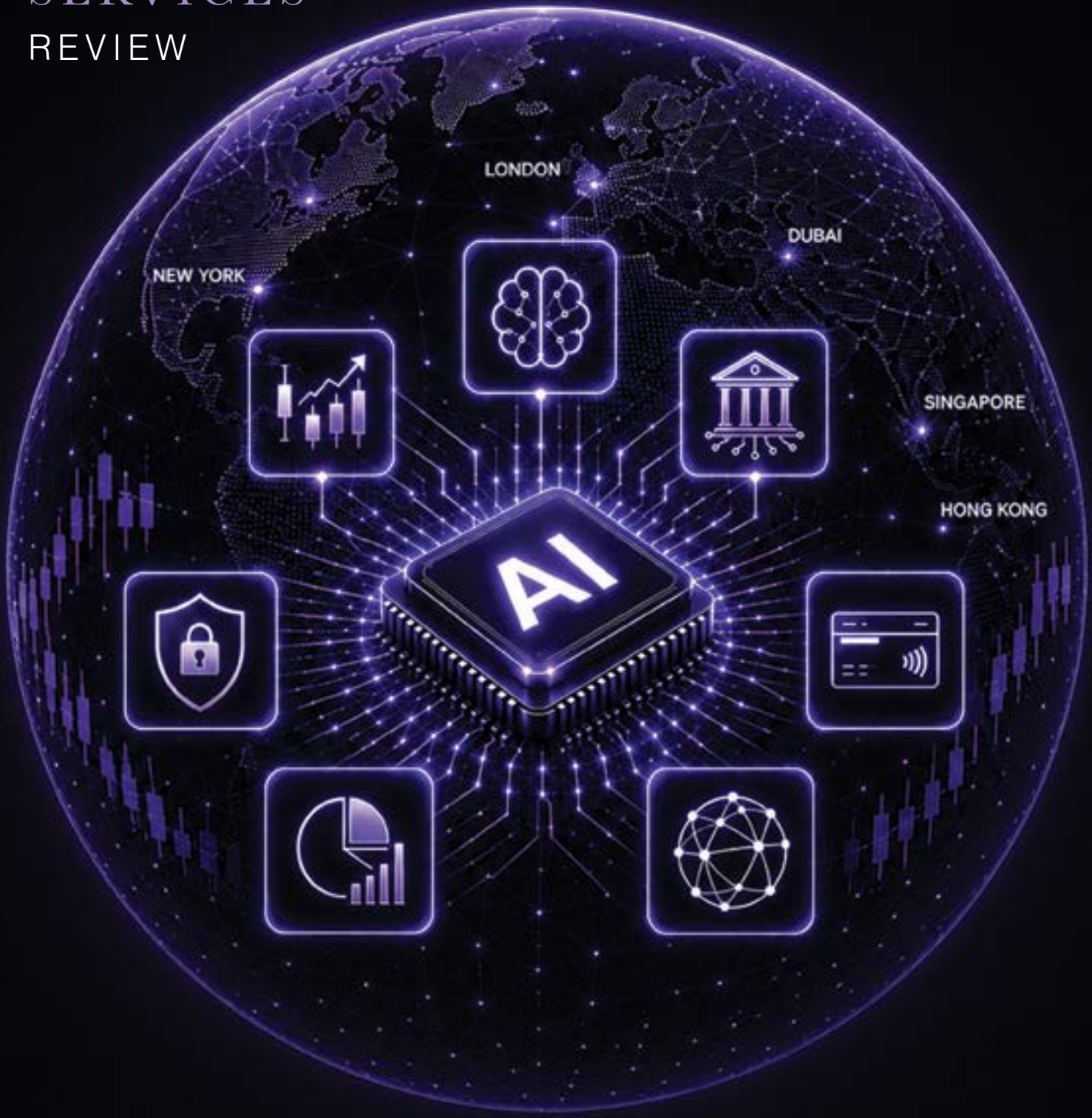
INTERNATIONAL AI IN FINANCIAL SERVICES REVIEW

2026/27 EDITION



Beaumont Capital Markets

www.beaumont-capitalmarkets.co.uk

The International AI in Financial Services Review 2026/27

Content

Country chapters

WORLD BANK

Dessislav Dobrev
Senior Counsel

ddobrev@worldbank.org

www.worldbank.org

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

BIO

Dessislav Dobrev is a leading expert in artificial intelligence and the law and has spearheaded AI matters in both academia and practice. His recent book Artificial Intelligence and the Law: A Comprehensive Guide for the Legal Profession, Academia and Society, published by Thomson Reuters, has been endorsed by high-profile academics, leading law practitioners and global AI leaders. This timely volume is one of the first comprehensive studies of how AI will dramatically affect the law, both as a profession and a regulatory domain, as well as society at large.

As an Adjunct Professor Dessislav created and taught new law school courses on AI at McGill Law School, has previously taught at Georgetown Law School, and has provided seminars on AI at other leading institutions such as the Berkeley Center for Law & Technology and the University of Toronto's Faculty of Law. In addition, Dessislav serves on the New York State Bar Association's Task Force on Artificial Intelligence. The task force is dedicated to identifying potential benefits, dangers, and opportunities surrounding AI and to make regulatory recommendations for how to manage and integrate the technology in legal practice.

Mr. Dobrev is educated and has practiced law in both civil and common law jurisdictions and has worked in both public and private sector contexts. He is currently a Senior Counsel at the World Bank Group, where he is exposed to AI issues at the global level. Previously, he practiced corporate law at Davis Polk in New York.

This foreword explores in general terms the strategic evolution of artificial intelligence (AI) in financial services over the last year. Some of the key trends to highlight over this period include the wider adoption and integration of AI, the increased and more complex risks associated with it, and the evolving regulatory environment. Each of these will be briefly examined, as well as on what to keep a keen eye looking ahead.

> One of the most significant developments has been the emergence of AI systems capable of carrying out complex sequences of tasks with limited human intervention.

Wide-Ranging Adoption and Integration of AI

AI has embarked upon a new phase within the financial services sector. The discourse is no longer centered merely on whether AI has potential value for financial institutions. It is now becoming increasingly evident that it has. Rather, attention has shifted toward how organizations can more optimally integrate increasingly sophisticated AI systems into their core business functions while maintaining effective governance, risk management and regulatory compliance. Over the past twelve months, advances in generative and agentic AI have accelerated adoption across banking, insurance, capital markets and payment systems, transforming AI from a specialized analytical tool into a broader operational capability. Recent industry research indicates that AI deployment has become widespread across financial institutions worldwide, with technology-driven firms continuing to lead traditional market participants in both experimentation and implementation. This is separate from, and in addition to, financial institutions achieving higher revenues as AI is driving demand for financing, from data centers and power infrastructure to capital markets activity (such as IPOs).

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

One of the most significant developments has been the emergence of AI systems capable of carrying out complex sequences of tasks with limited human intervention. Earlier deployments focused primarily on prediction and classification. Today's systems increasingly assist with conducting compliance checks, reviewing documentation, drafting reports, supporting software development and coordinating multi-step business processes. Financial institutions are employing these capabilities to reduce operational friction, improve productivity (e.g., through improvements in internal workflows, knowledge management and operational execution) and enhance internal decision-making.

In parallel, AI is becoming increasingly embedded in market-facing activities. Financial firms are using advanced models to support portfolio management, market surveillance, anti-money laundering controls, fraud monitoring and customer engagement. In lending, richer data analysis is enabling more granular assessments of borrower characteristics, while insurers are applying AI to streamline underwriting and claims administration. Capital markets participants are leveraging AI-assisted research and market intelligence tools capable of processing vast amounts of structured and unstructured information in real time. International securities regulators have observed particularly strong growth in the use of AI for investment analysis, surveillance and regulatory compliance functions. Further, consumers worldwide are increasingly using AI to guide investment decisions.

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

Complex Risk Landscape

Alongside these opportunities, the risk landscape has become more complex. Concerns regarding inaccurate outputs, algorithmic bias and privacy remain relevant, but the policy debate has broadened considerably over the last year. Regulators and industry participants are increasingly focused on issues such as model concentration, dependence on external technology providers, cybersecurity vulnerabilities and the growing reliance on a limited number of cloud and foundation-model platforms. These concerns reflect the reality that AI-related risks may arise not only from the models themselves but also from the broader technological ecosystem on which financial institutions depend. The International Organization of Securities Commissions has highlighted third-party dependency, model governance and malicious uses of AI among the principal emerging concerns for financial markets.

Evolving Regulatory Environment

A notable trend over the last twelve months has been the greater involvement of international standard-setting bodies. Policymakers are increasingly approaching AI as a matter of financial resilience and institutional governance rather than merely a technology issue. The Financial Stability Board's 2026 consultation on responsible AI adoption emphasizes enterprise-wide oversight, board accountability, risk management and lifecycle controls for AI systems. This reflects an emerging consensus that successful AI implementation requires governance frameworks capable of keeping pace with technological innovation.

Around the world, supervisory authorities are moving beyond observation and exploratory studies toward more structured expectations regarding transparency, accountability, monitoring and human oversight. Rather than creating entirely new regulatory regimes, many authorities are adapting existing prudential, consumer protection and market conduct frameworks to address AI-related challenges. The resulting trajectory suggests increasing convergence around common principles such as explainability, documentation, testing and ongoing supervision.

Looking Ahead

Some of the main risk-based trends to watch for in the next year include the following. *First*, there appears to be a growing sense of potential over-investment in the AI sector in certain jurisdictions. Its magnitude, dependence on debt, and circular financings raise questions about sustainability and financial stability. If the current boom turns into bust, the financial services industry will also be impacted significantly. *Second*, as financial services companies are widely using AI agents for common business operations, there will be a growing concern about the implications of advanced AI systems hacking into digital financial infrastructure. This is about AI agents subverting human controls and displaying unauthorized or downright deceptive behavior. *Third*, there currently is a widening structural gap in AI adoption, integration and knowledge as between financial institutions and regulatory and supervisory authorities. Such divergence in adoption and integration could engender a limited ability to assess AI models used by regulated entities and difficulties in monitoring emerging or fastmoving risks, thereby undermining supervisory effectiveness over time. These and other trends will determine broadly the direction of international AI in financial services going forward.

> “
>
> Policymakers are increasingly approaching AI as a matter of financial resilience and institutional governance rather than merely a technology issue.

BELIZE

AI AND THE FUTURE OF INTERNATIONAL BANKING: WHY TRUST, TECHNOLOGY AND GLOBAL DIVERSIFICATION MATTER MORE THAN EVER

CAYE INTERNATIONAL BANK

Dr. Luigi Wewege
President of Caye International Bank

president@cayebank.bz

+501 226 2388

www.cayebank.bz

BIO

Dr. Luigi Wewege is the President of Caye International Bank, a Belize based, multi-award-winning institution recognized among the leading international banks within the Caribbean and Central America. Under his leadership, Caye has grown from the fifth largest to the largest international bank in Belize. Luigi also serves on several international advisory boards across asset protection, education, financial technology, media and wealth management.

A frequent international speaker and recognized thought leader, Wewege is an accomplished author whose publications include The Digital Banking Revolution, now in its third edition and "Disruptions and Digital Banking Trends" published in The Journal of Applied Finance & Banking. He also serves as an Instructor with the FinTech School, contributing to executive learning initiatives.

Luigi holds a Doctor of Philosophy (PhD) from the International School of Management in Paris, where his doctoral research focused on retail banking crises in Central America and the region's future financial stability. He also holds an MBA in International Business from the MIB Trieste School of Management and a BSBA from the University of Missouri–St. Louis, where he graduated with a triple major in Finance, International Business, and Management.

Caye International Bank®

AI AND THE FUTURE OF INTERNATIONAL BANKING: WHY TRUST, TECHNOLOGY AND GLOBAL DIVERSIFICATION MATTER MORE THAN EVER

International banking is undergoing one of the most significant transformations in its history. Advances in artificial intelligence, rising geopolitical uncertainty, increasingly sophisticated cybersecurity threats, evolving regulatory expectations, and the growing international mobility of both individuals and businesses are fundamentally reshaping how clients think about managing their wealth.

> “
>
> Yet despite these profound technological developments, one principle remains constant: banking has always been, and will continue to be, a business built on trust.

Artificial intelligence sits at the centre of much of this change. It is altering how banks assess risk, monitor transactions, identify fraud, support regulatory compliance and engage with clients across borders. Yet despite these profound technological developments, one principle remains constant: banking has always been, and will continue to be, a business built on trust.

For decades, international banking was often viewed through a narrow lens. Many assumed it was designed primarily for the ultra-wealthy or associated it with secrecy and regulatory arbitrage. That perception no longer reflects reality. Modern international banking is defined by transparency, rigorous compliance and helping globally minded individuals and businesses manage risk across an increasingly interconnected and technology-driven world.

Today’s clients are entrepreneurs expanding internationally, investors diversifying across jurisdictions, family offices protecting generational wealth, and professionals whose financial lives span multiple countries. They require banking partners capable of navigating complex regulatory environments while providing sophisticated cross-border solutions that combine security, flexibility and increasingly intelligent use of data.

