COLUMBIA | SIPA
School of International and Public Affairs

Generative AI, Trust, and the Financial Sector:

Aligning Consumer Expectations, Executive Strategies, and Sectoral Standards

May 2025

Reported By

Ruchira Ramkrishna
Anthony Tokarz
Ziqi Yang
Dolly Zhang
Melody Zhang
Ting Zhang

Table of Contents

Executive Summary

Part 1 - Background and Context

1. Introduction to FS-ISAC and the Project
2. Regulatory Landscape
3. Literature Review

Part 2 - Research Design

4. Methodology
5. Interview Findings
    a. Consumer Interviews
    b. Executive Interviews
    c. Comparative Inference
6. Interview Insights

Part 3 - Future Research

7. Strategic Agenda for Follow-Up Research
8. Mapping Friction Zones in AI Adoption
9. Governance Models for Responsible AI in the Sector

Part 4 - Conclusions

10. Recommendations
11. Operationalizing Trust

12. Appendices
    - Appendix A: Consumer Interviewee Demographic Table
    - Appendix B: Executive Interviewee Demographic Table
    - Appendix C: Detailed Overview of Regulatory Landscape Across the U.S., EU, UK, and Asia-Pacific
    - Appendix D: Technical Approaches to Privacy Protection & Hallucination Control

13. Bibliography

Executive Summary:

This report, a collaboration between the Financial Services Information Sharing and Analysis Center (FS-ISAC) and Columbia University's School of International and Public Affairs (SIPA), examines Generative Artificial Intelligence's (GenAI) impact on consumer trust in the financial services sector. FS-ISAC's trusted role as a neutral, member-driven consortium, rooted in its global cybersecurity leadership, informs our analysis of governance challenges and strategic recommendations to pioneer responsible AI adoption in the sector.

Drawing on interviews with consumers and executives across the U.S. and Asia-Pacific, complemented by regulatory and industry analyses, we uncover trust dynamics and institutional practices shaping GenAI deployment. Consumers trust AI conditionally, favoring human oversight for complex transactions and valuing institutional reputation over regulatory frameworks. Regional nuances emerge: U.S. consumers demand transparency and human override, while Asia-Pacific consumers prioritize outcomes, reflecting pragmatic trust. Data privacy and opaque AI use fuel concerns, underscoring the need for clear communication and recourse mechanisms.

Executive insights reveal stark contrasts. Asia-Pacific institutions, notably in China and Singapore, pursue agile, performance-driven AI integration, supported by national innovation strategies and selective human oversight. U.S. institutions adopt a cautious, compliance-first approach, driven by regulatory uncertainty and expectations for transparency and comprehensive human supervision. Both regions prioritize responsible AI, but operational goals differ—efficiency and modernization take priority in the Asia-Pacific while risk containment and auditability shape the adoption of AI tools in the United States and European Union. Systemic

risks, such as model concentration, highlight the need for sector-wide collaboration across all regions.

Successful GenAI integration hinges on meeting consumer expectations for transparency and control, maintaining robust human-in-the-loop protocols, and embedding trust into both system design and organizational practices.

Established practices, such as Model Risk Management (MRM) and ethical frameworks like Singapore's FEAT Principles, provide a blueprint for effective governance of GenAI in the financial sector. Standards for data integrity, fairness, and accountability can strengthen trust, enabling FS-ISAC to lead sector-wide frameworks that balance innovation with resilience.

Recommendations:

- Establish AI use-case typologies and risk classifications to tailor governance to application sensitivity.
- Implement model registries for traceability, accountability, and predictable outputs.
- Adopt privacy-preserving technologies (e.g., differential privacy, federated learning) to safeguard data integrity and consumer privacy.
- Expand regulatory sandboxes, as piloted in Singapore and the UK, to foster controlled innovation.
- Enhance consumer transparency with plain-language disclosures, Trust Dashboards, and human override options to build confidence.
- Safeguard human-in-the-loop protocols for critical financial decisions.
- Promote consumer education through accessible explainers to demystify AI use.

Future Research: Research should track regulatory convergence on global standards (e.g., NIST AI RMF, OECD principles), shifts in consumer trust, progress on workforce reskilling for AI oversight, and the effectiveness of governance measures. FS-ISAC can lead by developing trust

metrics and public communication strategies to ensure consumers trust and understand AI-enhanced services.

By leveraging its global membership and threat intelligence expertise, FS-ISAC has the potential to shape responsible GenAI adoption by fostering innovation, resilience, and consumer confidence across the financial services sector.

In a fragmented regulatory environment– not only across US states but also across Europe and the Asia-Pacific region– **consumer trust shows greater alignment across jurisdictions** and has the potential to serve as **the centripetal force that unites financial services institutions** in aligning objectives and standards for the protection of data and responsible AI governance.

Part 1 - Background and Context

Section 1. Introduction to FS-ISAC and the Capstone Project

The Financial Services Information Sharing and Analysis Center (FS-ISAC) constitutes one of the global financial sector's hubs for conducting research, sharing information, and disseminating best practices in the context of cybersecurity. Since 1999, it has operated under a congressional recommendation to safeguard critical infrastructure through public-private collaboration. The Capstone project with Columbia University's School of International and Public Affairs dovetails with that public-private partnership by bringing together a group of students of diverse backgrounds and academic interests to examine the implications of generative AI for consumer trust and safety within the financial services sector, with an emphasis on how FS-ISAC's clients can apply the graduate student consultants' findings to entrench their positions as custodians of consumer relationships.

FS-ISAC is an independent, member-funded consortium representing more than 5,000 financial institutions across 75 countries. It serves as a neutral platform for real-time information exchange, cyber threat intelligence, and coordinated crisis response. Critically, FS-ISAC does not operate on behalf of any government; rather, it supports its members through collaborative, industry-led initiatives. Participation in its cyber-resilience exercises is voluntary, with many institutions opting in to stress-test preparedness. The consortium's operational model blends anonymized data sharing from major banks and financial infrastructure providers with strategic collaboration with entities such as the U.S. Department of Homeland Security and the Department of the Treasury. Originally established in 1999 under a U.S. Presidential Directive, FS-ISAC began as a domestically focused initiative but expanded globally following the 2008

financial crisis, recognizing cyber risk as a borderless systemic challenge. Today, FS-ISAC functions as the financial sector's principal threat intelligence hub, not only for the United States, but globally.

This report stands at the confluence of various trends in emerging technologies relevant to the financial services sector and builds upon the work that FS-ISAC has done to study the effects of generative AI (GenAI) adoption across its network of members and the financial services industry as a whole. Recent FS-ISAC initiatives include cryptographic security in the context of quantum-resistant security frameworks and systematic algorithm replacement, the publication of six white papers on adversarial AI risks and responsible implementation in the context of AI governance, and the creation of customizable assessment tools for GenAI supply chain risks in the context of vendor risk management. This report, undertaken within the context of the Capstone program at Columbia University's School of International and Public Affairs, takes AI governance as its focus and seeks to triangulate the key issues pertaining to consumer trust across three lines of effort: gauging consumer perceptions of AI trust factors, assessing executive strategies for balancing AI innovation with ethical AI deployment, and a broader scan of emerging standards, such as National Institute for Standards and Technology's (NIST) AI Risk Management Framework.

The initial findings of the team assessed four central gaps in consumers' perceptions of AI governance: unclear training data accountability, inconsistent disclosure rules across regions and institutions, imbalances in the respective risk tolerances of banks and customers, and the need to update legacy systems to pass modern AI audits. The goal of this report is to assist FS-ISAC in leveraging its neutral position in the sector to develop and disseminate useful trust

metrics and standardized global transparency rules. Furthermore, the report seeks to lay the groundwork for an AI security toolkit that might position FS-ISAC's members for success in the emerging world of AI by highlighting the questions that practitioners ought to ask with their consumers in mind. One potential fruit of further research grounded in this report is the feasibility and usefulness of a voluntary certification for AI vendors as well as in-house tools.

Objectives & Significance

The study is designed to evaluate consumer perceptions of GenAI in financial services, including trust thresholds, privacy concerns, and preferences for human vs. AI-driven interactions, analyze institutional strategies for GenAI deployment, governance frameworks, and risk mitigation practices, map regional regulatory landscapes (U.S., EU, Asia) to identify compliance challenges and opportunities for alignment, and provide FS-ISAC with evidence-based recommendations to help guide financial institutions the world over and foster trust in AI-driven finance.

Trust is the cornerstone of financial systems. As GenAI reshapes customer experiences and operational workflows, understanding its impact on trust is vital to ensuring equitable access, ethical innovation, and systemic stability. This project bridges the gap between technological advancement and stakeholder confidence, offering a roadmap for responsible AI adoption.

Triangulation Strategy

Data from consumer sentiment, executive strategies, and institutional threat analysis are cross-referenced to:

1) Identify gaps between consumer expectations and institutional practices.
2) Highlight systemic vulnerabilities requiring FS-ISAC-led collaboration.

Section 2: Regulatory Landscape

The regulatory landscape for artificial intelligence (AI) in financial services is rapidly evolving, marked by a combination of foundational cybersecurity laws, sector-specific regulations, and emerging AI-specific frameworks. While the United States, European Union, and Asia-Pacific regions each pursue distinct approaches, several cross-cutting themes and actionable insights emerge for financial institutions seeking to implement trustworthy AI.

Comparisons among emerging AI regulatory frameworks reveal varying strategies yet a shared recognition that trustworthy AI is essential for sustained public acceptance. The European Union's approach, crystallized in its AI Act, represents a comprehensive ex-ante framework for mitigating AI risks, demanding transparency, human oversight, and fairness. The United States, by contrast, relies heavily on extending existing consumer protection and anti-discrimination laws, supplemented by federal guidance (such as the NIST AI RMF), rather than a sweeping legislative act. China has enacted algorithm- and content-focused rules that integrate national security and ethical considerations, particularly around controlling deepfakes and generative AI outputs. Other jurisdictions, Japan and Singapore, combine data-centric laws and targeted guidelines to strike a balance between fintech innovation and maintaining public trust.

International organizations are actively shaping the discourse as well. IOSCO, for instance, published guidelines (2021, updated 2023/24) for securities regulators on AI governance in trading and asset management, with updates planned to address newer AI advances like large language models (The Board of the International Organization of Securities Commissions, 2025). Policymakers worldwide increasingly cite the concept of "Trustworthy

AI," referencing revised OECD AI Principles (2024) and other ethical frameworks, signaling global convergence on transparency, accountability, and safety. At forums such as the G20, leaders in 2023–2024 emphasized a "responsible AI innovation" agenda, reinforcing that effectively managing AI's risks in sectors like finance is vital to harnessing its benefits without compromising user confidence.

Regulatory frameworks globally demonstrate a deliberate interaction between innovation and enforcement. Recognizing that advanced AI can lead to substantial enhancements in financial services, including fraud detection and tailored customer experiences, policymakers emphasize the necessity of strong governance. The prevailing theme across all jurisdictions indicates that cybersecurity, transparency, and accountability are fundamental to fostering consumer confidence, thereby enabling generative AI and other advanced technologies to develop responsibly within the financial sector.

Appendix C provides further detail with full in-text citations and an integrated overview of regulatory frameworks across the U.S., EU, UK, and Asia-Pacific.

Global Trends and Takeaways

State of Play

Global North	Global South
Emphasis on binding regulations (EU), sector-specific guidance (US), and principles-based frameworks (UK). Priorities	Focus on leveraging AI for development, often with less emphasis on stringent regulatory oversight. Priorities include

include: ● Mitigating bias and discrimination. ● Ensuring transparency and explainability. ● Protecting data privacy. ● Addressing cybersecurity risks.	Promoting AI adoption in key sectors (agriculture, healthcare, education). ● Bridging the digital divide. ● Building local AI capacity. ● Adapting global standards to local contexts.

Cross-Cutting Concerns
1. Balancing Innovation and Risk Mitigation ● Global South: Struggles to foster AI adoption for development (e.g., financial inclusion) while lacking infrastructure and governance for risk management. Example: Kenya's draft AI policy prioritizes healthcare and agriculture but lacks enforceable safeguards for fintech. ● US: Sectoral enforcement (e.g., guidance from Consumer Financial Protection Bureau, U.S. Securities and Exchange Commission) risks stifling AI-driven innovation in lending and trading, with firms navigating conflicting state and federal rules. ● UK: Principles-based guidelines promote flexibility but face criticism for enabling "ethics washing" without accountability **2. Regulatory Fragmentation and Compliance Costs:** ● Global South: Weak digital governance (e.g., data privacy laws) and geopolitical pressures (e.g., US-China tech decoupling) complicate cross-border AI deployments.

- US: State laws (e.g., Colorado AI Act) clash with federal agency guidance, forcing firms to maintain parallel compliance frameworks.
- UK: Post-Brexit divergence from EU AI rules creates uncertainty for firms operating in both markets.

3. Explainability and Bias

- Global South: Limited technical expertise and data quality exacerbate risks of discriminatory AI in credit scoring and insurance.
- US: SEC and CFPB increasingly mandate explainability for AI-driven decisions, but deep learning models remain opaque.
- UK: FCA's focus on "fairness" lacks standardized metrics, leaving firms to self-assess bias risks

4. Cybersecurity and Data Privacy

- Global South: Weak cybersecurity frameworks (e.g., Indonesia, Nigeria) heighten risks of AI-powered fraud and data breaches.
- US: NYDFS Cybersecurity Regulation and FTC enforcement target AI vulnerabilities, but legacy systems hinder compliance.
- UK: DORA-like operational resilience rules are underdeveloped for AI-specific threats (e.g., adversarial attacks).

5. Capacity Gaps

- Global South: Regulators lack resources to audit AI systems, relying on foreign frameworks ill-suited to local contexts.
- US: Smaller fintechs struggle with NIST AI RMF compliance costs, favoring

incumbents.

- UK: FCA's reliance on industry self-reporting risks under-detecting AI harms.

6. Geopolitical Pressures

- Global South: Caught between US-led "risk-based" norms and China's state-centric AI governance, complicating regulatory alignment.
- US: Export controls on AI chips (e.g., NVIDIA) limit Global South's access to cutting-edge tools, perpetuating dependency.
- UK: Post-Brexit AI strategy risks isolation without strong EU or US alignment.

Emerging Flashpoints

1. Technology Dependence vs. Sovereignty

- Foreign AI Dominance: Reliance on Global North tech (e.g., U.S. cloud providers, EU AI tools) risks eroding local control and exacerbating data colonialism. Example: Kenyan fintechs depend on foreign fraud-detection AI, limiting customization for local fraud patterns.

2. Geopolitical Alignment Pressures

- US-China Decoupling: Export controls on AI chips (e.g., NVIDIA) force countries to choose between U.S. or Chinese tech ecosystems, complicating financial AI adoption.
- Regulatory Imposition: EU AI Act and NIST standards are often adopted without adaptation to local contexts, risking misaligned compliance burdens.

3. Labor Exploitation

- AI Supply Chains: Content moderators in Kenya and the Philippines face exploitative

conditions while training AI models for Global North firms.

- Bias in Financial AI: Credit-scoring models trained on Global North data exacerbate inequalities in local markets
- Labor Strategies:
    - Unions: UK's Unite is lobbying for "right to human review" clauses in AI-driven HR decisions.
    - Banks: Citi's $1B upskilling program aims to transition employees to AI oversight roles.

4. **Infrastructure Gaps**

- Energy and Connectivity: Unreliable electricity and internet access hinder AI deployment in sectors like mobile banking.
- Data Governance: Weak privacy laws (e.g., Nigeria's GDPR-inspired rules) fail to address AI-specific risks like algorithmic discrimination

5. **Cross-Border Data Governance Conflicts**

- Data Localization Laws: China's PIPL vs. ASEAN's cross-border data flow rules force firms to maintain region-specific AI models, increasing compliance costs.
    - India's DPDP Act (2023): Requires AI training data to be stored locally, complicating cloud-based AI deployments for global banks.
    - Brazil's LGPD (2024 amendments): Mandates explicit consent for AI training on customer data, slowing fintech innovation

6. **Energy and Environmental Costs**

- Sustainability Risks: Large language models (LLMs) like GPT-4 consume energy

equivalent to 120 US households annually per deployment.

