still lack trust? There seem to be other drivers, beyond policy, driving the deficit.

The exemplary data architecture is critical for large-scale AI implementations. Processes alone do not mitigate all challenges, and some remain unaddressed, including data stored across different systems, delays in data processing, and data sovereignty laws that favor on-prem infrastructure over shared cloud computing. Financial services leadership is concerned about these, especially the possibility that the data repository may be partial and have low data integrity. Streamlining data architecture requires regulatory intervention and a policy stance. In the context of AI, third-party providers are both a boon and a bane. Financial services management needs to balance ease of implementation with model opacity and concentration risk arising from dependence on a limited number of suppliers.

The third-party risk would also need regulatory and policy intervention to enhance market trust. Currently, while financial services are regulated and under stringent systems resilience requirements, third-party vendors are largely not affected in case of resiliency or data leak failures. This difference in regulatory oversight and surveillance is affecting public trust in AI. Banking leadership is hesitant to trust third-party providers, thus limiting AI implementation potential. Overall, AI regulation functions as a sociotechnical mechanism that must align institutions, technologies, and human judgment to sustain trust. Next, we cover responsible AI frameworks that would help financial services build trust in AI.

## 5. Framework for Responsible AI in Asian Digital Finance

Besides regulations and technology, structured models are fundamental to building trust. From a sociotechnical perspective, these models function as the 'technical' bridge to the 'social' requirement of legitimacy. These are actionable models that detail guardrails designed to integrate trust drivers. This defines the space for responsible AI, which emerges from the definition of AI itself. AI in European terms is defined as human-centric, sustainable, inclusive, and trustworthy. Similarly, the US Government Accountability Office has defined its own AI framework (GAO, 2023). This framework stresses the four principles of governance, data, performance, and monitoring. Governance establishes processes that promote accountability; data establishes the quality and reliability of model inputs; performance stresses ensuring that the outputs are consistent with program objectives; and monitoring ensures that the results are future-proofed and relevant to any business or ecosystem changes. Given these definitions, implementing a

programmes that help businesses adopt digital and AI technologies responsibly. https://www.imda.gov.sg/about-imda/emerging-technologies-and-research/artificial-intelligence

[3] Otoritas Jasa Keuangan is a Financial Service and Authorities in Indonesia

responsible AI framework requires defining processes across data, models, human oversight, and the ecosystem.

Data governance is defined as a company-wide framework for assigning decision-making rights and duties to adequately handle data as a company asset (Alhassan et al., 2016). This involves processes related to data collection, data availability, data use, data storage & related security, and, finally, data destruction. Furthermore, regulated industries, such as financial services, may have additional localisation challenges.

The collection of data requires consent frameworks and adherence to national data privacy laws. Crucially, some regulations, like GDPR, also require compliance if part of the business is EU-driven. The basic tenets for data guidance were published by OECD in 2001, and subsequently updated in 2013 (OECD, 2013). The principles include the collection limitation principle, which states that data must be collected fairly and legally with the provider's full consent, and the data quality principle, which requires that data be fit for purpose, accurate, and kept up to date. The purpose specification principle and the use limitation principle are two other critical principles that limit data collection to the specified purpose and prohibit sharing the data for incompatible uses. Further principles stress the need to protect data, uphold openness, and uphold the individual ownership principle, under which individuals have a right to review their current data and request its deletion. Finally, the principles stress the data controller's role in building accountability. Country regulations are often derived from these principles.