Caye International Bank®

One of the defining characteristics of the coming decade will be the increasing importance of diversification. In an era marked by geopolitical tensions, economic uncertainty and rapidly changing regulations, concentration risk has become one of the greatest yet most overlooked threats to wealth preservation.

Diversification is no longer simply about investment portfolios. It extends to banking relationships, currencies, legal jurisdictions and financial structures. Increasingly, prudent clients are asking not only where they should bank, but where they can best protect their assets while maintaining optionality should global conditions change.

Artificial intelligence can support this process by helping financial institutions analyse large volumes of economic, regulatory and behavioural data more efficiently. AI-powered systems can assist banks in identifying emerging risks, assessing client exposure across jurisdictions and detecting patterns that may be difficult to recognise through traditional methods alone.

However, the use of AI does not remove the need for experienced judgment. Decisions involving cross-border wealth, geopolitical exposure, succession planning and long-term financial security cannot be reduced entirely to automated outputs. Technology can strengthen the decision-making process, but it must remain supported by human expertise and a clear understanding of each client’s circumstances.

Much discussion has focused on whether AI will replace bankers. The reality is considerably more nuanced. Artificial intelligence excels at processing enormous volumes of data, identifying unusual transaction patterns, strengthening fraud detection, enhancing compliance monitoring and improving operational efficiency. These capabilities allow banks to deliver faster, more accurate and more secure services.

In international banking, AI has particularly significant potential within anti-money laundering, know-your-customer procedures and sanctions screening. Cross-border institutions frequently deal with clients, transactions and corporate structures spanning multiple legal systems. AI can help analyse documentation, identify inconsistencies, flag unusual behaviour and prioritise higher-risk cases for human review.

AI can also improve the client onboarding process. International account applications often involve substantial documentation, source-of-funds verification and complex ownership structures. Intelligent systems can help organise and assess this information more quickly, reducing delays while maintaining robust compliance standards.

Used responsibly, this can improve both security and the client experience. However, banks must ensure that speed does not come at the expense of accuracy, fairness or appropriate oversight.

International banking remains fundamentally relationship-driven. Advising clients on cross-border wealth structuring, succession planning, geopolitical risk and long-term financial strategy requires judgment, experience and trust, qualities that cannot be fully replicated by algorithms. AI will undoubtedly transform banking, but it will serve most effectively as a powerful tool supporting experienced professionals rather than replacing them.

This is especially important where AI-generated recommendations may influence significant financial decisions. Clients must be able to understand how decisions are reached, particularly when an automated system affects onboarding, transaction monitoring, risk classification or access to services.

Banks therefore need strong governance frameworks for the use of AI. This includes clearly defined accountability, appropriate testing, ongoing monitoring and the ability for qualified professionals to challenge or override automated conclusions when necessary.

Technology also brings heightened responsibilities. Cybersecurity has become one of the defining challenges facing financial institutions and their clients alike. While banks continue investing heavily in advanced security infrastructure and monitoring systems, cybersecurity is increasingly a shared responsibility.

AI is strengthening cybersecurity by helping institutions detect unusual network activity, identify suspicious transactions and respond to threats more rapidly. At the same time, it is also increasing the sophistication of criminal activity. Fraudsters can use generative AI to create highly convincing phishing messages, impersonate trusted contacts, replicate voices and produce false documentation.

This creates a more challenging threat environment for both banks and clients. High-net-worth individuals and internationally active businesses must approach digital security with the same discipline they apply to managing their investments. Multi-factor authentication, secure communications, careful verification procedures and heightened awareness of increasingly sophisticated fraud attempts are no longer optional; they are essential components of modern wealth protection.

“

AI will undoubtedly transform banking, but it will serve most effectively as a powerful tool supporting experienced professionals rather than replacing them.

Financial institutions must also ensure that AI systems themselves are secure. Sensitive client information, transaction data and risk assessments must be protected against unauthorised access or misuse. Data governance will therefore become as important as the technology itself.

Meanwhile, the rapid growth of digital assets presents both opportunities and challenges. Blockchain technology offers remarkable transparency, yet regulatory frameworks remain fragmented across jurisdictions. Financial institutions must carefully balance innovation with robust compliance, ensuring they fully understand the source of funds while managing custody, volatility and evolving regulatory expectations.

> “
>
> Wealth management has evolved beyond maximising returns. It now encompasses preserving flexibility, resilience and freedom of choice.

AI can assist banks in analysing blockchain transactions, identifying suspicious wallet activity and tracing complex flows of digital assets. These capabilities may help financial institutions engage with the digital asset economy while maintaining appropriate controls.

However, the reliability of any AI-supported analysis depends on the quality of the underlying data and the effectiveness of the models being used. Banks must avoid treating automated analysis as infallible and should continue applying experienced human review, particularly where regulatory or reputational consequences may be significant.

This broader regulatory evolution is creating an interesting paradox. International standards surrounding anti-money laundering, know-your-customer requirements, financial transparency and increasingly the responsible use of AI are becoming more aligned through global organisations. At the same time, individual jurisdictions continue implementing these standards differently, creating greater operational complexity for international banks.

The regulation of AI is likely to add another layer to this challenge. Institutions operating internationally may be required to comply with different rules concerning data privacy, automated decision-making, consumer protection, model transparency and accountability.

Success therefore depends on maintaining compliance frameworks that exceed minimum regulatory expectations while remaining sufficiently flexible to accommodate local requirements. Institutions that adopt globally consistent standards while executing with local precision will be best positioned for long-term success.

For AI specifically, this means establishing clear standards governing how systems are selected, trained, tested and monitored. It also means ensuring that responsibility remains with the institution and its professionals rather than being transferred to the technology provider or algorithm.

Perhaps the most underestimated shift currently taking place is the convergence of banking with personal mobility. Financial decisions are no longer made independently of lifestyle decisions. More clients are integrating their banking relationships with second residency strategies, international business expansion, family succession planning and broader geopolitical diversification.

Wealth management has evolved beyond maximising returns. It now encompasses preserving flexibility, resilience and freedom of choice. AI can help banks develop a more complete understanding of internationally mobile clients by bringing together information across banking activity, currency exposure, jurisdictions and financial structures. This can allow institutions to provide more relevant support and identify potential risks earlier.

However, increased personalisation must be handled carefully. Clients must have confidence that their data is being used lawfully, securely and for legitimate purposes. International banking depends heavily on discretion, and the use of advanced technology must never undermine the confidentiality and trust expected by clients.

This convergence reflects a broader understanding that financial security increasingly depends upon optionality. Clients want banking relationships that complement how they live, work, invest and move across borders. They also expect technology to make those relationships more efficient without making them impersonal or opaque.

Ultimately, the successful international banks of the future will not necessarily be the largest institutions. They will be those that successfully combine four defining characteristics: trust, technological capability, adaptability and global relevance.

Trust will remain the cornerstone of every successful banking relationship. Technological capability will allow institutions to process information, manage risk and support clients more effectively. Adaptability will enable banks to embrace emerging technologies while navigating changing regulations and evolving client expectations. Global relevance will allow them to deliver seamless cross-border solutions that reflect the increasingly international nature of modern wealth.

The future belongs to institutions that combine AI-driven innovation with personalised service, strong governance with operational agility, and sophisticated infrastructure with trusted human relationships.

International banking has evolved beyond simply safeguarding assets. It now serves as a strategic platform that helps clients build resilience, manage uncertainty and confidently participate in an increasingly interconnected world.

Artificial intelligence will play a major role in that future, but its value will depend on how responsibly it is used. The strongest institutions will not be those that automate the most, but those that use technology intelligently while preserving accountability, transparency and human judgment.

Those that recognise this balance today will be best positioned to thrive tomorrow.

About the author

Dr. Luigi Wewege is the President of Caye International Bank, a Belize-based, multi-award-winning institution recognised among the leading international banks within the Caribbean and Central America. Under his leadership, Caye has grown from the fifth largest to the largest international bank in Belize.

Luigi also serves on several international advisory boards across asset protection, education, financial technology, media and wealth management. A frequent international speaker and recognised thought leader, Wewege is an accomplished author whose publications include *The Digital Banking Revolution*, now in its third edition, and "Disruptions and Digital Banking Trends", published in *The Journal of Applied Finance & Banking*.

He also serves as an Instructor with the FinTech School, contributing to executive learning initiatives. Luigi holds a Doctor of Philosophy from the International School of Management in Paris, where his doctoral research focused on retail banking crises in Central America and the region's future financial stability.

"

The strongest institutions will not be those that automate the most, but those that use technology intelligently while preserving accountability, transparency and human judgment.

He also holds an MBA in International Business from the MIB Trieste School of Management and a BSBA from the University of Missouri–St. Louis, where he graduated with a triple major in Finance, International Business and Management.

CHINA

AI IN CHINESE FINANCIAL SERVICES: FROM FINTECH TO FOUNDATION MODELS — MARKET PRACTICE, REGULATORY ARCHITECTURE AND THE COMPLIANCE ROADMAP

KING & WOOD

Yirui Zhang (Alicia)
Partner

zhangyirui@cn.kingandwood.com

+86 21 24126063

www.kingandwood.com

金杜律师事务所
KING&WOOD

BIO

Alicia is a partner in the Technology, Media, Entertainment and Cross-border Intellectual Property (TMEC) practice group at King & Wood. She specializes in intellectual property, technology commercialization, and strategic management of data and digital assets, with particular focus on artificial intelligence, fintech, and robotics.

Alicia advises leading enterprises on digital transformation, data transactions, AI compliance, algorithmic governance, and technology ethics frameworks. Her practice spans the application and regulation of emerging technologies at the intersection of finance and innovation. She serves as Director of the Digital Technology and AI Committee of the Shanghai Bar Association and holds advisory roles with the National Facility for Translational Medicine and Shanghai Data Service Provider Association.

Recognized as one of Shanghai's Outstanding Female Lawyers and ranked in IAM Patent 1000, Alicia holds an LL.B. and B.S. in Computer Science from East China University of Political Science and Law and an LL.M. from the University of Illinois Urbana-Champaign. She is admitted to practice in the PRC and New York, and works in Chinese and English.

金杜律师事务所
KING&WOOD

From FinTech to financial foundation models and AI agents

The application of artificial intelligence ("AI") in Chinese financial services ("FS") is best understood as the latest stage of a decades-long digitalisation trajectory. The first stage was financial electronification — the automation of bookkeeping, clearing and branch operations from the 1990s onwards. The second was internet finance, which from around 2013 brought mobile payments, online wealth management and platform-based lending to mass-market scale. The third, financial digitalisation, saw institutions rebuild core systems around data, cloud and APIs. Since around 2023, the sector has entered a fourth stage: the era of financial foundation models and AI agents, in which large language models ("LLMs") are no longer peripheral productivity tools but are being embedded into credit, trading, advisory, underwriting and claims workflows.

> "Chinese deployment has converged on the global enterprise stack — foundation model + knowledge base + AI agent + business systems. What sets China apart is the configuration, not the architecture.

This transition enjoys explicit top-level policy endorsement. The General Office of the State Council's Guiding Opinions on Advancing the "Five Major Articles" of Finance (Guo Ban Fa [2025] No. 8, March 2025)[1] elevated digital finance — alongside technology finance, green finance, inclusive finance and pension finance — to the status of national strategy, requiring financial institutions to accelerate digital transformation and calling for a sound digital finance governance system in which innovative businesses are brought within the regulatory perimeter. The National Financial Regulatory Administration ("NFRA") had earlier issued its own implementation opinions on the "Five Major Articles" (Jin Fa [2024] No. 11)[2], calling on banking and insurance institutions to advance digital transformation and encouraging technologically leading institutions to export risk-control tools and technology services to smaller institutions. Within this framework[2], AI is positioned not merely as an efficiency tool but as a component of the financial system's infrastructure.

Market landscape and infrastructure characteristics

Across leading Chinese financial institutions, deployment has converged on a four-layer stack — "foundation model + knowledge base + AI agent + business systems" — that mirrors the global enterprise norm for generative AI. What distinguishes China is not the architecture itself but the choices made within it. Rather than training general-purpose models from scratch, institutions typically deploy domestic open-weight or commercial foundation models (including DeepSeek, Qwen, GLM, Baichuan and Wenxin) on private infrastructure — frequently combining domestic accelerators with still-substantial imported GPU capacity; enhance them with proprietary financial data, knowledge graphs and retrieval-augmented generation ("RAG"); orchestrate task execution through agent frameworks; and wire the outputs into core banking, trading or claims systems. State-owned banks and the larger joint-stock banks lead adoption, while city and rural commercial banks, working with tighter budgets and scarcer talent, have pursued divergent paths. Many begin with narrow, single-scenario pilots, while some have scaled to broader deployment—for instance, Jiangsu Jiangnan Rural Commercial Bank now covers more than 30 scenarios. Regulatory guidance encourages larger institutions to export computing power and models to smaller peers, though concrete cases of such exports remain scarce in public reporting.

"Chinese deployment has converged on the global enterprise stack — foundation model + knowledge base + AI agent + business systems. What sets China apart is the configuration, not the architecture: domestic foundation models, private on-premises deployment, and mandatory human review in consequential decisions."