- Regulatory Pressure: EU Taxonomy Regulation (2025): May classify energy-intensive AI as "unsustainable," affecting ESG compliance for banks.Singapore MAS: Incentivizes green AI through grants for low-power edge computing mode.

7. **AI-Driven Systemic Risks in Financial Markets**

- Market volatility: Algorithmic trading models amplifying market volatility (e.g., flash crashes triggered by AI-driven liquidity mismatches).
- Regulatory Gaps: UK FCA AI Sprint (2025): Identified "herding risks" where multiple firms use similar AI models, creating correlated failures. SEC (2024): Exploring rules to mandate stress-testing for AI trading algorithms but faces pushback over implementation costs.
- Case Study: The 2024 "quant quake" in Asian markets, where AI models mispriced derivatives during geopolitical unrest, exposed vulnerabilities in unsupervised learning systems.

Takeaways

- EU and China lead in prescriptive rules, sparking debates over innovation costs.
- U.S. and UK face criticism for either overreach (state laws) or under-enforcement (principles).
- APAC’s fragmentation leaves firms navigating conflicting standards, while bias and liability remain universal pain points.

For FS-ISAC members, these controversies underscore the need for proactive engagement

with regulators and investment in explainability tools to preempt compliance risks.

Regional analysis — Convergence Toward Risk-Based Regulation:

Europe

The EU AI Act (formally adopted in 2024) is recognized as the world's first comprehensive, binding AI law, classifying AI systems by risk and imposing strict requirements on high-impact sectors, especially financial services (e.g., credit scoring, fraud detection). The Act mandates transparency, human oversight, and robust risk management, effectively setting a global benchmark (Council of the EU, 2023; European Commission, 2025). Meanwhile, the UK's 2024 AI Regulation White Paper advocates a sector-led, principles-based approach, emphasizing flexibility, innovation, and regulator coordination rather than prescriptive rules (AI Working Group A&O Shearman, 2024). Despite its pioneering status, the EU AI Act raises concerns about its high-risk classification for credit scoring, fraud detection, and insurance pricing, which mandates costly compliance measures (documentation, human oversight, and third-party audits). Critics argue this could stifle innovation for smaller fintech companies while favoring incumbents with the resources to comply. Financial institutions also question whether existing sectoral laws (e.g., GDPR, MiFID II) already address AI risks, potentially making the AI Act redundant for finance (Henderson, 2024). In the UK, the FCA's principles-based approach (fairness, transparency) has been criticized for lacking binding requirements, creating a risk of "ethics washing" without real accountability (Department for Science, Innovation Technology & Office for Artificial Intelligence, 2023). Furthermore, plans to introduce binding

rules for “highly capable” AI models could alienate pro-innovation advocates who prefer the UK’s more flexible stance.

Asia

In the Asia-Pacific region, Singapore’s Model AI Governance Framework (2024 update) and Hong Kong’s SFC guidance underline agile regulation, regulatory sandboxes, and sectoral codes of conduct (Abdullah, 2025). China pursues a national AI strategy emphasizing technological advancement and global competitiveness, with regulations focusing on data security, algorithmic bias, and content moderation, reflecting a strong element of state control and innovation (Bird & Bird, 2023). Mandatory government audits for financial AI systems, enforced under the Algorithmic Recommendation Management Provisions, are viewed as a tool for state surveillance and market control, while strict security assessments for tools like LLMs limit fintech experimentation and favor state-backed entities (Cyberspace Administration of China, 2023). Meanwhile, Singapore maintains an agile, performance-oriented approach, promoting AI adoption through regulatory sandboxes and the Model AI Governance Framework to balance innovation with consumer protection (Monetary Authority of Singapore, 2018b). MAS’s voluntary FEAT Principles receive praise for flexibility but face criticism for enabling inconsistent adoption, particularly among smaller firms. There remain no explicit AI governance laws for finance in Singapore, leaving cross-border risks such as bias largely unchecked (Bamberger et al., 2018). India is developing a national AI strategy focused on leveraging AI for socio-economic development, data privacy, and ethical considerations, with draft regulations under review. ASEAN countries vary considerably in strategy, with Indonesia and Vietnam prioritizing AI-driven economic growth, and Malaysia working on ethical guidelines and

frameworks. South Korea, through its Act on Fostering the AI Industry, takes an EU-style approach by classifying AI in lending and insurance as high-risk, thereby requiring extensive compliance. Critics contend this overlooks differences in regional market maturity.

North America

In the United States, no single federal AI law exists, but there is a growing patchwork of state laws, such as the Colorado AI Act, sectoral guidance (e.g., CFPB's AI in Lending), and federal executive actions (Executive Order 14110, NIST AI RMF). The overall trend is toward risk-based, outcome-focused oversight, prioritizing explainability, fairness, and consumer protection (Anderson et al., 2024). However, the Colorado AI Bill proposes holding developers liable for discriminatory AI outcomes, diverging from federal agencies' focus on user accountability and prompting concern from industry groups over retroactive liability for legacy systems (Colorado General Assembly, 2024). Additionally, the SEC's proposed "conflict of interest" rules for AI-driven brokerages could force firms to abandon predictive algorithms, viewed by trading platforms as regulatory overreach (Federal Register, 2023). In Canada, the forthcoming Artificial Intelligence and Data Act (AIDA) will emphasize high-impact AI systems and underscores ethical AI and human rights, reflecting a broader commitment to harness AI responsibly (Ernst & Young LLP, 2023). Overall, the region moves steadily toward more comprehensive oversight, albeit through a complex interplay of federal, state, and provincial regulations.

Latin America

Countries in Latin America are shaping their AI policies with an emphasis on ethics and

responsible deployment. Brazil has proposed an AI bill of rights, underscoring ethics and human rights in AI development and deployment (Ogunkeye, 2024). Mexico is developing a national AI strategy to spur innovation, highlighting ethical considerations and international cooperation. Chile has a national AI strategy focusing on ethics, governance, and promoting AI adoption across various sectors. This regional momentum demonstrates a growing interest in aligning AI innovation with social and developmental priorities (Tony Blair Institute for Global Change, 2025).

Africa

At a pan-African level, the African Union has drafted an AI strategy that emphasizes using AI for socio-economic development while addressing ethical considerations (Anthony et al., 2024). In individual countries such as Kenya, Nigeria, and South Africa, policy efforts center on leveraging AI for economic growth, tackling ethical challenges, and building local AI capacity. Although each nation is at a different stage, the overall trend points toward an increasingly coordinated approach to AI regulation and development across the continent (Tony Blair Institute for Global Change, 2025).

Insights for financial institutions

Across all jurisdictions, there is a convergence on transparency, accountability, human oversight, and risk management as fundamental principles (Aldasoro et al., 2024). Institutions should proactively align with global best practices, such as the NIST AI RMF and emerging EU AI Act requirements, rather than waiting for final, prescriptive local laws (European Commission, 2024). This anticipatory approach ensures readiness for regulatory harmonization

and cross-border operations. Moreover, regulatory arbitrage is shrinking as international standards converge, reducing the window for exploiting gaps; multinational financial institutions must therefore prepare to meet consistently high standards in every market (Adria, 2024).

Application and enforcement

A shift is evident from voluntary guidance to active supervision and enforcement, as regulators intensify oversight, visible in the EU and various US states, where non-compliance now carries rising reputational and financial risks (Council of the EU, 2023). Sector-specific agencies such as the EBA, CFPB, and MAS have issued AI guidance clarifying expectations for model risk management, data governance, and consumer protection, signaling that accountability and compliance obligations will only become more stringent (Monetary Authority of Singapore, 2021).

Noteworthy gaps

Despite increasing regulation, challenges remain. The explainability of complex AI models, especially deep learning and generative AI, continues to be a sticking point for both regulators and industry, leading to calls for standardized explainability metrics and tools (Cheong, 2024). Third-party and vendor risk is also under heightened scrutiny, with regulators advocating mandatory model registries and robust audit trails to address every stage of the AI supply chain. These gaps reflect the continuously evolving nature of AI oversight and the need for agile regulatory and institutional responses.

Robust safeguards alone do not address AI's explainability and vendor risk challenges. Institutions must develop clear disclosures and standardized metrics to enhance transparency. FS-ISAC can spearhead standardization by promoting sector-wide explainability tools and audit trails to bridge regulatory gaps and build consumer trust in AI-enhanced financial tools.

Section 3: Literature Review

Trustworthy AI and Governance in Financial Services

Generative AI (GenAI) is progressively embedded in financial products, from chat-based customer service to investment advisory systems. Yet this rapid uptake prompts fundamental trust questions: Will consumers trust AI-driven financial decision-making, and how are industry participants and policymakers mitigating associated risks? Recent evidence suggests a persistent trust gap, despite growing GenAI adoption. One 2024 U.S. household survey found nearly half of respondents had used GenAI tools, yet overall they reported lower trust in AI than in human experts when it comes to critical domains like banking (Aldasoro, Armantier, et al., 2024). The same study noted widespread fears around data misuse, support for stricter AI regulations, and a tendency to trust banks and regulators over technology firms in safeguarding financial data (Aldasoro, Armantier, et al., 2024).

A *BIS – FSI Insights (2024)* report, "Regulating AI in the financial sector: recent developments and main challenges," analyzes global regulatory attitudes toward AI in banking, insurance, and related fields. Researchers highlight a consensus around key trustworthiness principles, including reliability, accountability, transparency, fairness, ethics, privacy, and security (Crisanto et al., 2024). Few jurisdictions have passed explicit AI rules for financial services; instead, they rely on overarching risk management and consumer protection regulations. Regulators emphasize firm governance, bias control, and model oversight, and they increasingly warn institutions about the perils of using GenAI for direct client interactions. Common apprehensions involve AI's tendency toward errors, complexities in compliance,

accountability gaps for mistakes, and limited consumer willingness to rely on AI-advised decisions (Crisanto et al., 2024). As a result, many banks deploy GenAI primarily for back-office tasks, such as document processing or fraud detection, where trust risks are lower. This approach lets financial institutions gain experience with AI technologies without immediately jeopardizing consumer relationships.

Beyond private firms, a variety of non-academic organizations, including regulators, international standard-setters, and industry consortia, are taking steps to prevent AI-driven erosion of societal trust. Global policy forums increasingly place "trustworthy AI" front and center. The G20 declared a commitment to human-centric AI, adopting AI Principles (building on the OECD's guidelines) in 2019, while the G7's 2023 Hiroshima AI Process coordinates safe AI strategies among major economies (Crisanto et al., 2024). These efforts recognize that if GenAI becomes integral to sectors like finance, it must adhere to robust ethical standards to maintain public confidence.

A *BIS Working Paper (2024)*, titled *"Intelligent Financial System: How AI is Transforming Finance,"* notes that swift adoption of AI across banking, insurance, asset management, and payment platforms presents both operational opportunities and systemic risks (Aldasoro et al., 2024). The authors argue for updating financial regulations in line with widely endorsed AI governance principles, transparency, accountability, fairness, safety, and human oversight, so that AI remains trustworthy as it scales. They emphasize that each stage of AI advancement (including GenAI) introduces new challenges for bias, data privacy, and systemic

interdependence. Their conclusion urges closer global regulatory collaboration to address these complexities in a consistent manner (Aldasoro, Gambacorta, et al., 2024).

The *OECD (2024) – "Regulatory Approaches to Artificial Intelligence in Finance"* report collated perspectives from 49 jurisdictions and found near-universal concern over potential systemic risks that AI may introduce, such as concentration risk or herd behavior if many financial firms depend on the same AI models (OCED, 2024). Respondents point to "black-box" opacity and limited interpretability as core trust barriers. While most jurisdictions rely on existing regulatory frameworks (covering outsourcing, model risk, or operational risk) to address AI, some are crafting new codes of conduct around responsible AI. This global snapshot further illustrates the rising importance of trust, reliability, and transparency in AI's integration into finance (OCED, 2024).

A joint *Turing Institute & HSBC (2024)* study, *"The Impact of Large Language Models in Finance: Towards Trustworthy Adoption,"* builds on industry workshops and warns of specific challenges faced by banks deploying LLMs (Maple et al., 2024). Although LLMs offer promising utility across areas like customer engagement and risk analysis, participants stressed ongoing concerns over data leaks, inaccurate or biased outputs, and the need for more stringent governance. Many institutions' existing machine learning frameworks are not yet equipped to manage the higher stakes associated with large-scale language models. Practitioners thus advocate stricter policies on privacy, data security, and oversight to preserve consumer trust and comply with broad financial regulations on fairness and consumer protection (Maple et al., 2024).

In a similar vein, the *U.S. Treasury (2023/24) – "Artificial Intelligence in Financial Services: Uses, Risks, and Recommendations"* outlines governance imperatives drawn from industry feedback. Treasury advises that regulators, firms, and global standard-setters harmonize robust AI standards and identify any regulatory "blind spots." This includes clarifying how existing rules, from privacy statutes to fair lending laws, apply specifically to AI-driven systems (U.S. Department of the Treasury, 2024). The report also recommends adopting risk management frameworks (inspired by model risk oversight) to address AI-specific threats such as bias, cyber vulnerabilities, and uncertain accountability for automated decisions. These proposals reflect a U.S. policy pivot aimed at nurturing AI innovation "within guardrails," ensuring it does not erode consumer trust or broader financial stability (U.S. Department of the Treasury, 2024).

Comunale & Manera (IMF, 2024), in *"The Economic Impacts and the Regulation of AI: A Review…"*, offer a wide-ranging study of global AI governance. They find that few jurisdictions explicitly address financial stability risks introduced by AI, focusing instead on ethics, privacy, or antitrust concerns (Comunale, 2024). Approaches diverge significantly: the EU (and similarly Brazil) pursue ex-ante, risk-based regulation exemplified by the EU AI Act; the United States leans on existing regulatory agencies and sector-specific guidelines; China focuses on algorithmic oversight and ethical reviews; the UK employs a sector-led strategy outlined in its 2023 white paper; and Japan/India maintain more flexible stances with minimal AI-specific laws (Comunale, 2024). Yet all policymakers grapple with balancing innovation against potential biases, reduced explainability, or cybersecurity weaknesses. The IMF emphasizes that no

universal template has emerged, suggesting that international collaboration remains critical to address these evolving risks uniformly (Comunale, 2024).

Broad AI Governance Frameworks

A growing body of literature examines what "trustworthy AI" entails, identifying shared principles like human oversight, fairness, transparency, privacy, security, robustness, and accountability. For instance, Kowald et al. (2024) in *"Establishing and evaluating trustworthy AI: overview and research challenges"* (Frontiers in Big Data) compile six core requirements for AI to be considered trustworthy, spanning human agency, fairness, transparency, accuracy, privacy, and accountability, and detail practical and theoretical gaps in achieving them (Kowald et al., 2024). Their taxonomy aligns closely with global governance models, including those from the EU and NIST, thereby offering policymakers and AI practitioners a consolidated framework.

Meanwhile, Cheong (2024) in *"Transparency and accountability in AI systems: safeguarding well-being in the age of algorithmic decision-making"* (Frontiers in Human Dynamics) focuses specifically on two pillars: transparency (making AI operations open and explainable) and accountability (ensuring systems and their operators can be held responsible for outcomes) (Cheong, 2024). The author contends that realizing these features requires not just technical breakthroughs, like explainable AI and auditing tools, but also robust laws, inter-agency coordination, and ethical stakeholder collaboration. By weaving together research from engineering, jurisprudence, and social sciences, Cheong underscores that transparency and accountability are interdisciplinary challenges affecting individual and collective well-being (Cheong, 2024).