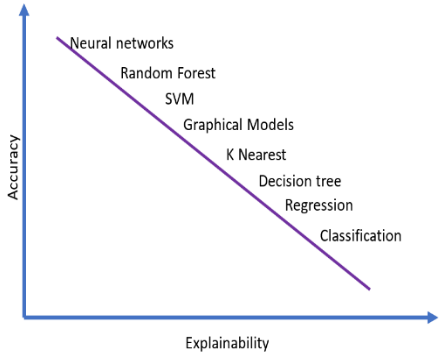

**Figure 2. Accuracy vs Explainability in AI and Machine Learning**

Data availability and use require a discussion on data infrastructure and localisation. Financial services data tends to be fragmented across legacy and modern systems, and across infrastructure, including multiple data centres. Therefore, data availability, an important component of governance, requires intelligent orchestration rather than just centralised management. While data centralisation led many financial services companies to initiate data lakes and cloud computing projects in the previous decade, current architectures are shifting toward Data Mesh and federated governance. The aim was originally to collocate analytics data in a data lake and build processing capabilities using cloud computing. However, in Asia, cloud computing could be regulated, and those regulations would need to be considered within the data governance framework. Consequently, Privacy-Enhancing Technologies (PETs)—such as homomorphic encryption and zero-knowledge proofs—are becoming essential. These technologies allow institutions to extract insights from data without moving the underlying assets across borders, thereby satisfying strict localization and sovereignty laws while maintaining operational efficiency.

Available, high-quality data is required to help build AI models. Yet, AI models can exacerbate the trust deficit. The recent pandemic has shown that financial services models, when stressed, can yield incorrect analysis; that risk will magnify as the preponderance of new models built to monetise AI proliferates. Consequently, model governance must include bias audits and regular monitoring to ensure that the results are consistent with the objectives. The models must also ensure compliance with explainable AI requirements. Explainable AI is a requirement under most central regulatory acts, such as the EU AI Act. Explainable AI is a method that helps humans understand and trust AI model outputs. However, this presents a challenge. Historical methods analysis has shown that as explainability increases, model accuracy drops. The relationship can be best explained in Figure 2. However, it is important to note that this trade-off is no longer absolute. Emerging techniques in 'Interpretability by Design' and post-hoc explainability are beginning to bend this curve, allowing high-performance models to function within regulated environments without remaining total 'black boxes.'

The model and data challenge is best addressed through a human oversight layer, and indeed, most regulators are now requiring it for high-risk processes. The central principle holds that a human is the final decision-maker, and that is where accountability stops. While it can be argued that human oversight may be unfathomable for all processes, suggesting that machines may need to be autonomous with some level of decision-making, i.e., the agentic AI principles, that argument is currently being challenged because of low AI trust.

Human oversight is particularly critical given the risk of hallucination (Ji et al., 2023; Zhou & Zafarani, 2020). Specifically, hallucination has posed challenges for compliance, accuracy, and authenticity. Notably, Lee (2023) argues that hallucination is an inherent trait of AI and that eradicating hallucination without achieving quality and cost performance is nearly impossible. Therefore, humans in the loop must be incorporated into the AI governance model. Overall, the governance model must include aspects such as fast redressal processes, vendor management, third-party information security audits, and AI model audits. Additionally, cross-border supervisory agreements and changes in vendor master service agreements may also be required, with vendors needing to document the models being used and their risk mitigation plans.

## 6. Case Studies: Trust in Action

While reducing the AI trust deficit may seem stringent, there are numerous case studies where industry and regulators have successfully addressed key AI implementation concerns. Three of these, which are particularly worth highlighting, include the proactive regulatory measures adopted by Singapore, the stress on human-AI interface in the EU Act, and Indonesia's localization initiatives.

Singapore represents a regulatory signalling-driven case study in which regulators are leading efforts to build AI technology and implement it in financial services. To achieve that objective, the MAS defined the Veritas framework as guidelines for financial services to implement FEAT (Fairness, Ethics, Accountability, and Transparency) AI principles (Monetary Authority of Singapore, 2022). While much still needs to be implemented, Singapore's success is validated by the recent AI preparedness index (International Monetary Fund, 2025), which ranks Singapore among the countries with the highest level of preparation.

The Veritas framework includes case studies to guide transparent AI implementation. For example, one of the case studies within the framework is credit decisioning, where the Veritas framework specifies a checklist to help financial services prepare for adoption and implementation (Monetary Authority of Singapore, 2022). Key elements of this checklist include a determination of whether transparency is required for external stakeholders, the proactive and reactive communications that must be sent to support transparency efforts, the features and factors required to create internal transparency, factors that will drive explainability, and, finally, the acceptable accuracy rate. Implementing this checklist would significantly help reduce the trust deficit.