Three structural features distinguish the Chinese market. First, private, on-premises deployment is the dominant norm — driven above all by data-localisation and financial data-security compliance requirements rather than by unconstrained technical preference; regulators require critical data to remain onshore and expressly encourage private deployment. Second, the policy orientation towards autonomous and controllable technology stacks is unambiguous: adoption of domestic chips and operating systems is advancing under regulatory pressure, but progress is uneven across institutions, core computing power still relies substantially on imported GPUs, and self-controllability remains a work in progress rather than an accomplished fact. Third, human review is systematically retained in consequential customer-facing decisions — initially a prudent response to model-hallucination risk, as Section 5 explains, has now been codified as a regulatory mandate.

Use Cases & Institutional Practices

Banking: from intelligent customer service to intelligent credit

Banking applications have moved well beyond chatbots. In credit, AI now supports automated underwriting for micro and consumer loans — with human review retained at defined checkpoints — while intelligent customer service handles the bulk of routine interactions.

Industrial and Commercial Bank of China ("ICBC") offers the flagship example. Its enterprise-level, hundred-billion-parameter financial foundation model system, "ICBC Zhiyong" (工银智涌), had empowered more than 20 major business lines and over 200 scenarios, with cumulative calls exceeding one billion, according to its 2024 annual report[3]; its 2025 interim report disclosed further deployment progress, including AI wealth assistants and investment research agents[4]. China CITIC Bank has built a three-in-one AI system combining its decision-making AI "CITIC Brain (中信大脑)" — deployed across a large number of business scenarios including credit decisioning, wealth management, risk compliance and operations[5] — with its generative "Cangjie (仓颉)" LLM, supporting an additional package of applications across wealth management, customer service, marketing, advisory and risk compliance. China Merchants Bank deploys the "Zhao Xiao Cai (招小财)" AI assistant in transaction banking — and has reported similar response accuracy on routine transaction-banking queries — and is reported to be applying generative-LLM capability to capital management and product pricing.

Securities and asset management: research, advisory and AI-native apps

In the securities sector, adoption has shifted from embedded tools to AI-native applications. Huatai Securities launched "AI Zhangle" (AI涨乐) — what the company describes as the industry's first AI-native trading application — in October 2025[6]: rather than layering AI functions onto a conventional app, it rebuilds the service logic around AI, embedding large models into stock selection, analysis, trading and alerting through multiple AI investment-assistant personas capable of executing complex operations via dialogue. Guotai Haitong Securities, pursuing an "All in AI" strategy, built the "Junhong Lingxi (君弘灵犀)" hundred-billion-parameter multimodal securities model[7] and, in July 2025, launched the securities industry's first all-AI application, "Lingxi (灵犀)" — with version 2.0 following in December 2025. In March 2026 the Junhong Lingxi model was reported to have received the top tier in the China Academy of Information and Communications Technology's ("CAICT") assessment of LLM capability in typical financial scenarios[8].

> Chinese brokerages have moved from embedding AI tools into existing apps to launching AI-native applications — a shift from 'functions looking for users' to 'AI delivering services'.

"Chinese brokerages have moved from embedding AI tools into existing apps to launching AI-native applications — a shift from 'functions looking for users' to 'AI delivering services'."

These deployments operate within the existing perimeter of investor suitability, advisory licensing and distribution rules — the use of an LLM interface does not exempt an institution from suitability management, record-keeping or content-review obligations, which apply on a technology-neutral basis.

Insurance and payments: underwriting, claims and anti-fraud

Insurance shows perhaps the deepest operational integration. Under Ping An Group's "AI in All" strategy, the group has reported an internal portfolio of 67 vertical foundation models (some at or above the hundred-billion-parameter scale) and over 23,000 deployed AI-agent applications, with company-disclosed adoption metrics across its approximately 230,000-strong workforce and more than 650 operational scenarios in the first half of 2025[9]. Its 2025 annual results disclose that AI agents handled 1.702 billion customer service interactions, covering 80% of total service volume[10]; 93% of policies issued through the auto dealer channel are now underwritten intelligently within one minute[11]; and AI-powered anti-fraud systems intercepted fraudulent claims, reducing losses by RMB 10.51 billion in 2025 — the third consecutive year above RMB 10 billion[12]. China Pacific Insurance (CPIC) P&C developed "Lingxi (灵析)" — what the company describes as the industry's first claims AI employee integrating workflow assistance, risk alerting and quality management across the full claims chain[13]: it extracts photo and document information and generates assessment reports in approximately one minute, with material reductions in manual data-entry volume and gains in pilot-team productivity as reported by the company.

> "Insurance shows the deepest operational integration: AI agents now handle the bulk of customer service, underwrite the majority of policies in under a minute, and intercept billions in fraudulent claims.

In payments, AI underpins real-time fraud interception, merchant risk scoring and customer service at massive scale, building on the infrastructure of the mobile payment era.

Value Creation vs. Legal & Compliance Risks

Core value

The value case is consistent across segments: efficiency in unstructured information processing (research reports, disclosures, claims documents); stronger risk early-warning and anti-fraud capability; lower operating costs (ICBC's digital workforce alone is reported to approximate tens of thousands of person-years of capacity); and financial inclusion, as automated credit scoring extends lending to small businesses and underbanked consumers.

Core legal and compliance risks

Data security and personal information protection. Financial data is heavily regulated: sensitive personal information (expressly including financial accounts) requires separate consent; the "minimum necessary" principle, together with the classification and grading requirements for important data under the Data Security Law, constrains training-data assembly; and cross-border data transfer ("CBDT") restrictions shape — and often rule out — architectures involving offshore models or group-level platforms.

Algorithm governance and model defects. The opacity of complex models ("black box") collides with explainability expectations; LLM hallucination creates mis-selling and misinformation risks in advisory contexts; and algorithmic discrimination in credit and pricing engages both the algorithm rules[14] and PIPL Section 24[35] on automated decision-making.

Intellectual property and data provenance. The legality of training corpora, licensing risks in third-party API arrangements, and the contested copyright status of AI-generated content[17] all remain live issues in Chinese law.

Novel cyber and systemic risks. Prompt injection, model poisoning and supply-chain compromise feature expressly in the latest regulatory guidance[34]; at the market level, homogeneous models raise concerns about convergent strategies, herding and the amplification of volatility — a systemic-risk framing now explicitly addressed by the financial regulator[21] (See Section 4 China's AI Regulatory & Legal Framework).

China's AI Regulatory & Legal Framework

China has no single comprehensive AI statute. Instead, obligations arise from four layers of rules, which stack upon one another. For FS institutions the practical question is never whether a single rule applies, but how the layers interact: a customer-facing chatbot at a bank may simultaneously be a "generative AI service" under the Cyberspace Administration of China ("CAC") rules[16], a "high-risk application" under the NFRA's June 2026 guidance[21], an "automated decision" under the Personal Information Protection Law ("PIPL")[35], and a regulated advisory activity under securities law. This section maps each layer and — critically — how each connects to financial regulation and practice.

Layer One: horizontal AI and algorithm governance (led by the CAC)

Administered principally by the CAC, this layer targets AI techniques directly. Four instruments form the core, and each has a distinct financial-services interface:

Algorithm Recommendation Provisions (effective 1 March 2022)[14] — China's first algorithm-specific regulation requires service providers with "public opinion attributes or social mobilisation capacity" to complete an algorithm filing, grants users the right to opt out of personalised recommendation, and prohibits unreasonable differential treatment in trading conditions. For FS institutions, this reaches personalised product recommendation in banking and wealth-management apps, robo-advisory distribution logic, and "big-data-enabled price discrimination" in loan pricing and insurance quotation — practices that overlap with, and are reinforced by, PIPL Section 24 and financial consumer-protection rules.

Deep Synthesis Provisions (effective 10 January 2023)[15] — impose labelling, real-name verification, biometric consent and security assessment duties on providers of deep synthesis technology (synthetic media, digital humans, voice synthesis). These rules cut both ways for finance: they govern institutions' own deployments — digital human tellers and livestream anchors, synthetic voice in telemarketing and collections — and they supply part of the legal toolkit against deepfake-enabled fraud targeting facial-recognition onboarding and voice authentication.

Interim Measures for Generative AI Services (effective 15 August 2023)[16] — the centrepiece. The Measures regulate GenAI services offered to the public; research, development and use that does not offer services to the domestic public are expressly outside their scope. Providers of public-facing services must ensure the legality of training data, protect IP and personal information, prevent discrimination, maintain content integrity, handle complaints, and — where the service has public opinion attributes — undergo security assessment and algorithm filing. The internal-use carve-out is the pivot of every FS deployment analysis: an internal coding assistant or document-summarisation tool generally sits outside the regime, whereas a customer-facing chatbot, AI research generator or marketing-copy engine falls squarely within it — and institutions are well advised to treat borderline "employee-facing but customer-impacting" deployments conservatively.

> China has no single comprehensive AI statute. Instead, obligations arise from four layers of rules that stack upon one another — and for financial institutions the practical question is never whether a single rule applies, but how the layers interact.

AI-Generated and Synthesized Content Labelling Measures (effective 1 September 2025)[17] — require dual explicit and implicit labelling of AI-generated content (visible markers plus embedded metadata), prohibit tampering with labels, and oblige dissemination platforms and app distribution channels to verify labelling. A companion mandatory national standard, GB 45438-2025 (Cybersecurity Technology — Labelling Method for AI-Generated Synthetic Content), supplies the technical specifications. In finance, this bears directly on AI-generated marketing materials, research summaries, digital-human video and customer communications — and creates a review checkpoint for compliance and investor-education content pipelines.

The filing regime has scaled rapidly: across CAC monthly publications the population of filed GenAI services grew from 346 in March 2025 toward 868 in April 2026, with a parallel group of registered AI applications and functions[18]; mid-2026 the same filing regime was extended to on-device large models, with the first cohort of smartphone-side generative AI services — including Apple Intelligence — recorded as filed[19]. The threshold question for an FS institution is therefore factual and structural: is the deployment public-facing, does it rely on an externally sourced model (and is that model filed?), and which entity in the group is the "provider"? As this Section shows, the financial regulator has now imported these thresholds directly into sectoral supervision[33].

> "The NFRA's June 2026 AI Guiding Opinions convert high-level principles into concrete, examinable expectations — requiring board-level governance, classified risk tiers, formal admission for high-risk applications, and mandatory human review at key decision nodes.

Layer Two: financial-sector rules (NFRA, PBOC, CSRC)

The policy trajectory. Sectoral AI governance did not appear overnight. The PBOC's FinTech Development Plan (2022–2025) first called for algorithm management rules, model security assessment and audit systems, disclosure of algorithmic decision logic[39], and stress-testing against black-box, herding and discrimination risks. The "Five Major Articles" framework then made digital finance a strategic priority, and the NFRA's December 2025 implementation plan for high-quality digital finance in banking and insurance committed the regulator to building a classified and graded management framework for secure AI application, with human-intervention mechanisms covering important business processes and key nodes[20]. That commitment crystallised six months later.

The NFRA's AI Guiding Opinions (June 2026) — the sectoral centrepiece[21]. On 18 June 2026, the NFRA issued the Guiding Opinions on the Secure Development and Application of Artificial Intelligence in Banking and Insurance (Jin Fa [2026] No. 8, the "AI Guiding Opinions"), setting out 32 guidance items across governance, development and application, data governance, computing power, risk management, capability building and supervision. The AI Guiding Opinions convert previously high-level principles into concrete, examinable expectations:

Governance: the board must designate a dedicated committee responsible for AI; institutions must establish full-lifecycle management systems covering demand analysis, data preparation, development, deployment, iteration and decommissioning[22];

Classified and graded risk management: AI applications must be inventoried and tiered by business criticality, scale, customer impact and model complexity and dependence — operationalising the December 2025 commitment[23];

High-risk application admission: GenAI applications involving fund transactions, asset valuation, credit approval, underwriting and claims, or risk management — or otherwise directly affecting customers' interests or the conclusion of financial contracts — are deemed high-risk and may be implemented only with the approval of the institution's risk management committee[24];

Human oversight: human supervision and intervention mechanisms must be established at key nodes of high-risk applications, with defined emergency shutdown and model-exit conditions and backup manual processes[25]; where a model's explainability is insufficient, it may be used in high-risk scenarios only as an auxiliary tool, with humans making the final decision; decisions affecting customer rights or carrying material financial impact require manual review nodes;

Model admission and external models: GenAI models are subject to admission management assessing efficacy, security and compliance; externally sourced GenAI models must have completed the CAC filing described in Section 4 — a direct interlock between Layer One and Layer Two that makes the CAC filing status of a vendor's model a gating due-diligence item in every FS procurement[26];

Data and personal information: names, ID numbers, phone numbers and bank card numbers and other personal and privacy data must not be used for GenAI model training or optimisation — a sectoral rule that goes further than PIPL's consent architecture[35,36] and effectively mandates de-identification pipelines before any training or fine-tuning[27];

Computing power: institutions must build secure, autonomous and controllable computing bases; large institutions are encouraged to export computing capacity and mature models to small and medium-sized institutions — an explicit policy lever against a two-tier market[28];

Ethics and fairness: institutions must establish AI ethics review and monitoring systems, and be able to justify the legitimacy of using protected characteristics[29];

Agent-specific security: for AI agents, the AI Guiding Opinions name the failure modes to be controlled — memory poisoning, identity privilege escalation, tool abuse and loss of operational control — and require defences against prompt injection, chain-of-thought injection and multimodal attacks[30];

Transparency and auditability: institutions must set explainability standards for high-risk scenarios, prominently label AI-generated content and proactively inform financial consumers[17]; logs must be retained for no less than the duration of the relevant business; periodic audits of models and algorithms are required[31];

Outsourcing and supply chain: named-list management of outsourcing partners, internal evaluation frameworks for external models, concentration-risk management and open-source component review, with express warnings against supply-chain poisoning[32];

Reporting duty: institutions deploying GenAI in public-facing services or high-risk scenarios must report to the NFRA or its local offices — note how the "public-facing" trigger borrows the CAC's threshold vocabulary[33].