Moreover, recent industry scholarship in the United States has further underscored the importance of ensuring trust and confidentiality when evaluating algorithmic fairness in finance. One such approach is the Privacy-Preserving Probabilistic Race/Ethnicity Estimation (PPRE) framework, proposed which draws on local differential privacy, secure two-party computation, and the Bayesian Improved Surname Geocoding (BISG) technique to measure race-based disparities without creating new disclosures of sensitive personal data (Badrinarayanan et al., 2024). In this method, firms can audit race/ethnicity fairness by combining limited self-reported demographics with probabilistic BISG-based estimates, rather than assigning explicit demographic labels. By encrypting, anonymizing, and minimizing sensitive identifiers, PPRE aligns with central privacy and transparency norms in U.S. policy. Such mechanisms could signal private-sector momentum to integrate advanced privacy protections into AI fairness assessments.

Part 2 - Research Design

Section 4: Methodology

This research project adopts a qualitative approach, drawing on in-depth interviews with two key stakeholder groups within the financial sector: consumers and industry executives responsible for overseeing service development and deployment. The objective is to capture diverse perspectives on the intersection of artificial intelligence (AI), cybersecurity, regulation, and consumer trust. By integrating insights from both participant groups, the final deliverable

will offer practical, forward-looking recommendations to guide financial institutions in strengthening their stewardship of consumer trust in the age of AI.

The first component of the research focuses on consumer interviews, with a sample of 15 participants drawn from both the United States and the Asia-Pacific region. This cross-regional approach is intended to capture differences in consumer attitudes, expectations, and concerns related to AI adoption in financial services, as well as varying perceptions shaped by local regulatory environments. The comparative analysis will help identify common trust-related themes as well as region-specific sensitivities, enabling a nuanced understanding of how financial firms might tailor trust-building strategies.

The second component involves executive interviews with financial industry leaders. These interviews, conducted in March, explore the strategic priorities, challenges, and practices related to AI implementation within financial institutions. Particular emphasis is placed on how executives manage regulatory compliance and ensure consumer trust in AI-powered products and services. These conversations provide an institutional perspective on balancing innovation with ethical and regulatory obligations, and will inform the development of best practice recommendations that align with both consumer expectations and organizational capabilities.

Section 5: Interview Insights

a. Consumer Interviews (APPENDIX Table 1)

This section synthesizes insights from fifteen consumer interviews conducted in the United States between February and March 2025, exploring how everyday users perceive and experience AI-enabled banking services.

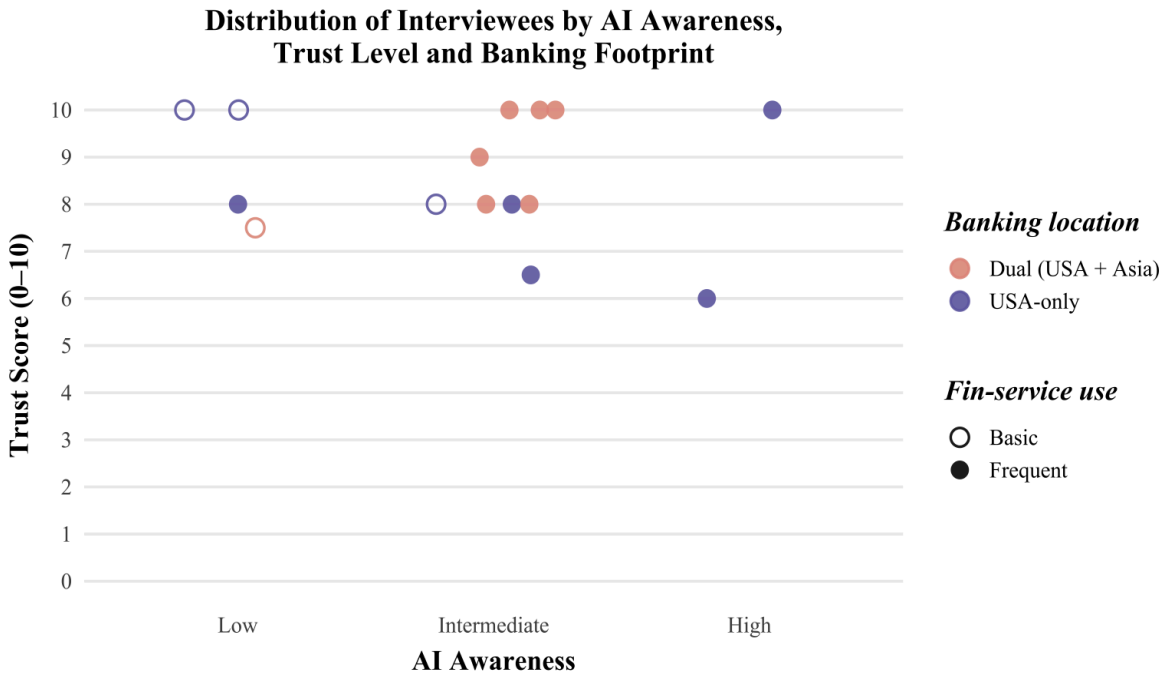

Figure 1. Distribution of Interviewees by AI-Awareness, Trust Level, and Banking Footprint

Note. *All respondents maintain at least one U.S.-based bank account. Color coding in Figure 1 differentiates those who also actively bank with an institution in Asia (coral) from those whose day-to-day banking is confined to the U.S. (purple).*

The consumer interviews revealed layered and often ambivalent perceptions of how generative AI is being integrated into financial services. Participants spanned the United States, China, South Korea, and Vietnam, representing diverse banking behaviors and varying levels of AI awareness. Most were regular users of online banking, with usage driven by either convenience or necessity. Their financial engagement ranged from routine actions, such as checking balances and paying bills, to more complex activities like managing credit cards, loans, and personal investments. AI awareness, as used here, denotes participants' familiarity with and personal use of general-purpose GenAI platforms; it does not measure their knowledge of banks' own AI deployments. Across respondents of diverse geographic and cultural backgrounds, no

material regional variation in attitudes toward AI emerged, and any observable differences reflected cultural nuance rather than fundamental divergence.

A prominent theme was the conditional nature of consumer trust in financial institutions that use AI. While most participants expressed mid-to-high levels of trust in their banks (typically rating them between 6 and 9 out of 10), this trust was not grounded in a deep understanding of AI technology. Instead, it was anchored in familiar proxies: institutional reputation, global presence, past positive experiences, and assumptions about regulatory oversight. Larger, more established banks were perceived as more trustworthy than digital-only or regional institutions, primarily due to their presumed investment in security infrastructure and institutional accountability. However, this trust remained superficial, many participants admitted they had not encountered situations that truly tested their bank's AI-driven systems or transparency protocols. Likewise, participants' trust in financial institutions appeared unaffected by their personal values or by perceptions of banks' AI usage, resting instead on an implicit belief that appropriate regulatory and organizational safeguards are already in place.

Participants consistently expressed a preference for human interaction in high-stakes financial matters, even while recognizing the advantages of AI in speed and efficiency. AI was widely accepted for transactional tasks, such as issuing fraud alerts, summarizing account activity, or sending payment reminders. However, resistance grew sharply when AI was involved in emotionally sensitive or complex services such as financial advising, loan decisions, or dispute resolution. Many emphasized that empathy, judgment, and contextual reasoning are essential in such cases, qualities they felt AI could not reliably provide. Several participants recounted frustrating experiences with chatbots that failed to understand their concerns or resolve

nuanced issues, reinforcing a belief that full automation lacks the subtlety required in human-centered services.

This tension reveals a strategic challenge for financial institutions: AI can deliver clear operational benefits, such as real-time responsiveness and reduced service costs, but over-reliance risks alienating consumers. The concept of "faceless banking" emerged in interviews, where decision-making feels opaque and impersonal, highlighting the psychological risks of removing human touchpoints. Some described AI interfaces as "annoying," especially when systems acted prematurely (e.g., flagging legitimate transactions) or failed to offer a straightforward path to a human representative. These frustrations indicate that even small lapses in user experience can undermine institutional credibility and erode the emotional capital banks have built through responsive customer service.

Data privacy and informed consent were also recurring concerns. While most consumers accepted that their data was being collected, few understood the extent or purpose of its use in AI-driven systems. Interviewees generally had not considered the issue of AI transparency prior to the conversation. However, when prompted with questions about AI use in the financial industry, many began to express concerns. Several participants suspected that banks were already using AI behind the scenes without explicit disclosure, which contributed to a sense of discomfort. As the discussion progressed, participants increasingly emphasized the importance of transparent communication around AI practices, particularly regarding how algorithms are trained, the types of data used, and the extent of human oversight involved. The concern was not

AI use per se, but the opacity surrounding it. Consumers sought assurance that AI would not act independently in sensitive areas without their knowledge or the option to opt out.

Compounding this issue was a perceived regulatory gap, which added another layer of anxiety to already tentative consumer perceptions. Many interviewees, regardless of geographic or demographic differences, voiced uncertainty about whether current legal and institutional frameworks were adequately equipped to address the complexities introduced by AI. There was a strong belief that while data privacy laws and financial regulations existed, they were designed for a pre-AI era and had not evolved in tandem with emerging technologies. This led to concerns that consumers might be left unprotected in scenarios involving algorithmic error, bias in credit scoring, or misuse of personal data.

However, while participants consistently called for "stricter regulation" or "more oversight," few could specify what form such regulation should take. This disconnect did not reflect apathy but rather a lack of clarity around how AI systems function and where accountability lies. AI was often perceived as a black box, difficult to interrogate, much less govern. The calls for regulation were rooted more in a general desire for fairness, transparency, and recourse than in concrete policy suggestions. Some floated ideas such as mandatory human review of high-stakes decisions, clearer data consent mechanisms, or greater algorithmic transparency, but these suggestions remained tentative and often based on intuition rather than legal or technical understanding.

Even participants with relatively high digital literacy admitted they would not know how to seek recourse if an AI system made a detrimental financial decision on their behalf, such as

denying a loan, misclassifying a transaction, or producing a flawed risk assessment. The absence of clear guidelines about who would be held accountable in such cases, banks, third-party vendors, or developers, fueled a perception of vulnerability. For some, this lack of visible or accessible oversight mechanisms made the use of AI feel more like a gamble than a guarantee of efficiency.

Moreover, several participants noted that regulatory agencies rarely communicate with the public about how they are adapting to AI advancements. This informational void intensified feelings of opacity and helplessness. Without assurances that regulators are actively monitoring and updating compliance protocols for AI, consumers were left to rely on institutional reputation and hope that "someone must be watching"—an uneasy foundation for trust in a high-stakes domain like finance.

When discussing AI-powered financial advice, opinions were sharply divided. A subset of participants categorically rejected AI-generated guidance, arguing that personal finance requires emotional intelligence, contextual understanding, and a long-term relationship, elements they believe AI cannot replicate. Others, particularly those with greater digital familiarity, were more open to AI-assisted recommendations, but under specific conditions. These included the presence of human review, the ability to trace and understand the algorithm's logic, and the option to override suggestions. Many viewed AI as a helpful second opinion or data source but resisted the idea of it serving as a standalone advisor. These distinctions point to the importance of designing AI systems that augment, rather than replace, human decision-making.

Taken together, the findings reinforce a growing consumer appetite for hybrid AI-human service models. Consumers are not rejecting AI outright, they show intention to acknowledge its utility in routine tasks and even as a supplement to financial guidance. However, they expect it to operate within well-defined boundaries, with transparent communication, built-in fail-safes, and the ability to escalate to a human agent when needed. Especially in scenarios involving financial risk or emotional stress, the human element remains indispensable.

While the interviews produced rich and valuable qualitative insights, several limitations must be acknowledged. First, the sample, though broad for a qualitative study, was not statistically representative of the global population. The findings are intended to highlight directional themes and patterns rather than make generalizable claims across all demographics or regions.

Second, although the sample intentionally spans multiple cultures, it remains geographically narrow, focusing on the U.S., China, South Korea, and Vietnam. Seven participants were born in mainland China, and three identify as Korean, Vietnamese, or Chinese-American; all fifteen, however, maintain U.S. bank accounts and conduct at least part of their financial lives in U.S. dollars. As a result, views shaped by European, Latin-American, African, or wholly non-U.S. banking environments are absent. Because trust, technology adoption, and customer service expectations are deeply influenced by local regulations and market experience, the findings should be read as indicative rather than comprehensive, capturing only a slice of the global diversity in consumer attitudes toward AI in finance.

Lastly, because the interviews were self-reported, findings may be influenced by social desirability bias. Participants may have overstated their trust in institutions or comfort with AI to appear informed or cautious, or understated concerns due to unfamiliarity. Additionally, attitudes toward AI are evolving rapidly, and views captured at this moment may shift as technologies mature, regulations develop, and consumers gain more direct experience interacting with AI-powered systems.

b. Executive Interview (APPENDIX Table 2)

This section synthesizes insights from ten executive interviews conducted across China, Singapore, and the United States, and supplements them with broader analysis of regional trends and policy contexts. The participating executives represent a broad cross-section of the financial services ecosystem, spanning investment banking, legal advisory, compliance, risk management, AI technology, and academia. The goal is to assess how institutions are approaching the adoption of GenAI, how risks are being managed, and how the interplay between innovation, trust, and governance is evolving in each jurisdiction.

Asia-Pacific

Financial institutions in Asia, particularly those based in China and Singapore, are embedding GenAI at both the strategic and operational levels. Regulatory frameworks like Singapore's Monetary Authority of Singapore (MAS) Veritas Toolkit and China's segmented enterprise governance frameworks provide an environment conducive to rapid yet bounded experimentation. Within this landscape, GenAI is increasingly treated as infrastructure, rather than innovation.

Executives across Asian markets see GenAI as a means of amplifying internal efficiency, enabling faster research, multilingual contract translation, and intelligent regulatory scanning. In many cases, GenAI is deployed to reduce the time spent on routine workflows by over 50%, with oversight personnel shifted toward reviewing and contextualizing AI-generated outputs. One executive (EX-02) noted, “The models don’t need to be perfect. They just need to save us time so we can focus on decisions.”

Trust in AI-augmented services is outcome-based rather than process-driven. Executives reported that clients, especially domestic ones, care more about delivery speed and reliability than whether a human or AI generated the initial result. Several firms report not disclosing AI involvement unless directly asked, and instead focus on the reliability of the final service. “ No client has ever asked, ‘Did AI do this?’ They only ask, ‘Is this accurate and on time?’” said a senior executive from a Singapore-based advisory firm (EX-02). While this highlights a prevailing emphasis on performance over provenance in client-facing services, it does not mean outputs go unchecked. Executives emphasized that internally, validation steps are still taken, albeit streamlined and context-specific. Typically, GenAI is deployed for first-draft generation, translation, or information retrieval tasks, with human review built into final delivery processes. The reliability of AI outputs is not assumed; instead, validation is integrated discreetly, ensuring that quality control occurs without disrupting perceived speed or simplicity for the client. In this model, GenAI acts as a productivity layer, but not a final authority, especially in legal, regulatory, or advisory contexts.

Ethical considerations are embedded into performance expectations rather than distinct frameworks. In practical terms, AI is expected to deliver results that are unbiased, secure, and confidential. Explainability is not prioritized unless it becomes relevant to a failure case. One executive (EX-03) summarized it as, “We don't need to explain how the sausage is made, just that it's safe to eat.” Firms mitigate ethical risk by deploying GenAI internally or in back-office operations first, and only later extending into client-facing functions. “Our legal team uses GenAI to create first drafts, but it never goes to a client without human editing,” explained a law firm partner in Beijing (EX-04).

Executives also highlighted external drivers such as national digital strategies, investor interest in “AI maturity,” and internal imperatives for modernization. Several institutions in Asia have rebranded their operational upgrades as part of broader AI transformation efforts to signal alignment with government priorities and industry trends. “AI adoption is not just operational, it’s reputational,” one executive stated “Being seen as an early mover helps us win clients and talent” (EX-04).