Like MAS, the European Union has been progressive in signalling through regulations. Specifically, Article 14 of the EU AI Act 2023 discusses human oversight for high-risk processes. Many of the financial services processes that lead to a trust deficit can be categorized as high-risk. The specific mandate under the law is that high-risk AI systems must build an appropriate human-machine interface tool that not only ensures human oversight but also protects fundamental rights. Human oversight, as detailed earlier, is thus an effective trust-building mechanism.

Our last case focuses on localization and the emerging paradigm of 'Sovereign AI'. With the United Nations estimating that there are over 7,000 languages worldwide, it is imprudent to assume that English-language large language models (LLMs), while they may be dominant, will be enough to build AI trust. Financial services organizations and their customers may require a local-language model that incorporates local context as a prerequisite for building AI trust. This goes beyond mere translation; it is about ensuring that the foundational models reflect local cultural norms, such as Pancasila in Indonesia, rather than importing Western biases that may not align with local social contracts.

Indonesia exemplifies this push for Sovereign AI. This involves a strategic collaboration between local language developers building Indonesian Bahasa-based models and multinational hyperscalers who understand the importance of local language nuance. Notably, some of these LLMs are now active in 11 regional languages. By securing 'digital sovereignty' through these localized models, regulators and firms can assure the public that their financial AI is not just technically competent, but culturally compatible. Localization is, therefore, one of the vital strategies that can help build trust.

## 7. Conclusion

While significant investments have been made in AI technology in financial services, adoption lags due to a range of factors, including technological limitations and a trust deficit. Between the two, there is an acceptance that technological limitations, such as hallucinations, are difficult to eliminate. As a result, AI adoption in financial services will depend on whether organizations can build trust in their solutions. The question, therefore, is: how can that trust be built?

It is certain that AI will be a transformative change, much like past experiences of such changes in history. Examples include the migration from horse-driven carts to motor vehicles. Such transformations have been successful because they integrated social acceptance with technological advancements. The same will need to happen with AI. Trust in AI will need to be earned and accelerated with an acceptance that trust requires sociotechnical signals. Both financial services organizations and regulators need to signal consistent policies. However, that requires accepting two facts. First, AI is not just a technology adoption but also has wider social implications. Secondly, contrary to skeptical viewpoints, regulations can play a positive role in AI adoption by signalling trustworthiness.

Consequently, a human-centric approach may be required to ensure social acceptance, and messaging needs to reinforce human enablement rather than human replacement. Human enablement needs to be built into a responsible AI framework. These frameworks need to be cognizant of privacy needs and the fact that AI requires oversight, with humans retaining final decision accountability for high-risk processes. Furthermore, guardrails need to be put in place to ensure the ethical and fair deployment of AI without deliberate bias. Crucially, algorithms must be explainable. A lack of explainability signals deep mistrust from both customers and management. Financial services were stress tested twice recently: once during the Great Recession and, secondly, during the recent pandemic. Stress situations require risk estimation, and it may not bode well for shareholders and management if these risk estimates cannot be explained.

This underscores the critical role of AI governance. AI governance defines the boundaries within which use cases are selected and implemented. These boundaries also prevent reputation losses by minimizing bias and ensuring that data is used fairly. The governance framework should include an audit requirement and specify the acceptable level of algorithmic accuracy. Effective AI policy could significantly reduce the trust deficit. Additionally, acceptable accuracy levels would help organizations plan for the human-machine interface. Since hallucination cannot be eliminated, an organization can thereby assess the extent of human oversight required through its processes.

Finally, the role of regulations cannot be overstated. Let us return to our example of horse-drawn carriages being replaced by motor vehicles. That would have been impossible if driving were unregulated. Fear of injury, loss of property, or death would have stopped the adoption. Similarly, people need to trust that AI will be used for good, not for unethical profiling or the misuse of privacy. On the positive side, there are significant examples of how trust can be built. Three of those were covered in this chapter, of which two are related to regulations. The learning from these success cases is that trust can indeed be built. Organizations must now implement some of these policies to ensure their AI spend delivers a positive ROI.