The AI Guiding Opinions also address market-level risk directly: institutions must guard against convergent investment strategies and the amplification of market volatility, and are prohibited from abusing AI to generate false information or manipulate prices[34]. The PBOC's FinTech Development Plan had already flagged herding risk among the dangers to be stress-tested; the AI Guiding Opinions go a step further, converting model-homogeneity risk into concrete supervisory obligations — a signal with obvious relevance for securities and asset-management business supervised by the CSRC as well.

Securities side. The CSRC has not issued an AI-specific rule; its IT management, algorithmic trading and outsourcing rules apply a technology-neutral lens, with suitability, record-keeping and content obligations applied to AI-generated research and advisory outputs. In practice, securities firms face the same Layer One duties as everyone else (GenAI filing for public-facing apps[16,17], labelling of AI-generated content), while the AI-native trading and advisory applications described in Section 2 remain subject to the full weight of investor-suitability and licensing rules.

> “
>
> Personal and privacy data — names, ID numbers, phone numbers, bank-card numbers — must not be used for GenAI model training or optimisation: a sectoral rule that goes further than PIPL's consent architecture and effectively mandates de-identification pipelines.

Layer Three: the data law trilogy

The Cybersecurity Law, Data Security Law and PIPL form the foundation on which both layers above rest. Four provisions matter most for AI in FS:

First, PIPL Section 24[35]: automated decision-making must be

transparent and produce fair and impartial results; unreasonable differential treatment in "transaction conditions, including price" is prohibited; where automated decisions materially affect an individual's rights, the individual has the right to an explanation and the right to refuse a decision made solely by automated means. Functionally, this captures credit scoring, automated underwriting, dynamic pricing — the exact scenarios the NFRA has now designated as high-risk[24] — and PIPL separately requires a prior personal-information-protection impact assessment for automated decision-making.

> "Even with consent, sectoral guidance bars the most identifying personal data fields from GenAI training corpora — creating a two-level constraint that goes beyond PIPL's general rules and reinforces the dominance of domestic models and private on-premises deployment.

Secondly, PIPL's sensitive-information rules [36]: financial accounts, identity documents and personal property status are expressly enumerated as sensitive personal information, processing of which requires a specific purpose, sufficient necessity and separate consent. Combined with the NFRA's prohibition on using names, ID numbers, phone numbers and bank card numbers in GenAI training[27], this creates a two-level constraint: even with consent, sectoral guidance now bars the most identifying fields from training corpora altogether.

Thirdly, the Network Data Security Management Regulation (State Council Decree No. 790, effective 1 January 2025)[37] consolidates data-handling obligations across the trilogy: it reiterates separate consent for sensitive personal information[36], obliges providers of GenAI services to strengthen training-data security management[16], and imposes annual risk-assessment and reporting duties on handlers of important data. For financial institutions — archetypal handlers of important data — this adds a recurring, board-visible compliance cycle that intersects with the NFRA's own periodic audit expectations[31].

Fourthly, the CBDT regime (security assessment, standard contract, or certification) shapes AI architecture choices: architectures that send prompts or customer data to offshore model APIs, or that consolidate data on group-level platforms abroad, will frequently be impracticable or require case-by-case structuring — a principal reason domestic models and private deployment dominate Chinese FS practice (Section 1), and a live issue for multinational institutions running global AI stacks.

Layer Four: technical standards and industry norms

The PBOC's financial industry standards supply the technical grammar for the rules above. JR/T 0221—2021, the Evaluation Specification of AI Algorithms in Financial Applications[38], sets the evaluation dimensions, including security, explainability, accuracy and performance — against which financial AI algorithms are assessed. JR/T 0287—2023, the Guidelines for Information Disclosure of AI Algorithms in Financial Applications[39], prescribes disclosure principles, forms and content across seven aspects — algorithm combination, algorithm logic, algorithm application, algorithm data, the responsible entity, algorithm changes and algorithm audits — to enhance explainability and protect financial consumers' right to know. Together with the mandatory national standard GB 45438-2025 on content labelling[45], these standards convert statutory abstractions ("transparency", "explainability") into testable engineering requirements. The NFRA's AI Guiding Opinions contemplate a further technical framework and security standards for GenAI in banking and insurance, to be developed with relevant departments — expect sector-specific GenAI security standards to follow[40].

Core Regulatory Principles & Operational Compliance Roadmap

Four core principles of Chinese AI-in-finance regulation

Balancing development and security. Adoption is state policy; the filing, assessment and labelling infrastructure is designed to enable adoption safely, not to impede it. The "AI Plus" Initiative (State Council Document No. 11 of 2025)[41] expressly names finance among the sectors for wide deployment of smart terminals and AI agents, targeting penetration of next-generation intelligent terminals and AI agents in over 70% of applications by 2027.

"Who uses, who is responsible"[42]. The financial institution — as service provider and technology user — bears ultimate

responsibility, regardless of outsourcing or third-party models[32]; internal accountability must be traceable at every node.
Full-lifecycle governance[43]. From development and testing through deployment, monitoring and exit, each stage carries defined controls: admission management, classified grading[23], human intervention[25], logging and audit[31], and decommissioning conditions.

Human-machine collaboration and human final decision[44]. Human-in-the-loop is not aspirational but codified: insufficiently explainable AI in high-risk scenarios is auxiliary only, and decisions affecting customer rights or with material financial impact require human review nodes[25].

"'Who uses, who is responsible': in China, the financial institution bears ultimate responsibility for its AI — outsourcing and third-party models dilute nothing[32]."

A compliance roadmap for deploying AI in Chinese financial institutions

Board level. Articulate AI principles reflecting the institution's values and risk appetite; designate the dedicated board committee contemplated by the AI Guiding Opinions[22]; approve high-risk deployments through the risk management committee[24].

Governance framework. Integrate AI strategy into overall risk management; establish an AI risk committee or equivalent; maintain a classified and graded AI application inventory[23]; define accountability and escalation lines; put ethics review on a standing footing[29].

Operational processes. Run use-case and impact assessments before deployment (mapping each use case against the CAC's "public-facing" threshold[16] and the NFRA's high-risk catalogue[24]); adopt policies for labelling[17], disclosure and consumer notification consistent with the Labelling Measures and JR/T 0287—2023[39]; build logging and information-management systems meeting the retention expectations[31] (no less than the business duration); train staff; and maintain continuing communication with regulators — including the NFRA reporting duty for public-facing or high-risk GenAI applications[33].

Model deployment. Decide deployment and compute architecture (on-premises vs. cloud) against data localisation and CBDT rules; map data flows, with particular care that names, ID numbers, phone numbers, bank card numbers and other personal and privacy data are excluded from GenAI training[27]; conduct supply-chain due diligence on procured models — verifying CAC filing status for external GenAI models[26] — and allocate responsibilities, audit rights and exit arrangements contractually[32].

Future Outlook & Strategic Conclusion

Several trajectories appear durable. First, financial foundation models are becoming general-purpose industry infrastructure, reshaping how products and services are delivered, how institutions operate and how risk is managed — the shift from "+AI" to "AI+", in the phrasing now common among Chinese bank technology leaders. Second, human-AI collaboration is likely to remain the dominant working model for the long term — a prediction now embedded in regulation[44] itself, which reserves final decision-making to humans in high-risk scenarios[25]. Third, the risk agenda — accuracy and explainability, privacy and IP, ethics and bias — will continue to demand proactive governance: risk assessment, human oversight and transparent communication.

"

'Who uses, who is responsible': in China, the financial institution bears ultimate responsibility for its AI — outsourcing and third-party models dilute nothing.

Two further developments warrant close monitoring. On the legislative front, the NPC Standing Committee's 2026 legislative

work plan retains AI legislation as a preparatory review item, while the State Council's 2026 legislative plan calls for improving AI governance and accelerating comprehensive legislation for the healthy development of AI — suggesting that today's layered, "small-incision" rules will progressively consolidate. On the international front, China is investing in global AI governance — the Global AI Governance Action Plan issued at the 2025 World Artificial Intelligence Conference, and the proposed World AI Cooperation Organization — which may shape future cross-border regulatory convergence relevant to multinational FS institutions.

> "Watch the consolidation: China is moving from fragmented, sector-specific rules towards comprehensive AI legislation — and from technique regulation towards a risk-tiered, full-lifecycle model already previewed by the financial regulator.

"Watch the consolidation: China is moving from fragmented, sector-specific rules towards comprehensive AI legislation — and from technique regulation towards a risk-tiered, full-lifecycle model already previewed by the financial regulator[43]."

For financial institutions, the strategic conclusion is straightforward: continuous monitoring of regulatory change is now a compliance function in its own right. Institutions that have already built use-case inventories, classification and grading[23], human oversight mechanisms[25], model risk management[38,39] and data governance[36,37] will find themselves substantially compliant with whatever form the eventual comprehensive statute takes — and best placed to capture the productivity gains that Chinese FS institutions are already reporting at scale.