Furthermore, the workforce is rapidly adapting. One executive (EX-04) described how junior analysts are now expected to “learn prompting before learning Excel” reflecting a significant reorientation of baseline digital fluency. Firms are investing in retraining programs focused on prompt design, GenAI troubleshooting, and client communication rather than traditional coding skills. Despite the aggressive pace, human oversight remains core to operations. As one compliance officer (EX-04) remarked, “AI helps us think faster, but we still think.” This statement captures a key insight from the Asia-Pacific region: GenAI is being

integrated not as a substitute for human decision-making, but as a catalyst for more efficient and focused cognition. In practice, GenAI tools are used to expedite repetitive or time-consuming tasks, such as document drafting, contract parsing, or regulatory scanning, allowing personnel to reallocate cognitive bandwidth to interpretive, relational, or risk-sensitive functions.

This human-in-the-loop approach reflects a distinct institutional logic: efficiency gains from GenAI are only valuable when paired with preserved human accountability. Even in jurisdictions where regulatory scrutiny is less prescriptive than in the U.S., firms voluntarily maintain human checkpoints as part of internal governance structures. The goal is not full automation, but informed acceleration, using GenAI to augment professional capabilities without compromising oversight or ethical standards. In this model, AI-generated outputs are reviewed and often refined by domain experts before they are acted upon or client-facing.

For financial institutions more broadly, this offers a replicable blueprint: GenAI should be embedded as a decision-support infrastructure, not as a decision authority. The strategic takeaway is that trust, whether institutional or client-facing, hinges not just on AI performance, but on the visible retention of human judgment at critical junctures. This approach not only mitigates reputational and compliance risks, but also aligns with consumer expectations for transparency and recourse in high-stakes financial interactions.

United States

U.S. financial institutions are more cautious in adopting GenAI, constrained by a multi-layered risk environment and fragmented regulatory oversight. Adoption is often siloed, with innovation teams pursuing GenAI pilots under the strict supervision of compliance and

legal departments. In the U.S., internal governance concerns specific to GenAI, such as model explainability, audit trails, risk classification, and regulatory compliance, frequently outweigh broader innovation ambitions. Rather than leading with experimentation, many institutions are prioritizing the development of internal oversight protocols that can preempt legal exposure or reputational risk. This results in a cautious deployment approach, where GenAI pilots are tightly scoped, heavily supervised, and often limited to low-risk, internal applications. Innovation is not absent, but it is subordinated to governance structures that emphasize control, traceability, and accountability.

Executives universally described a "trust, but verify" philosophy. GenAI is being introduced gradually, primarily for internal use cases like market research, risk flagging, and data summarization. Client-facing applications are permitted only when the institution can guarantee auditability, traceability, and full human oversight. As one executive (EX-06) explained, "We are still in the phase of training the trainer, ensuring our teams can catch when the model is wrong before it gets to the client."

"Human-in-the-loop is not merely a design feature but a mandated requirement. GenAI-generated content, even when used for internal purposes, must be flagged, verified, and logged (EX-06)". This infrastructure reflects a defensive posture, shaped by legal accountability and a deeply entrenched compliance culture. One executive (EX-06) referred to this approach as "an invisible scaffolding for trust." Executives emphasized that customer trust is the linchpin. Even tech-forward clients expect clear disclosures and human override options. Transparency isn't a differentiator, it's a baseline expectation. One executive (EX-05) stated, "We earn trust by

not giving anyone a reason to doubt us, whether AI is involved or not." and "Clients don't mind us using AI, they just don't want to feel like they're talking to a machine."

From a workforce standpoint, GenAI is reshaping skill requirements. Entry-level roles in data processing or research are diminishing in scope. Firms are retraining employees to act as AI interpreters, validating, refining, and deploying AI outputs in high-context environments. This transition reflects broader industry trends where GenAI acts not as a job eliminator, but as a role transformer. One executive (EX-05) noted, "We're not replacing humans, we're repurposing them. AI is doing the grunt work, but people still own the narrative."

Ethical oversight remains compliance-led. Internal AI committees, risk ratings, and fairness checks are embedded into AI tool development processes. Executives openly admitted that ethical policies are often modeled on existing regulatory templates. One executive (EX-09) commented, "Ethics is governance in this space. We don't need more principles, we need enforcement mechanisms." Transparency and documentation are emphasized not for philosophical reasons but to preempt litigation or regulatory action.

Executives remain optimistic but realistic. They cite hallucination risks, lack of model explainability, and regulatory ambiguity as barriers to rapid deployment. One interviewee (EX-05) described their approach as "cautious innovation," while another (EX-10) called it "fast in idea, slow in rollout." At present, GenAI is treated as a powerful assistant, not a decision-maker, and is strictly bound by governance protocols.

Several institutions are also grappling with data access issues. Because many financial documents contain proprietary or personally identifiable information, using cloud-based GenAI

models is off-limits. “If we can’t host it on-prem, we won’t use it at all,” one compliance executive (EX-10) noted, emphasizing the critical importance of data control and security in GenAI adoption. In many U.S. financial institutions, especially those handling proprietary models or sensitive client data, cloud-based AI services are viewed as too risky due to concerns over third-party access, data leakage, or regulatory non-compliance. As a result, these firms are prioritizing on-premises deployment, where AI models and data infrastructure remain entirely within institutional firewalls. This requirement, however, imposes significant technical and financial barriers, particularly for smaller firms that lack the in-house capacity to develop and maintain private large language models (LLMs).

While GenAI adoption is progressing, it is doing so along tightly controlled rails. The path forward will depend not only on technology readiness, but also on alignment between innovation teams and gatekeepers and, increasingly, on clearer regulatory signals from federal agencies.

c. Comparative Inference

The table below summarizes the structural and strategic differences between Asian and U.S. approaches to GenAI in financial services:

	Asia (China & Singapore)	**United States**
Regulatory Strategy	Principle-based, pro-innovation frameworks supported by national policy. Internal tools prioritized; public-facing systems more tightly governed.	Fragmented, rule-based oversight. Regulations emphasize auditability, explainability, and consumer protection.

Innovation Culture	Agile, aligned with government mandates. GenAI adoption signals modernization and competitiveness.	Cautious, shaped by reputational and litigation risk. Innovation often slowed by compliance gatekeeping.
Talent Strategy	National AI programs drive reskilling and creation of AI-native roles (e.g., prompt engineers).	Certifications and internal reskilling programs focused on governance and human-in-the-loop competencies.
Governance Models	Output-driven performance standards. Light-touch oversight for internal systems; compliance through outcomes.	Formalized governance boards and procedural ethics frameworks. Ethics is embedded in compliance infrastructure.
Public Trust Model	Trust is outcome-based and institutionally granted. Transparency is often optional.	Trust is process-based and legally bound. Clients demand disclosures, documentation, and human fallback.
Use Case Scope	Broad functional deployment across compliance, legal, research, and operations.	Narrow pilot use cases in low-risk areas. Few client-facing deployments without Human-in-the-Loop (HITL) systems.
Organizational Design	Centralized AI steering supported by top-down government initiatives.	Siloed innovation with compliance as final gatekeeper. Coordination challenges limit speed of adoption.
Feedback Loops	Limited formal feedback collection; success is measured by internal KPIs.	Reactive feedback loops often triggered by error or pushback. Limited preemptive client testing.
Strategic Purpose	GenAI is both a modernization tool and a symbolic asset aligned with national narratives.	GenAI is viewed as a compliance-bound efficiency driver rather than a strategic differentiator.

Financial institutions across both Asia and the U.S. are pursuing GenAI as a strategic enabler, but through divergent operational logics. Asian firms operate under state-aligned, efficiency-driven mandates that allow rapid experimentation and performance-based deployment. Trust is assumed if outcomes are strong, and innovation is tied closely to national strategy.

In contrast, U.S. institutions follow a governance-first pathway. GenAI systems are tightly monitored, client trust is earned through transparency and redundancy, and adoption is shaped by regulatory uncertainty. While the U.S. system provides strong legal safeguards, it also creates friction that slows organizational learning.

Across both regions, a shared truth emerges: GenAI is transforming financial services, but not replacing human judgment. Its most effective use cases are those that enhance decision-making, not automate it entirely. The future will not be AI-first, but AI-augmented, where GenAI serves as a tool that expands the capacity of financial professionals rather than displacing them.

Section 6: Interview Key Takeaways

The findings of this report underscore that the integration of GenAI into the financial services sector presents not only technological opportunities but also a series of governance, ethical, and communicative challenges that must be addressed to preserve institutional trust. While GenAI is increasingly employed to enhance operational efficiency, reduce costs, and improve response times, its successful deployment in consumer-facing contexts remains

contingent on reconciling divergent expectations between financial institutions and the public they serve.

Evidence from consumer interviews reveals that trust in financial institutions is often retrospective, reputational, and implicitly granted, grounded in perceived institutional legitimacy and brand familiarity. However, this form of trust is not synonymous with endorsement of AI-driven services. Consumers expressed conditional acceptance of GenAI, with a marked distinction between tolerable and unacceptable use cases. While automated systems are widely accepted for transactional functions, such as sending payment reminders or flagging suspicious activity, there is notable resistance to GenAI applications in more complex, high-impact areas such as credit assessments, financial advice, or dispute resolution. In such domains, consumers consistently prioritized human judgment, empathy, and recourse mechanisms, attributes they believe AI systems cannot reliably replicate.

A central theme that emerged across interviews was discomfort with the opaque and often undisclosed nature of GenAI deployment. Many participants were unaware that AI may already be operating in the background of their financial interactions. When made aware, they emphasized the importance of transparency, human fallback options, and the right to opt out. These responses suggest that trust in GenAI is not necessarily a function of technological accuracy or efficiency, but rather of users' ability to understand and control their interactions

with it. The absence of visibility into how AI systems operate, and the lack of clearly communicated safeguards, were perceived as significant contributors to user unease.

The executive interviews further illuminated the institutional logics shaping GenAI governance. Notably, a clear divergence was observed between institutions in the Asia-Pacific region and those in the United States. In jurisdictions such as China and Singapore, GenAI is frequently understood as an infrastructural asset, integral to national innovation agendas and deployed extensively to enhance internal efficiencies. Institutional trust in these contexts is often outcome-based: clients are seen to evaluate service quality based on delivery speed and reliability rather than the underlying processes or transparency of the tools involved. Consequently, explainability is not treated as a fundamental design principle unless precipitated by error or client inquiry.

In contrast, financial institutions in the United States exhibit a more risk-averse posture. Adoption of GenAI is incremental and often limited to internal applications. Compliance and legal departments maintain significant oversight, reflecting concerns over liability, reputational harm, and regulatory uncertainty. Trust is constructed through procedural guarantees, documentation, and human-in-the-loop systems. Executive accounts from U.S. institutions emphasize the primacy of auditability and accountability, with GenAI framed as a support mechanism rather than an autonomous agent. In this context, the pathway to adoption is not

merely technical but institutional, requiring alignment between innovation teams and risk gatekeepers.

Despite regional differences, institutions in both geographies converge on a common operational principle: GenAI is not deployed as a substitute for human decision-making but rather as a tool for augmenting professional judgment. The most effective implementations thus far have been those in which GenAI assists with document review, language translation, and regulatory research, allowing personnel to reallocate time toward higher-order analysis. However, this transformation is not without implications. The division of labor within financial institutions is shifting, requiring staff to develop new competencies such as prompt engineering, AI supervision, and ethical risk management. These evolving roles signal a broader institutional recalibration, in which trust is no longer invested solely in human expertise but also in the governance architecture surrounding machine outputs.

At a systemic level, this report identifies a growing misalignment between institutional governance structures and consumer expectations. Financial firms are investing heavily in internal GenAI governance mechanisms, but these efforts often remain invisible to end-users. As a result, consumers may continue to perceive AI systems as unaccountable and inaccessible, regardless of how well-structured the internal protocols may be. Conversely, institutions that operate under the assumption that high-quality output is sufficient to earn trust may underestimate the normative importance consumers place on transparency, control, and fairness.

The integration of GenAI into financial services is not merely a technical or operational transition; it is a trust transformation that demands new forms of institutional accountability and public engagement. The success of GenAI systems in this domain will depend not only on their efficiency or accuracy but on their perceived legitimacy, intelligibility, and ethical alignment. The current landscape suggests that institutions must move beyond a compliance-oriented approach and begin to treat trust as a design parameter, one that is embedded into the very architecture of AI systems and visibly communicated to the users they affect.

Part 3 - Future Research

Section 7: Strategic Agenda for Follow-Up Research

This section contains additional resources and recommendations for those readers that might wish to extend this research, taking into account developments in the global regulatory landscape and the interaction of public perceptions of AI and advances in its development by the private sector firms catering to the public. The recommendations flow from the methodology employed by the 2025 SIPA Capstone team. In the interest of transparency, which this report finds lies at the heart of efforts both to regulate AI and to render its use accessible to a broader public, the recommendations below replicate the process developed for the systematic review of relevant laws, regulations, and practices.

Monitor Evolving AI Governance Policies & Regional Compliance

(1) Track institutional adaptations: Review disclosures from global banks (U.S., EU, Asia-Pacific) to assess alignment with frameworks like NIST AI RMF, EU AI Act, and

MAS FEAT Principles. Focus on transparency, bias mitigation, and accountability mechanisms.

(2) Evaluate regional regulatory divergence: Compare enforcement of AI rules (e.g., EU's ex-ante risk-based approach vs. U.S. sectoral guidelines vs. China's algorithmic oversight) to identify compliance challenges and harmonization opportunities.

(3) Engage with FS-ISAC: Leverage its global network to analyze anonymized threat intelligence and best practices for cross-border AI governance.

Assess Consumer Trust Dynamics & Hybrid Service Models

(1) Conduct longitudinal surveys: Measure shifts in consumer trust thresholds across regions, emphasizing AI transparency, data privacy, and preferences for human-AI interaction in high-stakes decisions (e.g., loans, financial advice).

(2) Map emerging communication practices: Analyze product contracts and disclosures for AI use, focusing on clarity of opt-out mechanisms, explainability standards, and incident accountability.

(3) Pilot hybrid AI-human case studies: Partner with institutions to test hybrid service models (e.g., AI-generated advice + human review) and evaluate their impact on customer satisfaction and trust.

Investigate Technical Safeguards & AI Limitations

(1) Audit privacy-preserving techniques: Evaluate adoption of differential privacy, federated learning, and encryption in LLMs to mitigate data leakage risks. Collaborate with cybersecurity teams to benchmark effectiveness.

(2) Address AI reliability gaps: Research solutions for hallucinations and explainability (e.g., XAI tools, synthetic data validation) through partnerships with academia (e.g., Columbia SIPA) and tech innovators.

(3) Monitor third-party AI vendors: Assess compliance with emerging voluntary certifications and alignment with FS-ISAC's AI security toolkit guidelines.

Analyze Workforce Evolution & Ethical AI Integration

(1) Track workforce reskilling: Study how institutions retrain employees for AI oversight roles (e.g., model validation, bias auditing) and measure shifts in job functions (e.g., reduced rote tasks, increased strategic roles).

(2) Benchmark ethical frameworks: Document how firms operationalize fairness checks (e.g., PPRE framework for bias auditing) and align AI ethics committees with regulatory expectations (e.g., CFPB's "no AI exemption" stance).

Strengthen Global Collaboration & Standard-Setting

(1) Participate in policy dialogues: Engage with BIS, OECD, and G20 working groups to advocate for interoperable AI standards and address systemic risks (e.g., herd behavior from shared AI models).

(2) Leverage FS-ISAC's neutrality: Develop global trust metrics and crisis response protocols for AI incidents (e.g., data breaches, biased outputs) to foster sector-wide resilience.

Future research should track consumer trust, analyze AI disclosures, and audit privacy-preserving techniques. FS-ISAC can develop trust metrics and share best practices to standardize transparency, accountability, and alignment of AI innovation with consumer trust.