## References

Abdelsalam, O., Chantziaras, A., Joseph, N., & Tsileponis, N. (2023). *Trust matters: A global perspective on the influence of societal trust on bank market risk*. https://doi.org/10.2139/ssrn.4600115

AFCA. (2024). *Home page*. http://www.afca-asia.org/Portal.do?method=indexView

Afroogh, S., Akbari, A., Malone, E., Kargar, M., & Alambeigi, H. (2024). Trust in AI: Progress, challenges, and future directions. *Humanities and Social Sciences Communications*, *11*(1), 1568. https://doi.org/10.1057/s41599-024-04044-8

Aldasoro, I., Gambacorta, L., Korinek, A., Shreeti, V., & Stein, M. (2024). *Intelligent financial system: How AI is transforming finance* (No. BIS Working Paper No. 1194). Bank for International Settlements. https://www.bis.org/publ/work1194.pdf

Alhassan, I., Sammon, D., & Daly, M. (2016). Data Governance Activities: An analysis of the literature. *Journal of Decision Systems*, *25*(sup1), 64–75. https://doi.org/10.1080/12460125.2016.1187397

Ali, H., Mustafa, A. ul, & Aysan, A. F. (2025). Global adoption of generative AI: What matters most? *Journal of Economy and Technology*, *3*, 166–176. https://doi.org/10.1016/j.ject.2024.10.002

Basel Committee on Banking Supervision. (2023). *Report on the 2023 banking turmoil*. Bank for International Settlements. https://www.bis.org/bcbs/publ/d555.pdf

Bello, O. A., Ogundipe, A., Mohammed, D., Folorunso, A., & Alonge, O. A. (2023). *AI-driven approaches for real-time fraud detection in U.S. financial transactions: Challenges and opportunities*. https://www.researchgate.net/publication/381548442_AI-Driven_Approaches_for_Real-Time_Fraud_Detection_in_US_Financial_Transactions_Challenges_and_Opportunities

Bergh, D. D., Connelly, B. L., Ketchen, D. J., & Shannon, L. M. (2014). Signaling theory and equilibrium in strategic management research: An assessment and a research agenda. *Journal of Management Studies*, *51*(8), 1334–1360. https://doi.org/10.1111/joms.12097

Bradford, A. (2020). *The Brussels effect: How the European Union rules the world*. Oxford University Press. https://doi.org/10.1093/oso/9780190088583.003.0003

Chang, C.-Y., Park, S., & Dinh, M. C. (2026). Advice by algorithms: Dual dimensions of trust in AI financial service adoption. *Journal of Retailing and Consumer Services*, *88*, 104537. https://doi.org/10.1016/j.jretconser.2025.104537

Cheng, Z., Li, L., & Liu, J. (2018). The spatial correlation and interaction between environmental regulation and foreign direct investment. *Journal of Regulatory Economics*, *54*(2), 124–146. https://doi.org/10.1007/s11149-018-9366-x

CRIF. (2025). *Banking on banks 2025: Europe's financial services on the road to 2030*. https://crif.co.uk/media/01wjmjmn/crif-banking-on-banks-report-2-2025.pdf

Cristano, J. C., Leuterio, C. B., Prenio, J., & Yong, J. (2024). *FSI insights on policy implementation no. 63: Regulating AI in the financial sector*. Bank for International Settlements. https://www.bis.org/fsi/publ/insights63.pdf

Devlin, J. F., Ennew, C. T., Sekhon, H. S., & Roy, S. K. (2015). Trust in financial services: Retrospect and prospect. *Journal of Financial Services Marketing*, *20*(4), 234–245. https://doi.org/10.1057/fsm.2015.21

Edelman. (2025). *Flash poll: Trust and artificial intelligence at a crossroads*. https://www.edelman.com/trust/2025/trust-barometer/flash-poll-trust-artifical-intelligence

Escribá-Pérez, J., & Murgui-García, M. J. (2016). Do market regulations reduce investment? Evidence from European regions. *Regional Studies*, *51*(9), 1336–1347. https://doi.org/10.1080/00343404.2016.1182147