1. General Office of the State Council – Guiding Opinions on Advancing the "Five Major Articles" of Finance (Guo Ban Fa [2025] No. 8, 5 March 2025). See: https://www.gov.cn/zhengce/zhengceku/202503/content_7010605.htm
2. NFRA, Implementation Opinions on the "Five Major Articles" for Banking and Insurance (Jin Fa [2024] No. 11). See: https://www.nfra.gov.cn/cn/view/pages/governmentDetail.html?docId=1161211&itemId=861&generaltype=1
3. ICBC – 2024 Annual Report. See: https://v.icbc.com.cn/userfiles/resources/icbcltd/download/2025/ReportACh20250425.pdf
4. ICBC – 2025 Interim Report. See: https://v.icbc.com.cn/userfiles/resources/icbcltd/download/2025/Announcement202509242.pdf
5. China CITIC Bank –2024 Annual Report. See: https://www.citicbank.com/about/investor/financialaffairs/report/2024/202504/W020260408624632033941.pdf
6. China Daily – Huatai Securities, in partnership with Volcano Engine, has created China's first AI-native trading app, "AI Zhangle". See: https://ex.chinadaily.com.cn/exchange/partners/82/rss/channel/cn/columns/snl9a7/stories/WS692eabc9a310942cc49948fb.html
7. Cailian Press – The securities industry's first fully AI-powered intelligent app has been launched! Guotai Haitong Lingxi App creates ten different. See: https://www.cls.cn/detail/2094936
8. CAICT – March 2026 evaluation of LLM application capability in typical financial scenarios. See: https://www.jfdaily.com/sgh/detail?id=1718651
9. Xinhua Agency – Ping An Insurance: Deepening AI Empowerment and Fulfilling the Mission of "Finance for the People". See: http://www.news.cn/money/20251110/695c9d447b764585981a0d1fe54fde20/c.html
10. Ping An Group – 2025 Annual Results. See: https://www.pingan.com/app_upload/images/info/upload/17e8f0e4-4bd5-4f51-8415-037439c35ec6.pdf
11. Ping An Property & Casualty – Ping An Unveils Innovative AI Solutions in Healthcare, Insurance and Payments at WAIC 2026. See: https://markets.ft.markitdigital.com/data/announce/detail?dockey=600-202607210138PR_NEWS_USPRX____CN08814-1
12. Ping An Group – Technology Enablement as a Core Engine of Service Innovation. See: https://group.pingan.com/about_us/our_businesses/technology.html
13. CPIC P&C – From "trial run" to "deep cultivation," the property insurance industry enters the era of intelligent claims settlement. See: http://m.10jqka.com.cn/20260331/c675655142.shtml
14. CAC, MIIT, MPS & SAMR – Administrative Provisions on Algorithm Recommendation for Internet Information Services (CAC Order No. 9), effective 1 March 2022, sec. 21. See: http://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm
15. CAC – Provisions on the Administration of Deep Synthesis Technology, effective 10 January 2023. See: http://www.cac.gov.cn/2022-12/11/c_1672221949354811.htm
16. CAC – Interim Measures for the Management of Generative AI Services, effective 15 August 2023, sec. 2. See: http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm
17. CAC, MIIT, MPS & NRTA – Measures for Labelling AI Generated Synthetic Content, effective 1 September 2025, sec. 3. See: https://www.cac.gov.cn/2025-03/14/c_1743654684782215.htm
18. CAC – Announcement Regarding the Release of Filing Information for Generative Artificial Intelligence Services (January to March 2025). See: https://www.cac.gov.cn/2025-04/08/c_1745817775881843.htm
19. CAC – Announcement Regarding the Release of Filing Information for 7 Mobile-Side Generative Artificial Intelligence Services, published 15 July 2026. See: https://www.cac.gov.cn/2026-07/15/c_1785861480767004.htm
20. NFRA – Implementation Plan for High Quality Development of Digital Finance in Banking and Insurance, December 2025. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1239741&itemId=928
21. NFRA – Guiding Opinions on the Secure Development and Application of Artificial Intelligence in Banking and Insurance (Jin Fa [2026] No. 8), 18 June 2026. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1261784&itemId=928
22. NFRA – AI Guiding Opinions, sec. 1. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1261784&itemId=928
23. NFRA – AI Guiding Opinions, sec. 15. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1261784&itemId=928
24. NFRA – AI Guiding Opinions, sec. 16. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1261784&itemId=928
25. NFRA – AI Guiding Opinions, sec. 17. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1261784&itemId=928
26. NFRA – AI Guiding Opinions, sec. 5. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1261784&itemId=928
27. NFRA – AI Guiding Opinion, sec. 24. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1261784&itemId=928
28. NFRA – AI Guiding Opinions, sec. 12. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1261784&itemId=928
29. NFRA – AI Guiding Opinion, sec. 23. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1261784&itemId=928
30. NFRA – AI Guiding Opinions, sec. 25. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1261784&itemId=928
31. NFRA – AI Guiding Opinions, sec. 21, 22. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1261784&itemId=928
32. NFRA – AI Guiding Opinions, sec. 18, 19. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1261784&itemId=928
33. NFRA – AI Guiding Opinions, sec. 29. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1261784&itemId=928
34. NFRA – AI Guiding Opinions, sec. 14. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1261784&itemId=928
35. NPCSC – Personal Information Protection Law of the People's Republic of China, sec. 24. See: http://www.npc.gov.cn/npc/c2/c30834/202108/t20210820_313088.html
36. NPCSC – Personal Information Protection Law of the People's Republic of China, sec. 28–29. See: http://www.npc.gov.cn/npc/c2/c30834/202108/t20210820_313088.html
37. State Council – Network Data Security Management Regulation (Decree No. 790), effective 1 January 2025. See: https://www.gov.cn/gongbao/2024/issue_11646/202410/content_6980863.html
38. PBOC – Financial Industry Standard JR/T 0221—2021 – Evaluation Specification of AI Algorithms in Financial Applications. See: https://cfstc.pbc.gov.cn/bzgk/detail/?id=0&bzId=1912
39. PBOC – Financial Industry Standard JR/T 0287—2023 – Guidelines for Information Disclosure of AI Algorithms in Financial Applications. See: https://std.samr.gov.cn/hb/search/stdHBDetailed?id=0E01B7DCBB0C20A2E06397BE0A0A0C7F
40. NFRA – AI Guiding Opinions, sec. 28. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1261784&itemId=928
41. State Council – Opinions of the State Council on Deepening the Implementation of the "Artificial Intelligence+" Action (Guo Fa [2025] No. 11). See: https://www.gov.cn/zhengce/content/202508/content_7037861.htm
42. NFRA – AI Guiding Opinions, sec. 1. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1261784&itemId=928
43. NFRA – AI Guiding Opinions, sec. 2. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1261784&itemId=928
44. NFRA – AI Guiding Opinions, sec. 17, 22. See: https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1261784&itemId=928
45. CAC – Expert interpretation of mandatory national standard GB 45438-2025, Cybersecurity Technology – Labelling Method for AI-Generated Synthetic Content, September 2025. See: https://www.cac.gov.cn/2025-09/05/c_1758792061408012.htm

金杜律师事务所
KING&WOOD

BUILT IN CHINA
CONNECTED TO THE WORLD

A firm born in Asia, underpinned by world class capability. With over 2,300 lawyers in 23 global locations, we draw from our Western and Eastern perspectives to deliver incisive counsel.

We are focused on our clients – people and organisations with distinctive ambitions and challenges. We are driven to understand your needs, solve your problems and unearth the right opportunities for you. Whether you're expanding globally or strengthening locally, our service style is dynamic, insightful, and tailored for you.

AWARDS & RANKINGS

Chambers FinTech Guide 2026
FinTech Legal (PRC firms)
Band 1

2025 The Legal 500
FinTech (PRC firms)
Tier 1

Chambers Greater China Region Guide 2026
Banking and finance
Band 1

2026 The Legal 500
Banking and finance (PRC firms)
Tier1

Chambers Global Guide 2026
Artificial Intelligence
Global Market Leaders

Chambers Greater China Region Guide 2026
Intellectual Property
Law Firm of the Year

Chambers Greater China Region Guide 2022-2026
Intellectual Property (PRC-Firms): Litigation & Non-litigation
Band 1

2024-2026 The Legal 500
Intellectual Property (PRC Firms): Contentious & Non-Contentious
Tier1

Chambers Greater China Region Guide 2023-2026
TMT: Data Protection & Privacy (PRC Firms)
Band 1

Beijing · Shanghai · Hangzhou · Nanjing · Suzhou · Wuxi · Guangzhou · Shenzhen · Hong Kong · Hengqin · Haikou · Sanya · Chengdu · Chongqing · Qingdao · Jinan · Changchun · Tokyo · New York · Silicon Valley · Los Angeles · Vancouver

GERMANY

ARTIFICIAL INTELLIGENCE IN THE BANKING AND FINANCIAL SERVICES SECTOR – REGULATIONS GERMANY

MÖHRLE HAPP LUTHER

Dr. Joachim Jung
Partner

j.jung@mhl.de

040 85 301 - 0

www.mhl.de

BIO

Dr. Joachim Jung is a Partner at Möhrle Happ Luther and specialises in AI, data protection, intellectual property and emerging technologies. He advises established companies, high-growth businesses and investors on the legal and regulatory challenges arising from digital transformation, AI deployment and data-driven business models. As a Certified Specialist in Intellectual Property Law and Certified Data Protection Officer, he combines deep expertise in technology, privacy and regulatory compliance. Dr. Jung regularly supports clients in highly regulated industries, including financial services, healthcare and digital platforms, where innovation, data governance and AI regulation increasingly intersect.

MÖHRLE HAPP LUTHER

MÖHRLE HAPP LUTHER

Axel von Goldbeck
Partner

a.vongoldbeck@mhl.de

030 226 288 - 0

www.mhl.de

BIO

Axel von Goldbeck is a Partner at Möhrle Happ Luther, focusing on financial regulation, capital markets, digital assets and blockchain-based financing structures. He advises financial institutions, corporates, fintechs and investors on innovative financing solutions, tokenisation projects and regulatory frameworks for digital business models. Before joining Möhrle Happ Luther, he held senior positions at J.P. Morgan, White & Case, Luther and DWF Germany and served as Chief Executive of the German Property Federation (ZIA). With extensive experience at the intersection of finance, regulation and emerging technologies, he regularly advises on tokenised securities, crypto assets and digital capital markets.

MÖHRLE
HAPP
LUTHER

ARTIFICIAL INTELLIGENCE IN THE BANKING AND FINANCIAL SERVICES SECTOR – REGULATIONS GERMANY

MÖHRLE HAPP LUTHER

Introduction

The use of artificial intelligence (AI) in the banking and financial services sector has undergone dynamic development in recent years. AI applications have evolved from experimental technologies into integral components of the business model of many financial institutions. According to recent reports by the Association of German Banks, more than two thirds of banks in the German market already use AI applications. The financial industry expects AI to generate a significant increase in revenue over the coming years, with a 22 percent reduction in operational costs forecast by 2030. This technological transformation is, however, accompanied by a parallel development of the regulatory framework, which aims to harness the opportunities offered by AI while simultaneously controlling the associated risks. This article provides an overview of current and emerging legal developments at European and national level.

> “
>
> More than two thirds of banks in the German market already use AI applications.

European Regulatory Frameworks

The provisions of the AI Regulation do not all apply at once. Instead, they are being phased in over several years. The first set of rules, covering general principles and the prohibition of certain unacceptable AI practices, has applied since February 2025. A second set of rules, covering AI models with broad, general-purpose capabilities and related oversight arrangements, has applied since August 2025. The remaining rules, including the detailed requirements for high-risk AI systems, were originally due to apply from August 2026 (or, for certain AI embedded in already-regulated products, from August 2027). However, a subsequent EU amendment that took effect in July 2026 (commonly referred to as the “Digital Omnibus on AI”) pushed these later deadlines back further: the requirements for high-risk AI systems generally now apply from December 2027, and those for AI embedded in regulated products now apply from August 2028. The earlier rules on general principles, prohibited practices and general-purpose AI models were not affected by this postponement and remain in force on their original timeline.

The AI Regulation follows a risk-based approach that distinguishes between different categories of AI systems. Article 5 generally prohibits certain AI practices, including systems for subliminal influence, the exploitation of vulnerabilities of certain groups of persons, social scoring systems, and AI systems for the risk assessment of natural persons with regard to criminal offences. These prohibited practices have already been banned since 2 February 2025.

Of particular relevance to the financial sector is the classification of AI systems as “high-risk AI systems” pursuant to Article 6 in conjunction with Annex III of the AI Regulation. The AI Regulation classifies, for example, AI-supported creditworthiness assessments as a high-risk area under Article 6(2) in conjunction with Annex III paragraph 5(b). High-risk AI systems are subject to strict requirements under Articles 8 to 15 of the AI Regulation, including the establishment of a risk management system, data governance requirements, technical documentation, record-keeping through automatic logging, transparency and provision of information to deployers, human oversight, and accuracy, robustness and cybersecurity. Providers and deployers are subject to further obligations under Articles 16 to 27, including a quality management system, conformity assessment, EU database registration under Article 49, and, for certain deployers, a fundamental rights impact assessment under Article 27. As set out above, these obligations for stand-alone high-risk systems under Annex III, originally scheduled to apply from 2 August 2026, now apply from 2 December 2027 following the Digital Omnibus on AI.

Another important European development is Regulation (EU) 2022/2554 on digital operational resilience for the financial sector (the "DORA Regulation"), which became fully applicable on 17 January 2025. DORA establishes a comprehensive framework for ICT risk management in the financial sector, which also covers the use of AI (Art. 64 DORA Regulation). The Federal Financial Supervisory Authority (BaFin) has published guidance in this regard to support financial undertakings in implementing the DORA requirements when using AI. With the implementation of DORA, BaFin's circular on the "Supervisory Requirements for IT in Financial Institutions" (BAIT) was repealed with effect from 17 January 2025.

At European level, enforcement of the AI Regulation is the responsibility of the newly established European Artificial Intelligence Office, which is located within the European Commission (Commission Decision of 24 January 2024 establishing the European Artificial Intelligence Office). This office develops the Union's expertise and capabilities in the field of the AI Regulation and performs tasks relating to market surveillance and control of general-purpose AI systems.

National Developments in Germany

At national level, Germany has taken an important step towards implementing European requirements with the Act on the Digitalisation of the Financial Market (FinMaDiG). The FinMaDiG entered into force on 30 December 2024 and consolidates the necessary provisions for implementing the European regulations in the field of digital finance. To implement the DORA Regulation and the DORA Directive, targeted amendments were made to the German Banking Act and to a number of other sector-specific statutes.

As early as 2021, BaFin made an important contribution to the regulation of AI use in the financial sector with its "Principles for the Use of Algorithms in Decision-Making Processes". These principles set out preliminary considerations regarding minimum supervisory requirements for the use of AI and serve as guidance for financial market participants supervised by BaFin. In this context, BaFin defines AI as a combination of big data, computing resources and machine learning.

The BaFin principles include overarching principles such as the clear responsibility of the management board, adequate risk and outsourcing management, the avoidance of bias in algorithm-based decision-making processes, and the exclusion of legally prohibited differentiations. The management board is responsible for company-wide strategies and guidelines on the use of algorithm-based decision-making processes and for establishing risk management adapted to the use of AI.

BaFin has also published guidance on ICT risks in the use of AI in financial undertakings, which is intended in particular to support CRR institutions and Solvency II insurance undertakings in implementing the DORA requirements. This guidance addresses ICT risk management and ICT third-party risk management, including the Delegated Regulation on ICT risk management and the Delegated Regulation on the subcontracting of ICT services.

BaFin defines AI as a combination of big data, computing resources and machine learning.

The German legislative process implementing the AI Regulation has since been completed. On 29 July 2026, the "Act on the Market Surveillance and Innovation Promotion of Artificial Intelligence" (KI-Marktüberwachungs- und Innovationsförderungsgesetz, "KI-MIG") entered into force. Under the KI-MIG, the Federal Network Agency ("Bundesnetzagentur") is the central coordination and competence centre, notifying authority and, in most sectors not otherwise covered, the competent market surveillance authority, while BaFin retains supervisory responsibility for AI systems used in connection with regulated financial activities of the supervised institutions listed in the Act, and existing sector-specific market surveillance authorities (e.g. in fully harmonised product areas) continue to be used to avoid duplicate structures (one-stop-shop principle).