Section 8: Mapping Friction Zones for AI Adoption

Having sketched a picture of how consumers of financial services products perceive the integration of AI therein, and having delineated how firms deploy AI in actuality, it merits considering the areas wherein firms have proven unable, unwilling, or slow to deploy AI. In such cases, the firm experiences friction, which falls into four broad categories: technological, regulatory, ethical, and organizational. Technological friction describes situations where the AI tool does not function as intended or without the full understanding of the engineers tasked with overseeing its use. Thus, such friction encapsulates instances where the AI tool did not complete its objective, whether as a result of a partial solution or a total malfunction, and instances where the tool provided an output for which the engineering team could not fully account. Organizational friction refers to internal misalignments that impede AI adoption. In many cases, firms lack a coherent strategy for AI integration, resulting in fragmented initiatives and uneven governance. Resistance to change further compounds this issue: employees, particularly within compliance or legacy IT divisions, often exhibit caution toward AI-enabled transformation. Additionally, suboptimal communication between technical teams and business units can obstruct effective collaboration, with the effect of delaying or derailing implementation efforts. Regulatory friction emerges where the legal architecture that governs AI use remains underdeveloped, contradictory, or too rigid to meet business needs. Firms might struggle to interpret how legacy financial regulations apply to novel AI tools, and might thus hesitate to deploy AI more broadly until regulators issue clearer guidance. This caution reflects not only legal risk, but also reputational sensitivity in an industry subject to intense scrutiny from government actors, consumer groups, and even activist investors. Ethical friction involves

normative concerns that arise in the design and deployment of AI systems. On this front, data privacy remains a dominant issue, especially when consumer information powers training datasets or informs automated decision-making. Likewise, opacity in AI outputs, particularly those driven by complex models, might raise concerns about transparency and explainability. As one executive interview subject noted, one's firm does not allow the implementation of a model unless the technical engineers who constructed it can understand its behavior inside and out. In those instances where firms cannot articulate how a model functions or defends its outcomes, both clients and regulators might find reasons to suspect the technology of posing greater risks than rewards, with the effect of slowing adoption still further. As seen across the interviews in the study, these frictions may overlap and compound one another. However, the most cross-cutting and effective approach to mitigating such frictions is to err on the side of caution and optimize the development of AI tools for maximum transparency and privacy, so as to avoid unforeseen challenges and undue regulatory scrutiny as well.

Section 9: Governance Models for Responsible AI in the Finance Sector

a. Experts have approached AI governance through different lenses, ranging from technical risk management to broader ethical and accountability considerations.

- Example 1 – Risk-Management Lens. Organizations such as NIST view governance as a systematic process of identifying, measuring, and mitigating model-specific risks, including bias, model drift, privacy concerns, and cybersecurity vulnerabilities. FS-ISAC's publications on adversarial AI, cyber defense, vendor evaluations, and acceptable-use policies represent this pragmatic stance.

- Example 2 – Ethics-and-Accountability Lens. Other experts focus on foundational values, fairness, transparency, accountability, privacy, and safety, as the moral core of any oversight regime. This perspective is evident in FS-ISAC's Responsible AI Principles, the Monetary Authority of Singapore's FEAT framework, and LinkedIn's PPRE method (privacy-preserving fairness audits for bias assessment). From this standpoint, adherence to ethical standards is critical for earning public trust and acceptance.

b. A broader view of technology governance highlights distinct regional models

- Anu Bradford's *Digital Empires* identifies three paradigms, each grounded in different relationships among markets, states, and individual rights:
    - United States: Market-driven (innovation first, ex-post enforcement)
    - European: Rights-driven (fundamental-rights pre-screening, ex-ante controls)
    - China – State-driven (developmental oversight aligned with strategic objectives)

c. Our findings align with many established governance frameworks and also nuanced differences drawn from consumer and executive interviews:

- **Conditional Consumer Trust in Transparency and Explainability:** While we broadly advocate transparency and explainability as essential components of consumer trust, our consumer interviews indicate that the need for detailed explanations varies by context. Consumers judged trustworthiness by service speed, clarity of error messages, and human fallback, not by published fairness charters, suggesting that ethics-by-design must be proven in day-to-day operations. They also expressed higher trust in financial institutions' AI deployments when transparency involved clear communication of AI use-cases and

data privacy controls. However, explainability emerged as important primarily when AI applications impacted high-stakes financial decisions, such as loan approvals or investment advice. Unlike the theoretical expert stance of universal transparency, our findings underline that consumer demand for detailed explainability is scenario-dependent, becoming acute primarily in financially sensitive contexts.

- **Balancing Innovation with Human Oversight:** Experts have highlighted extensive human oversight as a blanket requirement for responsible AI governance. However, our executive interviews revealed variations in practice that reflect regional and operational differences. Executives from the Asia-Pacific region prioritize speed and efficiency, embedding human oversight selectively at critical junctures, whereas U.S.-based institutions uniformly emphasized comprehensive human review as part of risk mitigation strategies. Our findings thus suggest that effective governance may require adaptable oversight mechanisms rather than uniform frameworks, challenging the notion of a single global standard advocated in some expert literature.
- **Regulatory Expectations and Institutional Practices:** While rigorous compliance frameworks are widely seen as central to responsible AI, in practice, institutions interpret and apply these frameworks in ways that reflect their perception of regulatory clarity and tolerance for risk. U.S. executives, facing a fragmented regulatory environment, tend to adopt a cautious, compliance-first mindset. In contrast, executives in the Asia-Pacific region describe more flexible, innovation-driven governance strategies aligned with national priorities. Although compliance remains a shared baseline, our interviews

suggest that real-world governance practices often diverge from purely compliance-centric models.

Bradford's three regulatory paradigms provide a useful lens. Building on this framework, our field research suggests the following implications for generative-AI deployment.

Model	Core Logic	Governance posture (GenAI)
US – Market-driven / ex-post	Prioritises innovation and competition; government intervenes mainly after harm occurs.	High freedom to experiment with generative AI, but the "wait-and-see" approach can lead to liability gaps. Institutions are responsible for demonstrating solid internal controls.
EU – Rights-driven / ex-ante	Anchors technology to fundamental rights (privacy, non-discrimination, explainability).	Clear, risk-tiered regulations can build consumer trust but increase upfront compliance costs. Strong documentation and human-oversight mandates complicate third-party model use.
China – State-driven / developmental	Uses AI to drive growth and maintain social stability under state supervision.	Rapid scaling of AI solutions is possible, provided they align with state guidelines. Transparency obligations focus on regulatory bodies rather than end-users.

d. Therefore, our research suggests there is no universal approach to AI governance in the financial services sector; rather, governance models should be tailored according to regional regulatory environments, institutional risk tolerance, and consumer expectations. Nonetheless, responsible AI governance should adhere to basic guidelines:

Duty	What must be disclosed / done	Where / to whom

1 Visibility	*Three-layer disclosure*: 1. *Consumer-facing notice* - Plain-language banner / chatbot splash with: ● "AI in use – tap for details" ● Purpose (e.g., credit scoring, fraud alert) ● Human escalation link 2. *Annual public "AI-Use & Risk" statement* (website + regulatory filing) 3. *Technical documentation* (model card, training-data lineage) lodged with primary supervisor, available to sector bodies	**Where**: (1) On-screen at the point of decision (mobile app, chatbot, web portal, account statement) (2) In adverse-action notices (e.g., loan denials, fraud holds) (3) In the public privacy policy and model-risk summary **To Whom**: (a) Affected consumers – so they understand that AI contributed to the decision and may request human review. (b) National or state supervisors – via existing model-risk portals for supervisory audits.
2 Verification	*Continuously monitor* production models: real-time drift alarms, quarterly bias dashboards, and incident logs submitted within 72 hours of a material error.	**Where:** (1) Internal risk-analytics dashboards and automated alerting systems; (2) material-incident reports and audit files submitted via regulator incident-reporting portals within prescribed timelines (e.g., 72 h under CPS 234 or DORA) . **To Whom:** (a)Internal model-risk & compliance teams (first recipients) (b) Financial-sector supervisors and regulators when thresholds are breached.
3 Containment	*Safeguard data & outputs*: Ensure privacy-preserving computation (e.g., encryption, secure enclaves) for sensitive financial or personal information, and require third-party providers to meet the same security, testing (e.g., red-teaming), and privacy standards as in-house systems.	**Where**: (1) Internal deployments and private-cloud environments (2) Mandatory clauses in third-party contracts; vendors subject to the same audits and controls **To whom**: (a) All internal model owners (b) External model/analytics vendors & their regulators

E. We anticipate that FS-ISAC can leverage our findings to expand its existing frameworks and guidelines, incorporating explicit provisions for procedural transparency, user-centric disclosures, and human oversight mechanisms. In doing so, FS-ISAC can reinforce trust in the financial sector's adoption of generative AI, balancing innovation opportunities with robust governance standards that protect consumers and promote systemic resilience.

Part 4 - Conclusions

Section 10: Recommendations for Institutions

The decisive challenge for the next wave of GenAI deployment in finance is no longer algorithmic horsepower but human confidence. Banks and insurers that can translate ethical principles into everyday product experiences will capture the trust premium; those that cannot will confront rising regulatory capital and reputational drag. What follows is a seven-pillar roadmap for turning "trust" from a marketing slogan into a measurable, lived reality. These recommendations outline a strategic framework to support trustworthy and scalable deployment of GenAI in financial services.

1 Data Integrity.

GenAI advice is only as good as the information that shapes it. Institutions should hard-wire continuous data-quality audits, anomaly detection, and tamper-evident provenance logs into their ML pipelines, backed by independent attestations. Such controls mirror long-standing model-risk practices in quantitative trading desks and give customers defensible assurance that their financial profile is neither corrupted nor spoofed.

2 Methodological Transparency.

Consumers will not trust what they cannot inspect. Publishing concise *model cards* that describe training data, intended use-cases, performance bounds, and known failure modes, an approach endorsed in the NIST AI Risk-Management Framework, lets users and regulators understand "what's under the hood." Embedding plain-language "Why this answer?" pop-ups at decision time further demystifies GenAI outputs and reduces automation surprise.

3 Predictability & Consistency.

Advice and decisions must behave the same way today and tomorrow. Scenario stress-tests, variance-threshold alarms, and scheduled re-training windows with pre-release regression tests help spot drift before it harms customers. Publishing aggregate stability metrics (e.g., month-to-month answer variance) in annual stewardship reports turns internal QA hygiene into an external trust signal.

4 Human-Centred Design & Guided Information Access.

GenAI should *augment*, not replace, relationship banking. Every interface that can materially affect a customer's finances should surface a 24/7 human-escalation channel and step-by-step recourse instructions. An in-app *Trust Dashboard*, linking audit summaries, privacy policies, and regulator trustmarks, plus a QR code to a consumer explainer microsite, meets users where they are and counters the "black-box" perception.

5 Privacy-by-Design.

Because GenAI systems often ingest highly granular behavioural data, privacy cannot be bolted on. Differential privacy, federated learning, and secure multi-party computation should be the default in high-sensitivity contexts such as biometric authentication or behavioural scoring.

These techniques allow fairness testing and model improvement without leaking personal attributes, reinforcing stewardship and minimisation commitments.

6 Safe Innovation through Regulated Sandboxes.

To balance speed and safety, firms should channel new GenAI use-cases through supervisory sandboxes that permit live testing under controlled conditions. The UK Financial Conduct Authority's Digital Sandbox and Singapore's Generative AI Evaluation Sandbox provide templates: synthetic or masked data, opt-in participants, rollback plans, and public post-mortems that strengthen collective learning. Article 53 of the draft EU AI Act is set to codify similar sandbox privileges across Member States.

7 Strategic Vision & Sector Alignment.

Finally, fragmented controls cannot manage systemic risk. Boards should map every GenAI asset to a *sector-wide* risk typology that blends the EU's tiered "high-risk" designations with NIST's AI-RMF categories, then disclose progress annually. A standardised model-registry protocol, capturing architecture, data lineage, red-team results, and post-deployment drift, turns governance paperwork into a living memory system that auditors and regulators can interrogate. When incidents occur, anonymised failure modes should flow to industry collectives such as FS-ISAC, shortening feedback loops and enhancing sector resilience.

Trust is a product feature, not a compliance checkbox. Firms that embed the seven pillars above, visible in user journeys, measurable in KPIs, and anchored in cross-border standards, will convert GenAI from a regulatory headache into a durable competitive moat.

Section 11: Operationalizing Consumer Trust

As noted in several executive interviews, consumer trust is the cornerstone of financial services. The integration of Generative Artificial Intelligence (GenAI) introduces new complexities in maintaining confidence. Operationalizing trust—embedding it into consumer experiences through transparent frameworks, accessible tools, and clear communication—is critical to ensuring responsible AI adoption. This section outlines strategies for financial institutions to translate trust from an abstract principle into a measurable, consumer-facing outcome, with FS-ISAC leading the standardization of trust-building practices across the sector.

Consumer trust in GenAI is conditional, rooted in institutional reputation rather than regulatory oversight, as evidenced by interviews with U.S. and Asia-Pacific consumers. Users accept AI for transactional tasks like fraud alerts but demand human oversight for high-stakes decisions such as loan approvals or financial advice. Concerns about data privacy and opaque AI deployment underscore a critical gap: consumers seek visibility into AI's role and assurances of accountability. These expectations align with global regulatory trends that emphasize transparency, fairness, and human-in-the-loop protocols, yet many institutions' governance efforts remain invisible to end-users, risking misalignment with consumer needs.

To operationalize trust, institutions must embed it into both AI systems,from the development to the deployment phases, and consumer interactions. First, **transparent frameworks** should draw on established practices like Model Risk Management (MRM) and ethical guidelines such as Singapore's FEAT Principles. MRM's standards for data integrity and model validation ensure reliable AI outputs, while FEAT's focus on fairness and accountability addresses consumer concerns about bias and recourse. Implementing model registrie can enhance

traceability and predictability, reassuring consumers that AI decisions are auditable. Second, **consumer-facing tools** bridge the visibility gap. Plain-language disclosures, trust dashboards, and in-app explainers can demystify AI use, clarify data practices and offer human escalation options. Such tools align with consumer demands for control and align with regulatory calls for explainability. Third, **privacy-preserving technologies**, such as differential privacy and federated learning, serve to safeguard sensitive data and thus reinforce trust in AI's security. Finally, **consumer education** through accessible microsites or QR-coded explainers empowers users to understand AI's role, transforming trust from an assumption into an informed choice.

FS-ISAC can leverage its global membership and cybersecurity expertise to lead this transformation. By developing sector-wide trust metrics—standardized measures of transparency, fairness, and human oversight—FS-ISAC can serve as a trusted authority, guiding institutions to align governance with consumer expectations. Collaborative initiatives, such as shared model registries or crisis response protocols, can mitigate systemic risks like model concentration, fostering resilience. Through regulatory sandboxes, FS-ISAC can facilitate controlled innovation, balancing agility with accountability.

Financial institutions must not see the operationalization of consumer trust as a compliance exercise but rather as a strategic imperative. Institutions that embed trust into AI design and communication will secure a competitive advantage, especially among privacy-minded consumers. Meanwhile, FS-ISAC's leadership can entrench trust as a sector-wide standard, ensuring that AI enhances, rather than erodes, consumer confidence. The obstacles to achieving this do not result from inadequate technological capabilities; rather, they flow from a persistent underestimation of consumer demand for responsible AI governance.