Fatima, S., Desouza, K. C., Denford, J. S., & Dawson, G. S. (2021). What explains governments' interest in artificial intelligence? A signaling theory approach. *Economic Analysis and Policy*, *71*, 238–254. https://doi.org/10.1016/j.eap.2021.05.001

Ferrara, E. (2023). Fairness and bias in artificial intelligence: A brief survey. *Sci*, *6*(1), 3. https://doi.org/10.3390/sci6010003

Flint Global. (2025). *Accelerating AI adoption in financial services in Asia-Pacific*. https://flint-global.com/news/accelerating-ai-adoption-in-financial-services-in-asia-pacific/

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Friederiszick, H., Grajek, M., & Röller, L.-H. (2008). *Analyzing the relationship between regulation and investment in the telecom sector*. https://www.researchgate.net/publication/228353806_Analyzing_the_Relationship_between_Regulation_and_Investment_in_the_Telecom_Sector

GAO. (2023). *Artificial Intelligence*. https://www.gao.gov/artificial-intelligence

Geels, F. W. (2005). Processes and patterns in transitions and system innovations. *Technological Forecasting and Social Change*, *72*(6), 681–696. https://doi.org/10.1016/j.techfore.2004.08.014

Gonzalez, A. (2025). *Does AI regulation crowd in responsible investment? Evidence from 27 countries*. https://doi.org/10.2139/ssrn.5264355

Gunasekar, S., Zhang, Y., & Aneja, J. (2023). *Textbooks are all you need*. https://doi.org/10.48550/arXiv.2306.11644

Hinds, J., Williams, E. J., & Joinson, A. N. (2020). "It wouldn't happen to me": Privacy concerns after Cambridge Analytica. *International Journal of Human-Computer Studies*, *143*, 102498. https://doi.org/10.1016/j.ijhcs.2020.102498

Huang, H. (2023). Performance of ChatGPT on RN licensing exam in Taiwan. *Healthcare*, *11*(21), 2855. https://doi.org/10.3390/healthcare11212855

International Monetary Fund. (2025). *AI Preparedness Index*. https://www.imf.org/external/datamapper/datasets/AIPI

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, *55*(12), 1–38. https://doi.org/10.1145/3571730

Joshi, S. (2025). *Compensating for the risks and weaknesses of AI/ML models in finance*. https://doi.org/10.2139/ssrn.5206475

Jussupow, E., Benbasat, I., & Heinzl, A. (2024). An integrative perspective on algorithm aversion and appreciation. *MIS Quarterly*, *48*(4), 1575–1590. https://doi.org/10.25300/MISQ/2024/18512

Kindermann, B., Kraljev, M., & Flatten, T. (2025). FinTechs playing in the regulatory sandbox: The effect of interacting signals on funding. *Information Systems Journal*. https://doi.org/10.1111/isj.70004

KPMG. (2025). *Trust, attitudes and use of artificial intelligence: A global study 2025*. https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2025/05/trust-attitudes-and-use-of-ai-global-report.pdf

Lee, M. (2023). A mathematical investigation of hallucination and creativity in GPT models. *Mathematics*, *11*(10), 2320. https://doi.org/10.3390/math11102320

Liew, T. W., Lim, C. T., Khan, M. T., & Tan, S.-M. (2025). Banking on voice: AI attributes and trust in voicebot acceptance. *Computers in Human Behavior Reports*, *20*, 100812. https://doi.org/10.1016/j.chbr.2025.100812

McCallum, S. (2022). *Meta settles Cambridge Analytica scandal case for $725M*. https://www.bbc.com/news/technology-64075067

Molleman, E., & Broekhuis, M. (2001). Sociotechnical systems: Towards an organizational learning approach. *Journal of Engineering and Technology Management*, *18*(3–4), 271–294. https://doi.org/10.1016/S0923-4748(01)00038-8

Monetary Authority of Singapore. (2022). *Veritas Document 4*. https://www.mas.gov.sg/-/media/mas-media-library/news/media-releases/2022/veritas-document-4---feat-principles-assessment-case-studies.pdf

Monetary Authority of Singapore. (2025). *Consultation paper on guidelines on artificial intelligence risk management (P017-2025).* Monetary Authority of Singapore.