Areas of Application of AI in the Financial Sector

Financial undertakings use AI along the entire value chain: credit institutions use AI in sales to predict customer churn, while in lending AI applications can support case handlers in examining annual financial statements. In fund management, AI is used to summarise large volumes of analyst reports on investment instruments.

> “A particular problem is the so-called ‘black box’ nature of many AI systems.

Insurance undertakings use interactive AI assistants (chatbots) in sales and customer communications. Product pricing can be tailored more precisely to insured risks by means of complex models, potentially using real-time data (dynamic pricing or telematics). In underwriting, AI applications can assist in assessing risks. Automated input management allows a large number of incoming documents to be routed efficiently. In claims management, insurers use AI applications to support claims handlers in claims settlement, for example in the automated payment of small claims. In benefit processing, AI applications are used for fraud detection.

Specific use cases also include stock market forecasts, market analyses, sentiment analyses, algorithmic trading, portfolio management, risk management, compliance, insolvency forecasts, modelling of creditworthiness assessments (credit scoring and rating), lending, robo-advising, chatbots and virtual assistants. Other typical areas of application include the creation of scores, automated handling of anti-money laundering processes, algorithmic trading and portfolio management.

Across sectors, the use of AI assistants can be observed, most of which are large language models that process and generate unstructured data such as text and images. Such assistance systems can be used broadly to create presentations, program code or videos. The size of the AI fintech market is estimated at USD 44.08 billion in 2024 and is expected to reach as much as USD 50.87 billion by 2029.

Challenges and Risks

The use of AI in the financial sector also entails significant challenges and risks. AI applications carry risks of misuse, breaches of data protection requirements, inadequate data infrastructure, behavioural manipulation of persons, lack of transparency in decision-making, and vulnerability to cyberattacks. A particular problem is the so-called “black box” nature of many AI systems. Since AI can produce results that even developers cannot fully understand, this creates a transparency issue that is particularly critical in the highly regulated financial sector.

AI algorithms also carry the risk of adopting, reinforcing or expanding existing discrimination. This becomes particularly vivid in the assessment of the creditworthiness of natural persons by means of AI. An AI system learns on the basis of the empirical data that its developer provides to it. Persons belonging to a group that has historically had difficulty obtaining credit will initially also be classified by the AI system as risky candidates. This may lead to violations of the General Equal Treatment Act and of fundamental rights.

The AI Regulation addresses these challenges through various requirements. Article 4 requires providers and deployers of AI systems to take measures to ensure, to the best of their ability, that their staff and other persons dealing with the operation and use of AI systems on their behalf have a sufficient level of AI literacy.

Strict transparency obligations apply to high-risk AI systems. Under Article 13 of the AI Regulation, high-risk AI systems must be designed to be sufficiently transparent to enable deployers to interpret and appropriately use the system's output, and providers must supply instructions for use accordingly. Deployers of certain high-risk systems must, under Article 26(11), inform natural persons where the system is used to take or assist in decisions concerning them. In addition, Article 86 grants persons affected by a decision based on the output of certain Annex III high-risk AI systems, where that decision produces legal effects or similarly significantly affects them, the right to obtain from the deployer a clear and meaningful explanation of the role of the AI system in the decision-making procedure and the main elements of the decision taken. The AI Regulation also contains extensive requirements regarding data quality and data governance under Article 10 in order to avoid bias and discrimination.

Data Protection Aspects

The use of AI in the financial sector is also subject to the data protection requirements of the General Data Protection Regulation (GDPR). AI systems often process large quantities of personal data in order to recognise patterns and make predictions. This raises questions regarding the lawfulness of data processing, data subject rights and the transparency of data processing.

In this context, Article 22 GDPR is particularly relevant as it gives individuals the right not to be subject to a decision made purely by a computer, without any human involvement, if that decision has legal or similarly significant effects on them. This applies unless the decision is needed to enter into or carry out a contract, is allowed by law, or the person has clearly agreed to it. Even then, the person must still be able to ask for a human to review the decision, explain their side, and challenge the outcome.

This is highly relevant for AI-supported credit decisions and other automated decision-making processes in the financial sector: in 2023, the Court of Justice of the European Union ruled that automatically calculating a person's credit score already counts as such a decision if a bank relies heavily on that score when deciding whether to grant credit. As of today, the GDPR itself has not been changed to reflect the rise of AI. The EU has proposed adjustments as part of the European Commission's Digital Omnibus proposal of 19 November 2025, but, unlike the AI-Act-specific part of the Digital Omnibus package, these GDPR amendments had not been adopted as at the date of this article and remain under negotiation in the Council and the European Parliament.

In its principles, BaFin emphasises that the focus of the supervisory authorities is not on the algorithm itself, but on the entire decision-making process based on the algorithms. This is also consistent with the data protection approach, which does not regulate the technology as such, but rather its use in the specific context.

“

Automatically calculating a person's credit score already counts as such a decision if a bank relies heavily on that score when deciding whether to grant credit.

Liability Issues

Liability for damage caused by the use of AI in the financial sector is a complex legal question that has not yet been conclusively resolved. The AI Regulation itself does not contain specific liability rules but focuses on preventing harm through the regulation of the development and use of AI systems.

At European level, on 28 September 2022 the Commission presented a proposal for a directive regulating AI liability. This proposal refrains from introducing strict liability for AI and instead provides for rules on the burden of proof for fault-based non-contractual liability claims. This is intended to make it easier for injured parties to assert claims for damages without fundamentally changing the liability system.

In the financial sector, the specific liability rules of the German Banking Act, the German Securities Trading Act and other special statutory provisions also apply. The management board bears responsibility for the use of AI systems and may be held liable for damage caused by defective AI systems. Liability extends both to the development and implementation of AI systems and to their ongoing operation and monitoring.

> “
>
> Financial undertakings must develop a comprehensive understanding of the AI systems they use, even where these are provided by third parties.

Outsourcing and Third-Party Risks

The use of AI in the financial sector is frequently outsourced to third parties, in particular to specialised technology providers and FinTech companies. This creates particular challenges for ICT third-party risk management, which is regulated in detail by DORA (Art. 64 DORA Regulation).

BaFin already published guidance on outsourcing to cloud providers in 2018, which is also relevant for AI applications. With the implementation of DORA, these requirements have been further tightened and systematised. Financial undertakings must now ensure that their ICT service providers, including providers of AI systems, comply with the requirements for digital operational resilience.

BaFin's guidance on ICT risks in the use of AI emphasises that financial undertakings must develop a comprehensive understanding of the AI systems they use, even where these are provided by third parties. This includes knowledge of how the AI systems work, the data used, the potential risks and the means of remediating errors.

Future Developments

The legal regulation of the use of AI in the financial sector remains in a dynamic process of development. In the coming years, the AI Regulation will be further specified by delegated acts and implementing acts of the Commission (Art. 6 AI Regulation). The Commission has already published guidelines on the definition of an AI system and on prohibited practices.

The European Commission also promotes the development of codes of practice at Union level in order to contribute to the proper application of the AI Regulation (Art. 56 AI Regulation). These codes of practice are intended to cover at least the obligations provided for in Articles 53 and 55, including the means of ensuring the currency of information, the appropriate level of detail in summaries of content used for training, and measures for assessing and managing systemic risks (Art. 56 AI Regulation).

At national level, the KI-MIG has restructured supervision of AI systems in Germany, entering into force on 29 July 2026. The Act follows a hybrid approach: the Federal Network Agency acts as the central coordination and competence centre, notifying authority and general market surveillance authority, while BaFin has been given extended supervisory competence for AI systems used in connection with regulated financial activities of supervised institutions, including credit servicers within the meaning of the Credit Servicing Directive Implementation Act (Kreditzweitmarktgesetz).

This sectoral allocation of supervisory responsibility to BaFin for AI in the financial sector avoids a fragmentation of supervision and demarcation difficulties vis-à-vis sector-specific supervisory law. Questions remain, however, as to the practical coordination between the Federal Network Agency and BaFin where a single AI system is used across multiple business areas, and as to the independence of AI market surveillance, since BaFin is subject to the legal and technical supervision of the Federal Ministry of Finance.

In the long term, the potential applications of AI in the financial sector could develop considerably further. Experts are already speculating about scenarios involving artificial general intelligence. One likely concept is that of so-called “self-driving finance”, in which an intelligent AI agent manages the finances of private or business customers by automating routine decisions or advising customers on more complex decisions.

Even more futuristic is the concept of a machine-to-machine (M2M) economy, in which smart, autonomous, networked and economically independent machines or devices act as participants that carry out the necessary activities of production, distribution and allocation with little or no human intervention. These scenarios raise entirely new legal questions that have not yet been conclusively resolved.

Conclusion

The legal developments relating to the use of AI in the banking and financial services sector in Germany and Europe are characterised by increasing regulatory density. The AI Regulation creates a comprehensive European framework, supplemented by DORA in the area of digital operational resilience. At national level, Germany has taken important steps with the FinMaDiG and the BaFin principles to implement and specify these European requirements.

The risk-based approach of the AI Regulation, which imposes particularly strict requirements on high-risk AI systems such as AI-supported creditworthiness assessments, reflects the fact that AI applications in the financial sector can have significant effects on the rights and interests of consumers and on the stability of the financial system. Following the deferral introduced by the Digital Omnibus on AI, financial undertakings now have until 2 December 2027 to achieve compliance with the Annex III high-risk obligations, rather than the originally envisaged 2 August 2026. At the same time, the legislator seeks not to unnecessarily hinder innovation in order to safeguard the competitiveness of the European financial sector.

For financial undertakings, this means that they must carefully plan and implement the use of AI systems. Compliance with regulatory requirements calls for a comprehensive understanding of AI technology, robust risk management and close cooperation with supervisory authorities. Only in this way can the opportunities offered by AI be used while the associated risks are controlled.

The coming years will show whether the chosen regulatory approach is suitable for appropriately balancing innovation and protection. The dynamic further development of AI technology and the increasing spread of AI applications in the financial sector will certainly require further adjustments and additions to the legal framework.

> “
>
> Financial undertakings now have until 2 December 2027 to achieve compliance with the Annex III high-risk obligations.

GHANA

THE LEGAL AND REGULATORY FRAMEWORK FOR ARTIFICIAL INTELLIGENCE IN FINANCIAL SERVICES IN GHANA: DEVELOPMENTS, CHALLENGES, AND EMERGING TRENDS

SUSTINERI ATTORNEYS

Harold Kwabena Fearon
Associate

harold@sustineriattorneys.com

+233576792728

www.sustineriattorneys.com

BIO

Harold is an Associate at SUSTINERI ATTORNEYS PRUC, Ghana's foremost technology, fintech and start-up focused law firm. Harold has a wealth of expertise in entrepreneurship and corporate management services, startups and SMEs, regulatory compliance, financial technology and innovation, corporate law and transactions, new media, intellectual property, entertainment, and sports.

Before joining SUSTINERI ATTORNEYS PRUC, he worked with other leading Ghanaian law firms, which helped him gain valuable experience, insights, and a strong foundation in corporate and company law and practice in Ghana.

He advises startups, SMEs, local and multinational companies on regulatory compliance issues relating to incorporation, licensing/ registration with industry regulators, business operations, products, and services, and exit strategies.

Harold is also an avid writer and a regular contributor to the Business and Financial Times (B&FT), Ghana's leading business newspaper. His passion and interest for startups, fintech and innovation has led him to undertake related courses to improve his knowledge and expertise in these new and emerging practice areas. Outside the practice, he serves as the Deputy Director - Legal of the Association of Ghana Startups ecosystem, the leading association representing startups in Ghana.

SUSTINERI
ATTORNEYS
CORPORATE · TRANSACTIONS · DISPUTES · TAX

THE LEGAL AND REGULATORY FRAMEWORK FOR ARTIFICIAL INTELLIGENCE IN FINANCIAL SERVICES IN GHANA: DEVELOPMENTS, CHALLENGES, AND EMERGING TRENDS

THE STATE OF ARTIFICIAL INTELLIGENCE IN GHANA'S FINANCIAL SERVICES SECTOR

Artificial Intelligence (AI) has transitioned from a futuristic concept to a defining force shaping the modern financial ecosystem. In Ghana, AI is gradually reshaping the delivery, governance, and regulation of financial services, from credit scoring and fraud detection to customer engagement, compliance, and risk management. While adoption remains nascent compared to global benchmarks, Ghana's financial sector has made notable strides in embedding AI-driven solutions within banking, insurance, and fintech operations.

> “
>
> AI regulation in Ghana remains at a formative stage.

The push toward AI adoption has been driven by two key realities. First, the country's financial system has become increasingly digital, following widespread use of mobile money and digital payment platforms. Second, the competitive pressures from fintech innovation have compelled traditional financial institutions to modernize their operations and decision-making processes. AI-powered automation, data analytics, and natural language processing tools now support several back-office and customer-facing functions in Ghanaian banks.