APPENDICES

Appendix A: Consumer Interviewee Demographic Table

ID	Country	Fin-Service Use	AI-Awareness	Trust	Key Trust Factors	Human ↔ AI Preference	Data-Privacy Stance
C-01	China	Frequent	Intermediate	8 / 10	Deposit insurance; bank credibility	Human for fraud; AI fine for routine	OK with anonymised, non-sensitive data
C-02	USA	Frequent	Low	8 / 10	Branded bank app; perceived safety	Human for everything but basic FAQs	Wary of LLM data harvesting
C-03	China	Frequent	Intermediate	6–7 / 10	Encryption; MFA; face-ID	AI for look-ups; human for investment	Consent first; fears misuse
C-04	USA	Frequent	High	6 / 10	Convenience; uptime	Strong human preference; AI "reference only"	Opposes sharing trading data
C-05	USA (Chinese-Am.)	Frequent	High	10 / 10	Big brands; strong cyber-policies	AI welcome; human only for judgement cases	Fine with chat-history, not holdings/PII
C-06	USA (Korean)	Basic	Low	10 / 10	Reputation; clean UI; Zelle	Wants human in all cases	"Don't care"; no privacy worry
C-07	China	Frequent	Intermediate	9 / 10	Security; transparency	AI for analysis; human final say	Consent & anonymisation required
C-08	China	Basic	Low	7–8 / 10	MFA; global-bank reputation	Human for most tasks; AI for search	OK with bank-owned AI, not open LLMs
C-09	USA (Vietnamese)	Frequent	Intermediate	8 / 10	2FA; FDIC; support	Human for complex / urgent	Strong privacy concern; wants audit trails
C-10	USA	Basic	Low	10 / 10	2FA; transfer confirmation	Human for nearly everything	Wary; wants purpose-limited

							use
C-11	USA (Korean)	Frequent	Intermediate	8 / 10	Fraud alerts; customer service	Human for big/urgent issues	Opposes training on personal data w/o consent
C-12	USA (Korean)	Basic	Intermediate	8 / 10	Cyber-security history	Human for co specifics; AI draft OK	Chat history fine; PII not
C-13	China	Frequent	Intermediate	10 / 10	Big-bank stability	OK with replacing humans eventually	"Concerned but can't control"
C-14	China	Frequent	Intermediate	10 / 10	Brand credit; encryption	Human for complex; AI cost-cutting fine	Training use OK if no PII
C-15	China	Frequent	Intermediate	10 / 10	Bank scale; clean record	Human for complex, urgent	Only non-PII via AI; bank-controlled LLM

Appendix B: Executive Interviewee Demographic Table

ID	Institution Type	Country (HQ)	Sectoral Focus	Role / Title
EX-01	Risk- & compliance-consultancy	China	Integrity-risk, compliance, AI	Analyst
EX-02	Audit / tax consulting	Singapore	Tax advisory	Partner, Head of Tax
EX-03	AI / tech start-up	China	AI products	CEO
EX-04	Law firm	China	Legal services (law)	Partner
EX-05	Investment bank	USA (global)	Investment banking (cross-border, IB)	Executive Director
EX-06	Tech / platform	USA	Information-security, platform trust	Distinguished Engineer, InfoSec
EX-07	Investment bank / advisory	USA	Healthcare IB	Vice President
EX-08	Health-tech / virtual clinic	USA	Security & compliance	Head of Security & Compliance
EX-09	Investment bank and financial services	USA	Global cyber and information security and risk management	MD & Chief Information Security Officer
EX-10		USA	Capital markets,	Legal Entity Controller

			legal-entity control	
EX-11	Independent / industry expert	USA		

Appendix C: Detailed Overview of Regulatory Landscape Across the U.S., EU, UK, and Asia-Pacific

1. United States

The United States has not enacted a single comprehensive AI law comparable to the European Union's approach. Instead, it has relied on a combination of executive actions, agency guidance, and existing statutes to oversee AI, particularly within the financial sector. In theory, the Congress passes laws that regulators apply according to their specialized expertise, with the President weighing in to defend critical interests.

Such a demarcation of bounds mirrors the separation of powers that undergirds the American legal system, with Congress representing the legislative branch and bearing accountability to the American public, the regulators representing the judicial branch and bearing accountability to the federal judiciary, and the President representing the executive branch and, as delineated in American legal doctrine, defending the Constitution. In practice, the representatives of these three branches might act outside of this cycle of continual evaluation in areas where powers are not clearly defined and delimited, as in the case of breakthrough technologies or geopolitical exigencies.

As of April 2025, however, the development of AI regulation followed market forces and technological advancement, which are outpacing regulatory frameworks rather than following the traditional legal cycle of legislative enactment followed by executive enforcement and judicial oversight. The geopolitical importance of American firms in the AI value chain, as well as the American firms that produce the inputs necessary for AI development, such as chips, has handed the initiative to the private sector.

1.1 Evolving Federal AI Governance

After 2022, U.S. federal governance concerning AI has evolved through new frameworks and executive actions. The AI Bill of Rights Blueprint (Oct 2022), issued by the White House Office of Science and Technology Policy, serves as a voluntary framework highlighting five fundamental principles: safe and effective systems, freedom from prejudice, data privacy, notice and explanation, and human alternatives. Though non-binding, it sets expectations relevant to AI in financial services, such as ensuring credit decisions remain free of algorithmic bias and are explainable to consumers (The White House, 2022).

In the realm of artificial intelligence, the NIST AI Risk Management Framework (AI RMF 1.0), published by the National Institute of Standards and Technology in January 2023, defines "trustworthy AI" through seven characteristics: validity and reliability, safety, security and resilience, accountability, transparency, explainability, privacy enhancement, and mitigation of harmful bias (Maple et al, 2024; NIST, 2024). This voluntary framework supports organizations in managing AI-related risks by embedding trustworthiness in AI system development and use. Financial regulators broadly endorse this guidance. By July 2024, NIST released a Generative AI

Profile outlining specific risk-management considerations for large language models, including content controls and testing for misinformation (NIST, 2024).

In October 2023, President Biden issued Executive Order 14110 – "Safe, Secure, and Trustworthy Development and Use of AI," directing a whole-of-government approach to AI risks (Federal Register, 2023). EO 14110 prioritizes civil rights, prevention of AI-driven harm, and support for public trust in AI. It calls for federal agencies to develop safety standards (including red-teaming for foundation models), mitigate algorithmic discrimination in areas such as lending, and ensure AI transparency and accountability. The Order shows the need to protect consumers in financial services from fraud and discrimination and relies on existing statutes (e.g., fair lending laws) to ensure that "hard-won consumer protections are more important than ever" in the AI era (Federal Register, 2023). It also references and builds upon the voluntary NIST AI RMF, urging financial regulators to consider new guidance or regulations consistent with these principles (Harris & Jaikaran, 2024).

A significant shift occurred in January 2025, when President Trump issued Executive Order 14179 – "Removing Barriers to American Leadership in Artificial Intelligence." This EO explicitly revoked numerous AI policies from the prior administration, including EO 14110, citing concerns over "overregulation" and declaring a need to "clear a path for American AI innovation and supremacy" (The White House, 2025). It maintains that U.S. AI development should advance national competitiveness and remain free from "ideological bias or engineered social agendas," thereby emphasizing a market-driven approach (The White House, 2025). While it loosens federal-level guidance, prior legislation and agency oversight remain in force.

1.2 Agency and State-Level Initiatives

Beyond NIST, sector-specific regulators have also shaped the U.S. AI landscape. Banking agencies (Federal Reserve, CFPB, FDIC, OCC) issued a Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence in 2021 (Federal Reserve Board, 2021). FINRA (Financial Industry Regulatory Authority) released a 2020 report on AI in securities, focusing on transparency around investment advice (FINRA, 2020). The CFPB has reminded lenders that AI models must still comply with fair lending laws (e.g., Equal Credit Opportunity Act), a stance reinforced in April 2023 via a Joint Statement with the FTC, DOJ, and EEOC confirming there is "no AI exemption" under existing civil rights statutes. These actions emphasize regulators' readiness to penalize firms for discriminatory or deceptive AI-driven practices, reflecting a broader goal of sustaining public trust (Khan, 2023).

States have begun to enact their own AI governance rules. Colorado's AI Act (May 2024), for instance, applies to "high-risk" AI, which encompasses applications such as credit, housing, insurance, and employment. It obliges annual impact assessments on algorithmic discrimination risks, demands transparency when users interact with AI systems, and grants enforcement authority to the state Attorney General (Anderson et al., 2024; Colorado General Assembly, 2024). This aligns with the EU AI Act and certain federal guidelines, directly tackling public trust by preventing discriminatory outcomes and preserving user agency.

1.3 Impact on Consumer Trust and Accountability

Federal authorities are also increasingly attentive to how AI tools affect consumer trust, transparency, and accountability in finance. A notable example is the CFPB's Chatbots in

Consumer Finance report (June 6, 2023), which explores the use of AI-powered chatbots (including generative models) by major banks and examines risks to consumers. According to the CFPB, every one of the top 10 U.S. banks now employs chatbots for customer service, with estimates suggesting that over one-third of Americans interacted with a bank's chatbot in 2022 (CFPB, 2023). As financial institutions shift from simple scripted bots to advanced, large language model (LLM)-driven systems, the CFPB identifies multiple trust-related issues that, if unaddressed, could undermine consumer confidence. For example, the CFPB warns that next-generation, large language model (LLM) chatbots may offer inaccurate "hallucinations" that can misinform customers, potentially causing financial harm if users follow flawed guidance potentially causing financial harm if users follow flawed guidance. Privacy and data security risks arise when chatbots store or process sensitive information, and generative models might inadvertently disclose personal data seen during training (CFPB, 2023). Moreover, overreliance on AI could erode the human-centered service many consumers expect, particularly if customers feel trapped in chatbot loops that fail to solve complex issues (CFPB, 2023). Thus, while large banks see operational benefits, the CFPB stresses the need for robust oversight, privacy safeguards, and human backup.

1.4 Foundational Statutes Relevant to AI

The Gramm-Leach-Bliley Act (GLBA, 1999) plays a foundational role in governing data privacy and security in the U.S. financial system, with clear implications for AI applications. Section 501(a) of GLBA affirms that "each financial institution has an affirmative and continuing obligation to respect the privacy of its customers and to protect the security and confidentiality of those customers' nonpublic personal information" (15 USC 6801: Protection of Nonpublic Personal Information, 1999). This obligation underpins much of the FS-ISAC's work on threat intelligence

sharing and drives institutional investment in cybersecurity and information governance. In the AI context, GLBA's requirements imply that AI systems which process, store, or generate insights from customer data must implement rigorous safeguards, including training data selection and model output audits, to preserve confidentiality and trust. Thus GLBA forms an essential legal baseline for secure and privacy-conscious AI deployment across U.S. financial institutions.

Meanwhile, the Dodd–Frank Act, though broader in scope, includes provisions to enhance financial system stability and consumer protection through improved oversight and transparency, reinforcing the trust needed for digital innovation in finance (Text - H.R.4173 - 111th Congress (2009-2010)). Both laws reinforce trust in financial markets, albeit more broadly than AI-focused measures. Taken together, they support a regulatory environment in which AI solutions in finance must respect consumer privacy, protect data, and avoid destabilizing practices.

1.5 Cybersecurity Regulations in the United States

Note: The following cybersecurity regulations are essential for protecting the data and infrastructure that AI systems rely upon. However, cybersecurity risks do not always overlap directly with AI misuse. Where AI systems face threats such as data leakage, subverting of systems, or malicious manipulation, these laws and guidelines can indirectly apply, but they were not designed specifically for Generative AI regulation.

Before 2022, foundational laws such as the Computer Fraud and Abuse Act (CFAA, 1986) were pivotal in addressing cyber intrusions. The CFAA imposes penalties for unauthorized access to protected computers, setting strict boundaries to uphold trust in computer systems (H.R.4718 - 99th Congress (1985-1986): Computer Fraud and Abuse Act of 1986, 1986).

Specifically, 18 U.S.C. § 1030 prohibits “access[ing] a computer without authorization or exceed[ing] authorized access” to obtain information. By criminalizing hacking and insider misuse, the CFAA aims to deter breaches that undermine consumer trust in financial systems.

The Cybersecurity Information Sharing Act (CISA) facilitates voluntary information-sharing between private entities (including financial institutions) and government agencies (CISA, 2016). The law encourages a trusted environment for threat data exchange while shielding participating organizations from legal repercussions. As described in the legislation, CISA promotes cybersecurity through “enhanced sharing of information about cybersecurity threats.”(Text - S.754 - 114th Congress (2015-2016)).

NYDFS Cybersecurity Regulation (23 NYCRR Part 500) Introduced in 2017 (amended 2023) under the New York Department of Financial Services (DFS), this rule mandates robust cybersecurity programs for banks, insurers, and other covered financial institutions, requiring governance, risk assessments, incident reporting, and board-level accountability (NYDFS, 2017). It states, “Senior management must take this issue seriously and be responsible for the organization’s cybersecurity program.… A regulated entity’s cybersecurity program must ensure the safety and soundness of the institution and protect its customers.” New York’s approach encapsulates the value of consumer protection which, in large part, defines the American financial services sector.

The FFIEC IT Examination Handbook – Information Security (June 2015) offers interagency guidance for banks on implementing effective information security programs. It details a life-cycle

approach including risk identification, measurement, mitigation, and monitoring, while emphasizing the significance of information-sharing communities, such as FS–ISAC,to enhance trust through collective defense and transparency. According to the handbook, “Participation in an information-sharing forum, such as FS–ISAC, should be a component of the risk identification process because sharing information may help the institution identify and evaluate relevant cybersecurity threats” (FFIEC, 2015).

2. European Union

The European Union has passed an extensive set of regulations interlinking cybersecurity and operational trust, mirrored prominently in financial services. The Digital Operational Resilience Act (DORA, 2022), which entered into force on 16 January 2023 and becomes applicable on 17 January 2025, focuses on enhancing cybersecurity and operational resilience within the financial sector (European Union, 2022b). It mandates that banks, insurers, asset managers, and other financial entities manage and report ICT risks, conduct resilience testing, and supervise third-party ICT service providers. Article 45, in particular, permits and encourages cyber threat intelligence sharing within “trusted communities,” thereby institutionalizing trust-based cooperation among financial institutions.

In parallel, the EU Artificial Intelligence Act (AI Act), proposed in 2021, finalized in late 2023, and in force as of August 2024, has become the world’s first broad AI law (Regulation (EU) 2024/1689) that explicitly seeks to foster “trustworthy, human-centric AI” (European Commission, 2025). Adopting a risk-based framework, the AI Act bans certain “unacceptable-risk” AI applications

(e.g., social scoring) and imposes stringent requirements on “high-risk” AI, including credit scoring and algorithmic trading. These mandates address data quality, transparency, human oversight, accuracy, and cybersecurity (Council of the EU, 2023). Generative AI systems must disclose AI-generated content, preventing deception and reinforcing public confidence in AI-driven processes (European Commission, 2024; EU Artificial Intelligence Act, 2024c). Moreover, providers of large language models or other general-purpose AI are subject to baseline safety and accountability obligations (EU Artificial Intelligence Act, 2024b). Notably, AI used in creditworthiness evaluation is deemed high-risk, obligating banks or fintechs in the EU to comply with stringent fair and transparent lending protocols once the act is fully applicable (EU Artificial Intelligence Act, 2024a; Moody’s, 2024).

Other EU regulations augment this trust-centric approach. The NIS2 Directive (2022) extends EU cybersecurity directives to a wider set of “essential” and “important” entities, including those handling GenAI systems, while the General Data Protection Regulation (GDPR, 2016) consistently enforces personal data protection to build trust in digital ecosystems (Recital 7 - The Framework Is Based on Control and Certainty, 2016). For generative AI, GDPR considerations emerge around lawful training, data usage, consent, data minimization, and compliance with “the right to be forgotten.” In 2023, Italy’s data protection authority temporarily banned ChatGPT, prompting clarifications on data handling and user opt-out mechanisms (Mukherjee & Vagnoni, 2023). The Payment Services Directive 2 (PSD2, 2015) promotes Strong Customer Authentication and open banking, bolstering consumer safeguards in digital payments (Directive - 2015/2366 - EN - Payment Services Directive - EUR-Lex, 2015). Meanwhile, the EBA Guidelines on ICT and

Security Risk Management (2019) guide banks and payment service providers on governance, risk assessments, and effective cyber risk controls (European Banking Authority, 2025).