Ning, X., Lu, Y., Li, W., & Gupta, S. (2024). How transparency affects algorithmic advice utilization. *Decision Support Systems*, *183*, 114273. https://doi.org/10.1016/j.dss.2024.114273

Nishant, R., Nguyen, T., Teo, T., & Hsu, P.-F. (2023). Role of substantive and rhetorical signals in AI adoption announcements. *European Journal of Information Systems*. https://doi.org/10.1080/0960085X.2023.2243892

OECD. (2013). *The OECD Privacy Framework 2013*. https://www.afapdp.org/wp-content/uploads/2018/06/oecd_privacy_framework.pdf

OECD. (2024). *Regulatory approaches to artificial intelligence in finance*. https://www.oecd.org/en/publications/regulatory-approaches-to-artificial-intelligence-in-finance_f1498c02-en.html

Otoritas Jasa Keuangan. (2025). *Artificial intelligence governance for Indonesian banks*. https://ojk.go.id/en/Publikasi/Roadmap-dan-Pedoman/Perbankan/Pages/Indonesia-Artificial-Intelligence-Governance-for-Banking.aspx

Pasmore, W., Winby, S., Mohrman, S. A., & Vanasse, R. (2019). Reflections: Sociotechnical systems design and organization change. *Journal of Change Management*, *19*(2), 67–85. https://doi.org/10.1080/14697017.2018.1553761

Pombal, J., Cruz, A. F., Bravo, J., Saleiro, P., Figueiredo, M. A. T., & Bizarro, P. (2022). *Understanding unfairness in fraud detection through model and data bias interactions*. https://arxiv.org/abs/2207.06273

Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes*, *188*, 104405. https://doi.org/10.1016/j.obhdp.2025.104405

Siegmann, C., & Anderljung, M. (2022). *The Brussels effect and artificial intelligence*. GovAI. https://www.governance.ai/research-paper/brussels-effect-ai

Somu, B. (2022). AI and machine learning for predictive banking. *International Journal of Science and Research*, 1441–1456. https://doi.org/10.21275/ms2212141805

Sony, M., & Naik, S. (2020). Industry 4.0 integration with socio-technical systems theory: A systematic review and proposed theoretical model. *Technology in Society*, *61*, 101248. https://doi.org/10.1016/j.techsoc.2020.101248

Sun, Y., Sheng, D., Zhou, Z., & Wu, Y. (2024). AI hallucination: Classification of distorted information. *Humanities and Social Sciences Communications*, *11*(1). https://doi.org/10.1057/s41599-024-03811-x

The Guardian. (2018). *Revealed: 50 million Facebook profiles harvested for Cambridge Analytica*. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

Thoma, D. (2025). *The fintech experience gap: Why customers feel unseen*. https://fintechmagazine.com/news/the-fintech-experience-gap-why-customers-feel-unseen

Trist, E. L., & Bamforth, K. W. (1951). Some Social and Psychological Consequences of the Longwall Method of Coal-Getting: An Examination of the Psychological Situation and Defences of a Work Group in Relation to the Social Structure and Technological Content of the Work System. *Human Relations*, *4*(1), 3–38. https://doi.org/10.1177/001872675100400101

U.S. Department of the Treasury. (2024). *Uses, opportunities, and risks of artificial intelligence in financial services*. https://home.treasury.gov/system/files/136/Artificial-Intelligence-in-Financial-Services.pdf

Vuković, D. B., Dekpo-Adza, S., & Matović, S. (2025). AI integration in financial services: A systematic review. *Humanities and Social Sciences Communications*, *12*, 562. https://doi.org/10.1057/s41599-025-01654-4

World Economic Forum. (2025). *Artificial intelligence in financial services*. https://reports.weforum.org/docs/WEF_Artificial_Intelligence_in_Financial_Services_2025.pdf

Yuan, Y.-P., Tan, G. W., & Ooi, K.-B. (2025). What shapes mobile fintech consumers' post-adoption experience? *Technological Forecasting and Social Change*, *217*, 124162. https://doi.org/10.1016/j.techfore.2025.124162

Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys*, *53*(5), 1–40. https://doi.org/10.1145/3395046