Regulated Financial Institutions such as ABSA Bank Ghana, Stanbic Bank, Ecobank, and several leading fintech startups now deploy AI tools for chatbots, predictive customer analysis, and anti-fraud mechanisms. The 2025 PwC Ghana Banking Survey highlights that more than half of surveyed financial institutions intend to integrate machine learning into credit and operational risk assessment models within the next three years. Similarly, the Bank of Ghana's (BoG) Fintech and Innovation Office has identified AI and digital automation as key pillars of the evolving regulatory technology (RegTech) landscape.

SUSTINERI ATTORNEYS
CORPORATE . TRANSACTIONS . DISPUTES . TAX

From a national standpoint, the adoption of AI in financial services aligns with the Ghana National Artificial Intelligence Strategy, launched in 2026 by the Ministry of Communications and Digitalization. The strategy positions AI as a central enabler of Ghana's digital transformation agenda, with financial services, health, and education identified as priority sectors for targeted interventions.

Yet, despite these promising developments, AI regulation in Ghana remains at a formative stage. The country's current legal framework does not have a single statute dedicated exclusively to AI governance. Instead, regulation is spread across multiple laws and policy instruments covering data protection, cybersecurity, digital finance, and consumer protection. However, a clear national commitment toward responsible AI development is now evident, both in public policy and in regulatory planning within the financial services sector.

REGULATORY LANDSCAPE AND PRINCIPAL ACTORS

1. The Bank of Ghana (BoG)

The BoG plays a leading role in the governance of AI-driven financial services. Through the Payment Systems and Services Act, 2019 (Act 987), the Bank regulates electronic money issuers, payment service providers, and fintech companies. The Act provides the foundation upon which AI-enabled financial products are currently deployed.

Additionally, the BoG's Regulatory Sandbox, launched in 2022 htas become a key vehicle for testing AI-based financial innovations. The Sandbox allows fintechs and banks to experiment with AI-driven solutions in a controlled environment under regulatory oversight. Several projects tested within this framework involve predictive analytics for credit risk, biometric verification for KYC, and automated compliance solutions (RegTech).

2. The Data Protection Commission (DPC)

The Data Protection Act, 2012 (Act 843) governs personal data processing in Ghana and forms a critical pillar of AI governance. Given that AI systems rely heavily on large-scale data collection and processing, the DPC plays a key role in enforcing compliance with data minimization, purpose limitation, and user consent principles.

Financial institutions leveraging AI must therefore ensure that personal and transactional data used in training algorithms comply with the provisions of Act 843. Cross-border data transfers, a common feature of cloud-based AI services, are subject to prior approval by the DPC and restricted to countries offering adequate data protection.

3. The Cyber Security Authority (CSA)

The Cybersecurity Act, 2020 (Act 1038) complements AI regulation by safeguarding digital financial infrastructure. It mandates critical information infrastructure operators, including banks and payment institutions, to implement risk-management systems capable of detecting and responding to cyber threats - many of which now involve AI-driven attacks.

Given the growing use of AI for cybercrime detection and prevention, the CSA works closely with financial institutions to align cybersecurity standards with AI-driven automation. AI-enabled threat intelligence tools, anomaly detection systems, and fraud monitoring platforms are now common in Ghana's regulated financial sector, reflecting this integration.

4. The Ministry of Communications and Digitalisation (MoCD)

The MoCD is the policy lead for Ghana's National AI Strategy, which sets the broader direction for AI governance and ethical deployment. The strategy's financial services component focuses on four priorities:

- Enhancing the regulatory environment for algorithmic systems;
- Developing data infrastructure and digital identity frameworks;
- Building AI skills and research capacity; and
- Encouraging responsible innovation through public-private partnerships.

The Ministry has also announced plans to establish a National AI Centre of Excellence in partnership with academia and industry to promote standards and coordinate sectoral implementation.

5. Supporting Agencies

Other regulatory and policy actors include:

a. The **National Information Technology Agency (NITA)** – responsible for national ICT standards and interoperability, particularly for AI-driven platforms.
b. The **Financial Intelligence Centre (FIC)** – integrating AI tools into anti-money laundering (AML) monitoring and suspicious transaction reporting.
c. The **Ghana Revenue Authority (GRA)** – exploring AI systems for tax intelligence and compliance within the digital economy.

These institutions collectively form Ghana's emerging AI governance architecture in financial services - one that relies on collaboration, sectoral oversight, and adaptive policy evolution.

“

These institutions collectively form Ghana's emerging AI governance architecture in financial services.

LEGAL FOUNDATIONS AND RELEVANT LEGISLATION

Although Ghana does not yet have an AI-specific statute, existing legal frameworks already influence the development and deployment of AI in financial services. Key instruments include:

1. **Payment Systems and Services Act, 2019 (Act 987)** – Establishes regulatory requirements for digital financial service providers, under which AI-driven payment and lending models must operate.
2. **Data Protection Act, 2012 (Act 843)** – Provides privacy safeguards governing the collection, storage, and use of personal data within AI systems.
3. **Cybersecurity Act, 2020 (Act 1038)** – Imposes obligations on critical information infrastructure operators, ensuring that AI tools used for cybersecurity or fraud prevention comply with national resilience standards.

4. **Anti-Money Laundering Act, 2020 (Act 1044)** – Governs the application of AI in transaction monitoring, customer due diligence, and suspicious transaction detection.
5. **Electronic Transactions Act, 2008 (Act 772)** – Provides legal recognition for electronic communications, signatures, and contracts which is important for AI-driven financial automation.
6. **Companies Act, 2019 (Act 992)** – Imposes corporate governance duties that extend to risk oversight in AI adoption, including board-level accountability for ethical and compliance issues.

> “
>
> For the financial sector, the NAIS acts as both a policy compass and regulatory roadmap.

Together, these instruments provide a partial but evolving framework for AI regulation. The forthcoming Emerging Technologies Bill, expected by 2026 under the National AI Strategy, will likely consolidate these obligations and introduce risk-based oversight mechanisms similar to the EU's AI Act.

THE NATIONAL AI STRATEGY AND ITS IMPLICATIONS FOR FINANCIAL SERVICES REGULATION

The National Artificial Intelligence Strategy (NAIS) launched by Ghana's Ministry of Communications and Digitalization (MoCD) in April 2026 marks a historic policy milestone. It establishes Ghana as one of the first sub-Saharan African nations to articulate a comprehensive national framework for AI governance and adoption.

The Strategy's core vision is to leverage AI for *inclusive growth, productivity enhancement, and digital transformation* across all sectors of the economy. Within the financial services space, the NAIS envisions AI as a key enabler of innovation, customer inclusion, and operational efficiency, while emphasizing accountability, fairness, and human oversight.

The strategy identifies five areas relevant to financial regulation:

1. **Policy, Regulation and Governance** – Establishing a legal framework for AI ethics, risk classification, and accountability.
2. **Data and Digital Infrastructure** – Building interoperable and secure data systems for AI training and analytics.
3. **Research, Innovation, and Skills Development** – Strengthening capacity for AI-driven fintech solutions and compliance technologies.
4. **Industry Adoption and Partnerships** – Encouraging collaboration between regulators, financial institutions, and technology providers.
5. **International Cooperation** – Aligning Ghana's AI governance with continental and global norms, including the African Union's AI Continental Strategy and OECD guidelines.

For the financial sector, the NAIS acts as both a policy compass and regulatory roadmap. It mandates the development of an *Emerging Technologies Bill* to codify principles around risk management, data ethics, and algorithmic accountability. The Bill currently under consultation, will likely define AI system risk tiers, institutional liability, and standards for algorithmic explainability in financial decision-making.

ETHICAL AI, CONSUMER PROTECTION AND ALGORITHMIC TRANSPARENCY

The integration of AI into financial services presents unique challenges around ethics, fairness, and consumer protection. Ghana's regulatory response has been to extend traditional consumer protection principles: transparency, accountability, and fairness to algorithmic decision-making environments.

Under the Data Protection Act, consumers must be informed whenever their eligibility, credit limit, or transaction monitoring is determined by automated systems, and to object to the use of same. They must also have a right to human intervention if adversely affected by an AI-based decision.

The Data Protection Commission (DPC) has issued interpretative guidance encouraging financial institutions to ensure "algorithmic explainability"- that is, providing understandable reasons for decisions taken by machine learning systems. This complements global AI ethics principles such as the OECD's call for human-centric AI and the African Union's emphasis on fairness and inclusion.

Financial institutions are also expected to maintain internal AI ethics committees or designate senior officers responsible for AI governance. These bodies oversee compliance with ethical guidelines, assess bias in models, and review product designs to ensure consumer welfare.

Such approaches, while still evolving, illustrate Ghana's broader move from a *risk-tolerant* to a *risk-aware* model of innovation: one that encourages digital transformation while protecting the public interest.

REGULATORY SANDBOXES, INNOVATION HUBS, AND REGTECH INTEGRATION

The Bank of Ghana's Regulatory Sandbox, introduced in 2022, has proven instrumental in shaping Ghana's AI innovation ecosystem. Through the Sandbox, fintechs, banks, and startups can test new technologies including AI-driven products under relaxed regulatory conditions and direct supervisory oversight.

AI applications tested in the Sandbox have focused on areas such as:

a. Automated compliance and AML monitoring using machine learning;
b. Predictive analytics for SME credit scoring;
c. AI-powered customer service chatbots;
d. Biometric identity verification; and
e. Dynamic pricing and risk-based lending models.

Participants benefit from BoG's regulatory guidance, while the Bank gains insight into emerging technologies and their risk implications. This symbiotic model supports policy learning and ensures that regulatory frameworks evolve alongside innovation.

In parallel, the BoG has begun integrating RegTech and SupTech tools - AI-enabled regulatory and supervisory technologies into its operations. These tools enable real-time transaction monitoring, anomaly detection, and risk-based supervision.

Furthermore, Ghana's National Financial Inclusion and Development Strategy (NFIDS) encourages regulators to adopt AI tools for data analysis, promoting evidence-based policy and predictive supervision. Together, these developments signal the rise of AI-assisted regulation, where regulators and market actors alike leverage AI for efficiency and compliance.

> "Ghana's broader move from a risk-tolerant to a risk-aware model of innovation.

CHALLENGES AND POLICY GAPS

Despite notable progress, Ghana's regulatory environment for AI in financial services faces several challenges:

1. **Absence of Dedicated AI Legislation** – Current laws only indirectly address AI-related issues, leaving gaps in liability, accountability, and risk classification.
2. **Data Quality and Infrastructure Limitations** – Many AI models rely on fragmented or low-quality datasets, raising concerns about bias and accuracy.
3. **Skills and Capacity Deficits** – There is a shortage of AI professionals with expertise in both machine learning and financial regulation.
4. **Cross-Border Data Governance** – With most AI systems hosted on global cloud platforms, jurisdictional issues regarding data protection and compliance remain unresolved.
5. **Ethical and Social Risks** – The potential for algorithmic discrimination and exclusion of vulnerable groups continues to challenge policymakers.

Addressing these gaps will require a combination of legal reform, institutional capacity-building, and cross-sector collaboration.

EMERGING TRENDS AND THE FUTURE OF AI GOVERNANCE IN GHANA'S FINANCIAL SECTOR

Looking ahead, several trends are likely to define Ghana's AI regulatory and market landscape:

1. **Legislative Codification of AI Oversight** – The forthcoming Emerging Technologies Bill (2026) is expected to institutionalize ethical, risk-based regulation and assign formal oversight powers to a proposed Responsible AI Office.
2. **Integration of AI into Credit Reference and Risk Management Systems** – The BoG's Credit Reporting System will increasingly rely on AI models for credit scoring and fraud detection.
3. **AI-Driven Cybersecurity** – Financial institutions will deepen their adoption of AI-enabled cyber-defense systems, supported by the Cyber Security Authority.
4. **Sustainability and ESG Analytics** – AI will play a growing role in ESG compliance and sustainable finance reporting.
5. **Continental Harmonization** – Ghana's AI frameworks will likely align with African Union standards and ECOWAS digital policy, promoting cross-border interoperability.

In the longer term, Ghana's ambition is to position itself as a regional AI governance leader, combining strong regulatory frameworks with a thriving innovation ecosystem.

CONCLUSION

Ghana's approach to artificial intelligence in financial services demonstrates a deliberate shift from cautious observation to structured governance. The combination of the National AI Strategy (2026), and existing data protection and cybersecurity laws marks the foundation of an emerging, adaptive regulatory regime.

While Ghana's framework remains in development, its trajectory is clear: promoting innovation within a disciplined, ethically grounded, and risk-based governance structure. If effectively implemented, Ghana could become a model for responsible AI integration in financial services across Africa, balancing technological progress with trust, transparency, and inclusion.

HUNGARY

AI IN HUNGARIAN FINANCE: TURNING INNOVATION INTO ACCOUNTABLE SCALE

CMS

Katalin Horváth
Partner

katalin.horvath@cms-cmno.com

+36 1 483 4897

www.cms.law

BIO

Katalin Horváth is a partner in the commercial team at CMS Budapest and Co-Head of CMS CEE Fintech Competence Center. She specialises in software, IT and IP law, BankTech/FinTech law, outsourcing, data protection, and cybersecurity (NIS2, CRA, CSA) matters, as well as the legal and regulatory aspects of artificial intelligence.