As of April 2025, several months into President Trump's pivot toward protectionism in the United States, the European Union has signalled its intent to revisit some of its marquee digital regulations, such as the GDPR. A Politico report describes European Commission President Ursula von der Leyen as looking toward AI as a means of increasing the competitiveness of the EU tech sector and narrowing the innovation gap with the United States (Haeck, 2025). The push to streamline regulation stems from the need to attract investment in European AI companies and facilitate compliance, with a focus on reducing reporting obligations and their associated costs to companies.

3. United Kingdom

The UK's trajectory toward AI governance remains sector-focused, grounded in principle-based regulation. Its National AI Strategy (Sep 2021) outlines a decade-long plan to position the UK as a leader in AI innovation and governance, complementing safety and ethical considerations (Sherman, 2025).In March 2023, the Sunak government released the "A Pro-Innovation Approach to AI Regulation" White Paper, opting against a single overarching AI law; instead, it grants existing regulators in each sector the authority to develop guidance under five cross-cutting principles: safety, security, transparency, fairness, accountability, and contestability and redress(Department for Science, Innovation Technology & Office for Artificial Intelligence, 2023). This flexible model aims to prevent stifling AI growth through over-regulation.

In January 2025, the Department for Science, Innovation and Technology (DSIT) published the AI Opportunities Action Plan, which includes 50 proposals to leverage AI for economic benefits while emphasizing that public trust underpins successful adoption (Department for Science, Innovation Technology, 2025). Throughout these developments, UK regulators have repeatedly stressed that existing consumer protection, equality, and data protection laws apply equally to AI, affirming "no AI exemption" from statutory obligations (Bank of England, 2022). The Financial Conduct Authority (FCA) in particular focuses on "safe and responsible adoption," ensuring AI decisions do not adversely impact consumers. Innovations such as "digital sandboxes" allow AI-driven fintech to be tested in controlled settings (Artificial Intelligence (AI) Update – Further to the Government's Response to the AI White Paper, 2024). Hence, the UK's sector-led, principles-based system continues to guide AI innovation in finance, emphasizing accountability, transparency, and cybersecurity resilience.

4. Asia-Pacific

4.1 Singapore

In Singapore, the Monetary Authority of Singapore (MAS) has taken a leading role as the central bank and financial regulator, aiming to maintain high levels of trust in a rapidly evolving fintech landscape. The MAS Technology Risk Management Guidelines (2021) are influential, urging financial institutions to maintain strong cyber resilience, regular risk assessments, and comprehensive IT governance, despite not being legally binding. MAS has emphasized, "Strong

cyber resilience is critical for sustaining trust and confidence in financial services," emphasizing a defense-in-depth mindset (Monetary Authority of Singapore, 2021).

Complementarily, the FEAT Principles (Fairness, Ethics, Accountability, Transparency, 2018) guide responsible AI use in banking, insurance, and other financial domains by stressing fairness, unbiased outcomes, clarity in AI-driven decisions, and firm-level accountability (Bamberger et al., 2018; Monetary Authority of Singapore, 2018b). MAS integrates these principles into supervisory reviews, indicating the importance of ethical governance for public confidence (Monetary Authority of Singapore, 2018a).

4.2 China

In China, the regulatory environment addresses algorithms and generative AI with increasing specificity. The Algorithmic Recommendation Regulation (2022) mandates that service providers, including fintech apps, disclose the existence of recommendation algorithms, prevent manipulative or discriminatory practices, and register significant algorithms with the Cyberspace Administration of China (CAC) (Bird & Bird, 2023). The Deep Synthesis (Deepfake) Regulation (2023) mandates transparent labeling of AI-altered media, including synthetic voices and videos, guarding against fraud. Similarly, the Interim Measures for Generative AI Services (August 2023) apply to publicly accessible generative AI tools, enforcing content oversight, non-discrimination, transparency, and accurate output (Cyberspace Administration of China, 2023; Lai, 2023; Zhang, 2023). Article 4(4) highlights that generated content "shall be true and accurate," while providers must mitigate false information and discriminatory outputs (Zhang, 2023). These rules also reinforce the state's

authority to examine algorithms and training data. Proposed additional labeling requirements by 2025 indicate China's intent to systematically address deepfake risks (Quinn, 2025).

4.3 Japan

In Japan, the Act on the Protection of Personal Information (APPI), amended in 2020 and fully effective 2022, requires financial firms to implement security measures for data protection (Act on the Protection of Personal Information - English, 2003). Though not AI-specific, the APPI serves as a key legal foundation for trusted digital services.

Separately, the Japan Financial Services Agency (JFSA) contributed to the discourse with its 2025 "Discussion Paper on AI Utilization in Financial Services" (Financial Services Agency, 2025). While not statutory, this paper outlines both opportunities (efficiency gains, improved customer service) and potential pitfalls (fraud, bias, misinformation), encouraging industry feedback on AI governance. It shows that responsible adoption must encompass transparency, accountability, and safety to sustain public trust. The paper encourages industry dialogue and governance reforms and highlights that responsible adoption must encompass transparency, accountability, and safety to sustain public trust.

4.4 Australia

In Australia, the APRA Prudential Standard CPS 234 (2019) mandates that regulated entities maintain robust information security systems and notify the regulator of significant incidents within 72 hours. The regulation places direct accountability on senior management and seeks to ensure

institutions can withstand cyberattacks, thereby enhancing systemic resilience. The government has stressed using preexisting legal frameworks for AI to overcome the constraints of existing laws and regulations, but critics warn it may put Australia “at the back of the pack” without AI-specific legislation. In August 2024, the Australian Department of Industry, Science, and Resources announced the Voluntary AI Safety Standard, and in September, the Digital Transformation Agency released its strategy for responsible AI usage in government(Australian Signals Directorate’s Australian Cyber Security Centre, 2024). Such developments suggest an evolving approach where voluntary standards and existing regulations form the initial governance layer for AI until more targeted legislative measures are introduced.

Appendix D: Technical Approaches to Privacy Protection & Hallucination Control

Data Anonymization and Sanitization

A foundational strategy is removing personal identifiers from training corpora via automated detection and masking of names, emails, and other sensitive data (Feretzakis et al., 2024). Techniques like k-anonymity or pseudonymization reduce the likelihood that an LLM inadvertently reveals private details. However, perfect anonymization is theoretically unattainable, as attackers can potentially re-identify individuals when combined with outside data (Feretzakis et al., 2024).

Consequently, organizations typically employ data minimization principles, keeping only the necessary info and sanitizing PII as far as possible, to lessen risks.

Differential Privacy (DP)

DP provides a quantifiable safeguard that obscures individual training examples, often through DP-SGD (stochastic gradient descent with noise addition). This helps prevent an adversary from concluding whether a person's data was included. While DP is beneficial for mitigating memorized text leakage, it can degrade model accuracy or fluency (Bu & Wang, 2023). Researchers continue to refine DP methods to balance utility and privacy, including applying DP selectively to certain model parameters or at the decoding stage (Zhang et al., 2024).

Federated Learning

By training models locally on siloed data and exchanging only aggregated updates, federated learning avoids centralizing sensitive datasets (Feretzakis et al., 2024). This is especially relevant in domains with stringent data-sharing restrictions, such as healthcare or finance. Security can be further bolstered through secure aggregation or local differential privacy, although federated training for large LLMs remains computationally intensive.

Privacy-Preserving Fine-Tuning and Unlearning

When adapting an LLM to domain-specific or user data, approaches like differential privacy can be applied only to a subset of model weights (e.g., bias-only fine-tuning) to limit information leakage while preserving performance (Bu & Wang, 2023). Another strategy involves using synthetic data, generated by smaller, DP-trained models, for alignment tasks, sparing sensitive real data from

direct usage (Yu et al., 2024). Post-hoc unlearning techniques aim to remove or obscure memorized content if training data was included in error or if an individual invokes a legal right to be forgotten (Zhang et al., 2024). By combining these techniques, organizations can align LLMs to new tasks without compromising data privacy or trust.

Mitigating Hallucinations

Increasing hallucination (fabricated or inaccurate content) in LLMs can jeopardize reputation and safety by misleading users, making error reduction a crucial priority. Researchers are testing fixes such as chain-of-thought prompting with human feedback and active learning. Cohn et al. (2025) demonstrate that their “CoTAL” approach, combining chain-of-thought reasoning with iterative human-in-the-loop prompt engineering, can help large language models (LLMs) align more reliably with user-defined scoring rubrics and explanations in educational settings. Beyond education, organizations often utilize knowledge grounding (e.g., retrieval-augmented generation) to anchor model outputs in trusted data repositories, thereby reducing the model’s tendency to generate unsupported statements.

Bibliography

15 USC 6801: Protection of nonpublic personal information. (1999). Uscode.House.Gov. https://uscode.house.gov/view.xhtml?req=granuleid%3AUSC-prelim-title15-section6801&edition=prelim

Abdullah, I. N. (2025, February 12). An overview of AI regulations in financial services around the world. *Fintech Singapore*. https://fintechnews.sg/107530/ai/ai-regulations-asia-financial-services/

Act on the protection of personal information - English. (2003). Japanese Law Translation. https://www.japaneselawtranslation.go.jp/en/laws/view/4241/en

Adria, T. (2024, September 6). *Artificial Intelligence and its Impact on Financial Markets and Financial Stability*. IMF. https://www.imf.org/en/News/Articles/2024/09/06/sp090624-artificial-intelligence-and-its-impact-on-financial-markets-and-financial-stability

AI Working Group A&O Shearman. (2024, September 23). *Zooming in on AI - #5: AI under financial regulations in the U.S., EU and U.K. – a comparative assessment of the current state of play: Part 1*. A&O Shearman. https://www.aoshearman.com/en/insights/ao-shearman-on-tech/zooming-in-5-ai-under-financial-regulations-in-the-us-eu-and-uk-a-comparative-assessment-part-1

Aldasoro, I., Armantier, O., Doerr, S., Gambacorta, L., & Oliviero, T. (2024, April 23). Survey evidence on gen AI and households: Job prospects amid trust concerns. BIS. https://www.bis.org/publ/bisbull86.htm

Aldasoro, I., Gambacorta, L., Korinek, A., Shreeti, V., & Stein, M. (2024, June 13). Intelligent financial system: How AI is transforming finance. BIS. https://www.bis.org/publ/work1194.htm

Anderson, H., Nunes, I., & Oltean, J. (2024, June). White & Case LLP International Law Firm, global law practice. White & Case LLP. https://www.whitecase.com/insight-alert/newly-passed-colorado-ai-act-will-impose-obligations-d

evelopers-and-deployers-high?s=Newly%20passed%20Colorado%20AI%20Act%20will%20impose%20obligations%20on%20developers%20and%20deployers%20of%20high-risk%20AI%20systems

Anthony, A., Sharma, L., & Noor, E. (2024, April 30). Advancing a more global agenda for trustworthy artificial intelligence. *Carnegie Endowment for International Peace*. https://carnegieendowment.org/research/2024/04/advancing-a-more-global-agenda-for-trustworthy-artificial-intelligence?lang=en

APRA. (2019, July 1). Prudential Standard CPS 234 Information Security. APRA. https://www.apra.gov.au/sites/default/files/cps_234_july_2019_for_public_release.pdf

Artificial Intelligence (AI) update – further to the Government's response to the AI White Paper. (2024, April 19). FCA. https://www.fca.org.uk/publications/corporate-documents/artificial-intelligence-ai-update-further-governments-response-ai-white-paper

Australian Signals Directorate's Australian Cyber Security Centre. (2024, January 24). Engaging with artificial intelligence. Cyber.Gov.Au. https://www.cyber.gov.au/resources-business-and-government/governance-and-user-education/artificial-intelligence/engaging-with-artificial-intelligence

Badrinarayanan, S., Osoba, O., Cheng, M., Rogers, R., Jain, S., Tandra, R., & Pillai, N. S. (2024, September 6). Privacy-Preserving race/ethnicity estimation for algorithmic bias measurement in the U.S. arXiv.Org. https://arxiv.org/abs/2409.04652

Bamberger, N., Mahoney, C., & Whit, J. (2018, November 20). Meet FEAT: Singapore's new AI and data analytics principles for the financial sector. Cleary FinTech Update. https://www.clearyfintechupdate.com/2018/11/meet-feat-singapores-new-ai-data-analytics-principles-financial-sector/

Bank of England. (2022, October 11). DP5/22 - Artificial intelligence and machine learning. Bank of England. https://www.bankofengland.co.uk/prudential-regulation/publication/2022/october/artificial-intelligence

Bird & Bird. (2023). China: Data and evolving digital regulation: Algorithm regulation. Bird & Bird. https://www.twobirds.com/en/capabilities/practices/digital-rights-and-assets/apac-dra/apac-dsd/data-as-a-key-digital-asset/china/data-and-evolving-digital-regulation-algorithm-regulation

Bradford, A. (2023). Digital empires: The global battle to regulate technology. Oxford University Press.

Bu, Z., & Wang, Y.-X. (2023, May 11). Differential privacy for deep learning at GPT scale. Amazon Science.

https://www.amazon.science/blog/differential-privacy-for-deep-learning-at-gpt-scale

Cheong, B. C. (2024). Transparency and accountability in AI systems: Safeguarding wellbeing in the age of algorithmic decision-making. Frontiers in Human Dynamics, 6. https://doi.org/10.3389/fhumd.2024.1421273

CISA. (2016, February). Sharing of cyber threat indicators and defensive measures by the Federal Government under the Cybersecurity Information Sharing Act of 2015. Cybersecurity and Infrastructure Security Agency CISA. https://www.cisa.gov/resources-tools/resources/sharing-cyber-threat-indicators-and-defensive-measures-federal-government-under-cybersecurity

Cohn, C., Hutchins, N., S, A. T., & Biswas, G. (2025). CoTAL: Human-in-the-Loop prompt engineering, chain-of-thought reasoning, and active learning for generalizable formative assessment scoring. In arXiv.org. https://arxiv.org/abs/2504.02323

Colorado General Assembly. (2024). Consumer protections for artificial intelligence. Colorado General Assembly. https://leg.colorado.gov/sites/default/files/2024a_205_signed.pdf

Comunale, M. (2024). The Economic Impacts and the Regulation of AI: A review of the academic literature and policy actions. IMF Working Papers, 2024(065), 1. https://doi.org/10.5089/9798400268588.001

CFPB. (2023). Chatbots in consumer finance. In Consumer Financial Protection Bureau. https://www.consumerfinance.gov/data-research/research-reports/chatbots-in-consumer-finance/chatbots-in-consumer-finance/

Council of the EU. (2023, December 9). Artificial intelligence act: Council and Parliament strike a deal on the first rules for AI in the world. Council of the EU Press Release. https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/

Crisanto, J. C., Leuterio, C. B., Prenio, J., & Yong, J. (2024, December 12). Regulating AI in the financial sector: Recent developments and main challenges. BIS. https://www.bis.org/fsi/publ/insights63.htm

Cyberspace Administration of China. (2023, July 13). Interim Measures for the Administration of Generative Artificial Intelligence Services. Cac.Gov. https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

Delev, Z. (2024, April 30). *The future of finance: Adapting to AI and data privacy laws*. GDPR Local. https://gdprlocal.com/the-future-of-finance-adapting-to-ai-and-data-privacy-laws/

Department for Science, Innovation Technology. (2025, January 13). AI opportunities action plan. GOV.UK. https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan

Department for Science, Innovation Technology & Office for Artificial Intelligence. (2023, March 29). A pro-innovation approach to AI regulation. GOV.UK. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

Directive - 2015/2366 - EN - Payment services directive - EUR-Lex. (2015). https://eur-lex.europa.eu/eli/dir/2015/2366/oj

Ernst & Young LLP. (2023). *The Artificial Intelligence (AI) global regulatory landscape: Policy trends and considerations to build confidence in AI*. https://www.ey.com/content/dam/ey-unified-site/ey-com/en-gl/insights/ai/documents/ey-the-artificial-intelligence-ai-global-regulatory-landscape-final.pdf

EU Artificial Intelligence Act. (2024a). Annex III: High-Risk AI systems referred to in article 6(2). EU Artificial Intelligence Act. https://artificialintelligenceact.eu/annex/3/

EU Artificial Intelligence Act. (2024b). Article 49: Registration. EU Artificial Intelligence Act. https://artificialintelligenceact.eu/article/49/

EU Artificial Intelligence Act. (2024c). Article 50: Transparency obligations for providers and deployers of certain AI systems. EU Artificial Intelligence Act. https://artificialintelligenceact.eu/article/50/

EU Artificial Intelligence Act. (2024d). The AI act explorer. EU Artificial Intelligence Act. https://artificialintelligenceact.eu/ai-act-explorer/

European Banking Authority. (2025). Guidelines on ICT and security risk management. EBA. https://www.eba.europa.eu/activities/single-rulebook/regulatory-activities/internal-governance/guidelines-ict-and-security-risk-management

European Commission. (2024, August 1). AI Act enters into force. European Commission. https://commission.europa.eu/news/ai-act-enters-force-2024-08-01_en

European Commission. (2025). AI act. Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

European Parliament. (2023). EU AI Act: First regulation on artificial intelligence. Topics | European Parliament. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#ai-regulation-in-europe-the-first-comprehensive-framework-4:~:text=The%20Artificialt%20Intelligence,2%20February%202025

European Union. (2022a). Directive - 2022/2555 - EN - EUR-Lex. EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022L2555

European Union. (2022b). Regulation - 2022/2554 - EN - DORA - EUR-Lex. EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2554#art_45

Federal Register. (2023, November 1). Safe, secure, and trustworthy development and use of artificial intelligence. Federal Register. https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence

Federal Reserve Board. (2021). Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, including Machine Learning. https://www.federalreserve.gov/newsevents/pressreleases/files/bcreg20210517a1.pdf

Feretzakis, G., Papaspyridis, K., Gkoulalas-Divanis, A., & Verykios, V. S. (2024). Privacy-Preserving techniques in generative AI and large language models: A narrative review. Information, 15(11). https://doi.org/10.3390/info15110697

FFIEC. (2015). FFIEC cybersecurity assessment tool . https://omb.report/icr/201907-1557-001/doc/92940701

Financial Services Agency. (2025, April 20). Publication of AI discussion paper-English. FSA. https://www.fsa.go.jp/en/news/2025/20250304/aidp.html

Financial Services Agency . (25 C.E.). AI Discussion Paper (Version 1.0): Preliminary Discussion Points for Promoting the Sound Utilization of AI in the Financial Sector. FSA. https://www.fsa.go.jp/en/news/2025/20250304/aidp.html

FINRA. (2020). Artificial intelligence (AI) in the securities industry. FINRA.Org. https://www.finra.org/rules-guidance/key-topics/fintech/report/artificial-intelligence-in-the-securities-industry

FS-ISAC. (2024). *Adversarial AI frameworks: Taxonomy, threat landscape, and control frameworks* [White paper]. https://www.fsisac.com/hubfs/Knowledge/AI/FSISAC_Adversarial-AI-Framework-TaxonomyThreatLandscapeAndControlFrameworks.pdf

FS-ISAC. (2024). *Building AI into cyber defense* [White paper]. https://www.fsisac.com/hubfs/Knowledge/AI/FSISAC_BuildingAI-IntoCyberDefense.pdf

FS-ISAC. (2024). *Combating threats and reducing risks posed by AI* [White paper]. https://www.fsisac.com/hubfs/Knowledge/AI/FSISAC_CombatingThreatsAndReducingRisksPosedByAI.pdf

FS-ISAC. (2024). *Framework of an acceptable use policy for external generative AI* [White paper]. https://www.fsisac.com/hubfs/Knowledge/FrameworkOfAnAcceptableUsePolicyForExternalGenerativeAI.pdf

FS-ISAC. (2024). *Generative AI vendor evaluation and qualitative risk assessment* [White paper].

https://www.fsisac.com/hubfs/Knowledge/AI/FSISAC_GenerativeAI-VendorEvaluation%26QualitativeRiskAssessment.pdf

FS-ISAC. (2024). *Responsible AI principles* [White paper]. https://www.fsisac.com/hubfs/Knowledge/AI/FSISAC_ResponsibleAI-Principles.pdf

FSB. (2024, November 14). *The financial stability implications of artificial intelligence*. Financial Stability Board. https://www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/

Garg, A., Schoeman, D., Asaftei, G. M., Buehler, K., & Grennan, L. (2025, March 27). Amit garg. *McKinsey & Company*. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/how-financial-institutions-can-improve-their-governance-of-gen-ai

Giudici, P., & Raffinetti, E. (2023, February 28). *SAFE artificial intelligence in finance*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4362034

Haeck, Pieter. "EU Opens Door to Reworking AI Rulebook." *Politico*, April 9, 2025. https://www.politico.eu/article/how-eu-did-full-180-artificial-intelligence-rules/ .

Harris, L., & Jaikaran, C. (2024, March 4). Highlights of the 2023 Executive Order on Artificial Intelligence for Congress. Library of Congress. https://www.congress.gov/crs-product/R47843

Highlights of the 2023 Executive Order on Artificial Intelligence for Congress. (2025, April 13). https://www.congress.gov/crs-product/R47843

Henderson, A. (2024, August 9). EU AI act: Key points for financial services businesses. *Goodwin*. https://www.goodwinlaw.com/en/insights/publications/2024/08/alerts-practices-pif-key-points-for-financial-services-businesses

H.R.4718 - 99th Congress (1985-1986): Computer Fraud and Abuse Act of 1986. (1986, October 16). Library of Congress. https://www.congress.gov/bill/99th-congress/house-bill/4718

Intel. (2024, November 18). TEEs overview. Intel® Trust Authority. https://docs.trustauthority.intel.com/main/articles/concept-tees-overview.html

Johal, J. (2024, October 29). *Balancing Act: Managing AI governance risks in financial services*. Alvarez & Marsal | Management Consulting | Professional Services. https://www.alvarezandmarsal.com/insights/balancing-act-managing-ai-governance-risks-financial-services

Khan, L. (2023, April 25). Joint statement on enforcement efforts against discrimination and bias

in automated systems. Federal Trade Commission. https://www.ftc.gov/system/files/ftc_gov/pdf/EEOC-CRT-FTC-CFPB-AI-Joint-Statement%28final%29.pdf

Koide, M. (2022, August 4). AI for good: Research insights from financial services. *Brookings*. https://www.brookings.edu/articles/ai-for-good/

Kowald, D., Scher, S., Pammer-Schindler, V., Müllner, P., Waxnegger, K., Demelius, L., Fessl, A., Toller, M., Mendoza Estrada, I. G., Šimić, I., Sabol, V., Trügler, A., Veas, E., Kern, R., Nad, T., & Kopeinik, S. (2024). Establishing and evaluating trustworthy AI: Overview and research challenges. Frontiers in Big Data, 7. https://doi.org/10.3389/fdata.2024.1467222

LaForge, G. (2024, September 5). The dangers of imposing global north approaches to AI governance on the global south. *Tech Policy Press*. https://www.techpolicy.press/the-dangers-of-imposing-global-north-approaches-to-ai-governance-on-the-global-south/

Lai, H. (2023). Chinese AI regulation proposal.md. Gist. https://gist.github.com/FooBarWidget/201ea5e0983d05d21f6719bacb46795e

Maas, M. M., & Villalobos, J. J. (2023, September 23). *International AI institutions: A literature review of models, examples, and proposals*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4579773

Maple, C., Sabuncuoglu, A., Szpruch, L., Elliot, A., Reinert, G., & Zemaitis, T. (2024). The impact of large language models in finance: Towards trustworthy adoption. The Alan Turing Institute. https://www.turing.ac.uk/news/publications/impact-large-language-models-finance-towards-trustworthy-adoption

Ministry of Internal Affairs and Communications Ministry of Economy, Trade and Industry. (2024). AI Guidelines for Business Ver1.0. Meti. https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20240419_9.pdf

Monetary Authority of Singapore. (2018a, November 12). MAS introduces new FEAT Principles to promote responsible use of AI and data analytics. https://www.mas.gov.sg/news/media-releases/2018/mas-introduces-new-feat-principles-to-promote-responsible-use-of-ai-and-data-analytics

Monetary Authority of Singapore. (2018b, November 12). Principles to promote fairness, ethics, accountability and transparency (FEAT) in the use of artificial intelligence and data analytics in singapore's financial sector. MAS. https://www.mas.gov.sg/publications/monographs-or-information-paper/2018/feat

Monetary Authority of Singapore. (2021, January 18). Guidelines on risk management practices – technology risk. MAS.

https://www.mas.gov.sg/-/media/mas/regulations-and-financial-stability/regulatory-and-supervisory-framework/risk-management/trm-guidelines-18-january-2021.pdf

Moody's. (2024, July 31). European Union's AI Act: Implications for fraud prevention and anti-money laundering systems. Moody's. https://www.moodys.com/web/en/us/kyc/resources/insights/european-unions-ai-act-implications-for-fraud-prevention-anti-money-laundering-systems.html

Mukherjee, S., & Vagnoni, G. (2023, April 28). Italy restores ChatGPT after OpenAI responds to regulator. Reuters. https://www.reuters.com/technology/chatgpt-is-available-again-users-italy-spokesperson-says-2023-04-28/

NayaOne. (2024, February 16). *AI governance in financial services: Issues & best practices*. NayaOne. https://nayaone.com/blog/ai-governance-in-financial-services-challenges-and-best-practices/

NIS2 Directive: New rules on cybersecurity of network and information systems. (2022). Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/policies/nis2-directive

NIST. (2024). Artificial intelligence risk management framework : National Institute of Standards and Technology (U.S.). https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

NYDFS. (2017). NYCRR Part 500. Department of Financial Services. https://www.dfs.ny.gov/system/files/documents/2023/12/rf23_nycrr_part_500_amend02_20231101.pdf#:~:text=this%20regulation%20is%20designed%20to,an%20annual%20certification%20confirming%20compliance

OCED. (2024, September 24). *Regulatory approaches to artificial intelligence in finance: Oecd artificial intelligence papers*. OCED. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/09/regulatory-approaches-to-artificial-intelligence-in-finance_43d082c3/f1498c02-en.pdf

Ogunkeye, G. (2024, May 1). *A literature review on the regulation of artificial intelligence*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4803353

Phillips, T. (2024, September 26). *The risks of generative AI agents to financial services*. Roosevelt Institute. https://rooseveltinstitute.org/publications/the-risks-of-generative-ai-agents-to-financial-services/

Quinn, B. (2025, March 17). China and Spain introduce requirements on labelling of AI-generated content. IPTC. https://iptc.org/news/china-and-spain-introduce-ai-generated-content-labelling-requirements/

Recital 7 - The Framework is Based on Control and Certainty. (2016, July 14). General Data

Protection Regulation (GDPR). https://gdpr-info.eu/recitals/no-7/

Sandridge, T., Grant, R., Forster, S., & Harris, B. (2024, August 12). Maximizing compliance: Integrating Gen AI into the financial regulatory framework. *IBM*. https://www.ibm.com/think/insights/maximizing-compliance-integrating-gen-ai-into-the-financial-regulatory-framework

Sherman, N. (2025, January). AI regulations around the world. 2025. https://www.mindfoundry.ai/blog/ai-regulations-around-the-world

Statista Research Department. (2025, February). *Financial sector AI spending 202*. Statista. https://www.statista.com/statistics/1446037/financial-sector-estimated-ai-spending-forecast/

Stone, S., Pierides, M., & Mulligan, J. (2024, May 21). Global financial services: The sector's current AI regulatory landscape. *Morgan Lewis*. https://www.morganlewis.com/pubs/2024/05/global-financial-services-the-sectors-current-ai-regulatory-landscape

Su, N., Kahlen, L., Huang, S., & Lu, W. (2022). Trustworthy use of artificial intelligence in finance: Regulatory perspectives from Asia Pacific. In *Deloitte* . https://www2.deloitte.com/cn/en/pages/financial-services/articles/trustworthy-use-ai-finance-regulatory-perspectives-asia-pacific.html

Text - H.R.4173 - 111th Congress (2009-2010): Dodd-Frank Wall Street Reform and Consumer Protection Act. (2010, July 21). https://www.congress.gov/bill/111th-congress/house-bill/4173/text

Text - S.754 - 114th Congress (2015-2016): An act to improve cybersecurity in the United States through enhanced sharing of information about cybersecurity threats, and for other purposes. (2015, October 28). https://www.congress.gov/bill/114th-congress/senate-bill/754/text

The Board of the International Organization of Securities Commissions. (2025). Artificial Intelligence in Capital Markets: Use Cases, Risks, and Challenges. IOSCO. https://www.iosco.org/library/pubdocs/pdf/IOSCOPD788.pdf

The White House. (2022, October 4). Blueprint for an AI Bill of Rights. Bidenwhitehouse. https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/

The White House. (2025, January 23). Removing barriers to American leadership in artificial intelligence – The White House. The White House. https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/

ThuCCSLab. (2025). Awesome-LM-SSP/collection/paper/privacy/privacy-preserving_computation.md at main ·

ThuCCSLab/Awesome-LM-SSP. GitHub. https://github.com/ThuCCSLab/Awesome-LM-SSP/blob/main/collection/paper/privacy/privacy-preserving_computation.md

Tony Blair Institute for Global Change. (2025, February 6). How leaders in the global south can devise AI regulation that enables innovation. *Tony Blair Institute*. https://institute.global/insights/tech-and-digitalisation/how-leaders-in-the-global-south-can-devise-ai-regulation-that-enables-innovation

ucdlawreview. (2025, April 22). *AI and the global financial system: Innovative risks and regulatory challenges*. UCD Law Review. https://theucdlawreview.com/2025/04/22/ai-and-the-global-financial-system-innovative-risks-and-regulatory-challenges/

U.S. DEPARTMENT OF THE TREASURY. (2024, December). Artificial intelligence report on the uses, opportunities, and risks of artificial intelligence in the financial services sector. U.S. Department of the Treasury. https://home.treasury.gov/system/files/136/Artificial-Intelligence-in-Financial-Services.pdf

U.S. Department of the Treasury. (2024, December 19). Treasury releases report on the uses, opportunities, and risks of artificial intelligence in financial services. U.S. Department of the Treasury. https://home.treasury.gov/news/press-releases/jy2760

Yu, D., Kairouz, P., Oh, S., & Xu, Z. (2024, February 21). Privacy-Preserving instructions for aligning large language models. arXiv.Org. https://arxiv.org/abs/2402.13659

Yu, D., Rosenfeld, H., & Gupta, A. (2023, January 16). The 'AI divide' between the Global North and Global South. *World Economic Forum*. https://www.weforum.org/stories/2023/01/davos23-ai-divide-global-north-global-south/

Zarecki, I. (2025, February 21). *Protecting sensitive financial information in the age of gen AI*. BAI. https://www.bai.org/banking-strategies/protecting-sensitive-financial-information-in-the-age-of-gen-ai/

Zhang, L. (2023, July 19). China: Generative AI measures finalized. The Library of Congress. https://www.loc.gov/item/global-legal-monitor/2023-07-18/china-generative-ai-measures-finalized/

Zhang, Z., Jia, M., Lee, H.-P., Yao, B., Das, S., Lerner, A., Wang, D., & Li, T. (2024). "It's a Fair Game", or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents. Proceedings of the CHI Conference on Human Factors in Computing Systems, 1–26. https://doi.org/10.1145/3613904.364238