She has in-depth experience in complex TMT regulatory matters, commercial contracts, e-commerce and platform regulation, and e-signature issues with a particular focus on online and mobile services, web stores, FinTech, PayTech and payment solutions.

Katalin also has significant hands-on experience in AI and emerging technologies, including blockchain and cryptoassets, IoT, robotics and drones, as well as sector-specific work in space and defence. She supports the CMS Dispute Resolution team as an expert in technology-related disputes.

She regularly publishes and is a frequent conference speaker on BankTech/FinTech, AI, software law and data protection.

Before joining CMS, Katalin gained significant experience in drafting IT operating and software-related contracts, copyright and trademark litigation at a renowned Hungarian IP boutique law firm.

CMS
law·tax·future

AI is moving from isolated pilots into the operating fabric of banks, insurers and investment firms. In Hungary, the commercial opportunity is substantial, but so is the regulatory density: the EU AI Act now sits alongside DORA, the GDPR, sector-specific financial-services rules and an increasingly detailed supervisory framework issued by the Magyar Nemzeti Bank (MNB). For legal and compliance teams, the central task is no longer simply to identify whether an institution "uses AI". It is to understand which systems are used, for what purpose, in which legal role and under which combination of regimes.

> AI is moving from isolated pilots into the operating fabric of banks, insurers and investment firms.

I. From Pilots to Core Infrastructure

Financial institutions already use AI across a wide range of functions, including fraud and market-abuse detection, anti-money laundering, credit assessment, pricing, portfolio management, claims handling, cybersecurity, regulatory compliance and back-office automation. Customer-facing applications, from virtual assistants to personalised recommendations, are also becoming more sophisticated. The attraction is clear: institutions hold large volumes of data, operate repeatable processes and face constant pressure to improve speed, accuracy and cost efficiency.

The next phase is likely to be less about stand-alone tools and more about embedding AI into end-to-end processes. Banks are experimenting with internal AI academies, specialist governance teams and, increasingly, AI agents capable of coordinating tasks across several applications. This evolution can create material efficiencies, but it also moves AI closer to decisions and actions that affect customers, employees and the institution's own operational resilience.

Technology architecture remains a practical constraint. Many established institutions still depend on monolithic legacy systems, fragmented data estates and manual interfaces. Scalable AI therefore tends to accelerate wider modernisation: cloud adoption, containerisation, microservices and more disciplined data management. The decisive asset is not simply access to more data, but access to data that is accurate, traceable, appropriately governed and available at the right point in the process. In financial services, weak data architecture quickly becomes a legal, prudential and conduct risk.

II. The AI Act: The Rules That Matter Most

Regulation (EU) 2024/1689 (the AI Act) entered into force on 1 August 2024 and introduced a horizontal, risk-based framework for AI systems and general-purpose AI models. Its application is phased. Prohibited practices and the AI-literacy provisions have applied since 2 February 2025; the governance and general-purpose AI model rules have applied since 2 August 2025; and the Commission and national authorities began enforcing the generally applicable provisions, including key transparency duties, on 2 August 2026. Following the 2026 AI Omnibus, the high-risk rules for systems listed in Annex III apply from 2 December 2027, while those for AI embedded in regulated products listed in Annex I apply from 2 August 2028.

For financial institutions, the prohibited-practices regime is already operational. Relevant examples include AI that uses materially harmful manipulation or deception, social scoring that leads to unjustified detrimental treatment, emotion recognition in the workplace except within narrow medical or safety exceptions, and biometric categorisation intended to infer protected characteristics. The analysis must remain use-case specific. Conventional credit scoring based on relevant financial information is not prohibited merely because it involves profiling; by contrast, repurposing unrelated behavioural or sensitive data may trigger serious concerns under the AI Act, the GDPR and sectoral conduct rules.

The Act also imposes transparency obligations on certain interactive and generative systems. A customer should, subject to the statutory exceptions, be informed when interacting directly with an AI system. This is particularly relevant to chatbots, voice assistants and generative-AI interfaces used in customer service. Separately, Article 4 requires providers and deployers to take measures supporting AI literacy among staff and other persons operating AI systems on their behalf. It does not prescribe a uniform qualification or require institutions to guarantee a particular level for every individual; the measures should reflect the person's knowledge, role and the context in which the system is used.

The most demanding obligations concern high-risk systems. Financial-sector examples include AI used to evaluate the creditworthiness of natural persons or establish their credit score, except where used solely to detect financial fraud; AI used for risk assessment and pricing in life and health insurance; specified recruitment, worker-management and performance-monitoring tools; and emotion-recognition systems where their use is not prohibited. Classification depends on intended purpose and deployment context, not on the technology label chosen by the supplier.

A financial institution will commonly be a deployer, but its role must be mapped entity by entity. An institution may become a provider where it develops a system and places it on the market or puts it into service under its own name, rebrands it, substantially modifies it or changes its intended purpose in a way that brings it into the high-risk category. Group structures require particular care: the fact that a tool is developed by an internal technology centre does not, by itself, resolve which legal entity bears the provider obligations.

Deployers of high-risk systems must follow the instructions for use, assign effective human oversight, monitor operation, retain relevant logs and ensure that input data under their control is relevant and sufficiently representative. They must respond to risks and serious incidents and comply with information duties towards affected persons. For creditworthiness assessment and life or health insurance risk assessment and pricing, a fundamental rights impact assessment must be completed before first use once the high-risk provisions apply; where a data protection impact assessment is also required, the two should be coordinated.

Provider obligations are broader. They include lifecycle risk and quality management, data governance, technical documentation and traceability, logging, clear instructions, human-oversight design, accuracy, robustness and cybersecurity, conformity assessment, registration, post-market monitoring, corrective action and serious-incident reporting. For institutions building or materially adapting high-risk tools, these are product-compliance obligations, not merely internal policy requirements. The most efficient preparation is therefore a role-and-use-case inventory that links each system to its intended purpose, risk classification, responsible entity, applicable controls and evidence of compliance.

III. Hungary's Supervisory Overlay: Strategy, Data and Control

The MNB's Recommendation No. 13/2025 (XII. 3.) on the digital transformation of credit institutions adds a distinctly Hungarian supervisory layer. The Recommendation is not legally binding, but it expresses the MNB's supervisory expectations, and the MNB monitors and assesses compliance in its audit and monitoring activity. Credit institutions were expected to apply it from 1 April 2026, with the first AI sub-strategy due to the MNB by 30 June 2026. It superseded the earlier digital-transformation recommendation from 2021.

> “
>
> In practice, legal, compliance, risk, data, IT security, procurement and business functions need a common operating model rather than parallel policies.

Its importance lies in breadth. Whereas the AI Act reserves its most prescriptive system-level controls for defined categories, the MNB expects a coherent institutional approach to the entire AI portfolio. The AI sub-strategy should sit within the broader digital-transformation strategy and translate ambition into governance, measurable milestones and accountable ownership. In practice, legal, compliance, risk, data, IT security, procurement and business functions need a common operating model rather than parallel policies.

The Recommendation's expectations can be grouped around several themes. First, institutions should classify their AI systems, define acceptable risk levels and operate documented lifecycle risk management, with classifications reviewed regularly and at least annually. Secondly, data governance should cover the provenance, quality, preparation, access, storage and deletion of training, validation, test and production data. In several areas, the MNB treats the extension of high-risk-style controls to systems outside the AI Act's high-risk category as good practice.

> A single deployment may be subject simultaneously to the AI Act, DORA, the GDPR and sector-specific prudential and conduct requirements.

Thirdly, development and procurement controls should address validation, reproducibility, bias and ethical testing, guardrails, human intervention, ongoing performance monitoring and independent assurance where appropriate. For third-party systems, the institution should obtain enough information to assess technical, legal and ethical compliance rather than treating the supplier's standard documentation as conclusive. Fourthly, AI-specific security should protect training data, models, non-public inputs and prompts, and should include measures to detect manipulation and adversarial attacks. Finally, competence and communication matter: institutions should set clear rules for employees' use of third-party and generative AI, provide role-specific training, test key competencies where proportionate, maintain internal error-reporting channels and communicate transparently with customers.

Procurement is where these expectations meet DORA and established outsourcing law. AI delivered through a cloud platform, AI-as-a-Service arrangement or managed solution must be classified separately under the applicable outsourcing rules and DORA's ICT third-party risk framework. The relevant MNB instruments include Recommendation No. 7/2020 on external service providers and Recommendation No. 2/2025 on community and public cloud services, which replaced the former 2019 cloud recommendation.

Whether an AI service supports a critical or important function cannot be determined by a generic product label. The assessment turns on the consequences of disruption for the particular institution, including its regulatory compliance, financial performance and continuity of services. Where the threshold is met, enhanced due diligence, contractual safeguards, monitoring, business-continuity arrangements and a credible exit strategy are required. AI contracts should therefore address, at a minimum, permitted data use, data location and security, audit and access rights, incident notification, subcontracting, model or service changes, performance and validation, continuity, termination and the return or deletion of data. Vendor lock-in is especially acute where the institution's workflows, prompts, data formats or fine-tuning become model-specific.

IV. One Deployment, Several Rulebooks

The AI Act does not displace financial-services regulation. A single deployment may be subject simultaneously to the AI Act, DORA, the GDPR and sector-specific prudential and conduct requirements. The legal analysis should therefore begin with the activity and the risk, not with a search for one "lead" regulation.

DORA, applicable since 17 January 2025, is the primary EU framework for the digital operational resilience of covered financial entities. For the ICT-risk and incident-reporting matters it regulates, it operates as a sector-specific Union legal act in relation to NIS2. An externally supplied AI system may be an ICT service under DORA, and the same arrangement may also involve a high-risk AI system under the AI Act. The two regimes are cumulative: the AI Act focuses on the system, its intended purpose and the obligations of actors in the AI value chain; DORA focuses on the financial entity's governance of ICT risk, resilience testing, incidents and third-party dependencies. The designation of a supported function as critical or important intensifies the DORA requirements, but the assessment remains institution-specific.

Data protection remains equally central. Training and production data, prompts, outputs and monitoring logs may contain personal data, while data matching across internal and external sources can alter the purpose and risk of processing. Institutions must identify a lawful basis, apply purpose limitation and data minimisation, assess transparency and retention, and carry out a data protection impact assessment where required. Where AI is used to take decisions producing legal or similarly significant effects, Article 22 GDPR and its safeguards, including meaningful human intervention where applicable, require close attention. A nominal human sign-off will not cure an otherwise automated process if the reviewer lacks authority, time or information to challenge the result.

Sectoral expectations add further texture. ESMA's public statement on AI in investment services emphasises that existing MiFID II organisational and conduct duties continue to apply, particularly the obligation to act in clients' best interests and to manage risks such as bias, opacity, overreliance, privacy and security. EIOPA's 2025 Opinion clarifies governance and risk-management expectations for insurance AI that is neither prohibited nor classified as high-risk, with emphasis on data governance, fairness, record-keeping, cybersecurity, explainability and human oversight. These instruments reinforce a practical point: systems outside the AI Act's high-risk category are not unregulated.

The Cyber Resilience Act (CRA) adds another product-security layer. Its main obligations apply from 11 December 2027, while specified reporting duties apply from 11 September 2026. Direct CRA obligations are most relevant where a financial institution acts as a manufacturer or other economic operator placing a product with digital elements on the market, rather than merely procuring one. Where both regimes apply, CRA compliance may satisfy the cybersecurity limb of the AI Act's high-risk requirements under the conditions set by the legislation. For legal teams, the wider lesson is to create one integrated control map and evidence set, then identify which legal obligations each control satisfies.

V. Agentic AI: Autonomy with Guardrails

Agentic AI is likely to be the next major test of that integrated model. These systems move beyond generating content: they can plan a sequence of steps, retrieve real-time information, interact with databases and APIs, retain context and execute actions. Potential financial-services uses include dynamic insurance pricing, fraud investigation, credit workflows, automated KYC and AML processes, payment initiation and the coordination of complex customer-service cases.

The legal risk rises when a system can act, not merely advise. Hallucinated instructions, excessive functionality, over-broad permissions, data leakage, unintended transactions and opaque coordination between multiple agents can create a much larger "blast radius" than a conventional chatbot. The number of connected tools and data sources also expands the attack surface and makes responsibility harder to allocate. In customer or employee contexts, over-profiling and the stitching together of sensitive data may undermine data minimisation, fairness and the effectiveness of human intervention.

“

The legal risk rises when a system can act, not merely advise.

Agentic AI is not a separate legal category under the AI Act. Its classification still turns on intended purpose, functionality, the institution's role and the context of use. Its autonomy does, however, change the control design. Institutions should begin with tightly bounded use cases and apply least-privilege access, segregated environments, transaction limits, human approval for material decisions or payments, circuit breakers, tamper-evident logs, continuous monitoring and adversarial testing. Multi-agent deployments also need a defined orchestrator, clear escalation rules and controls preventing agents from repeatedly delegating or acting outside their mandate.

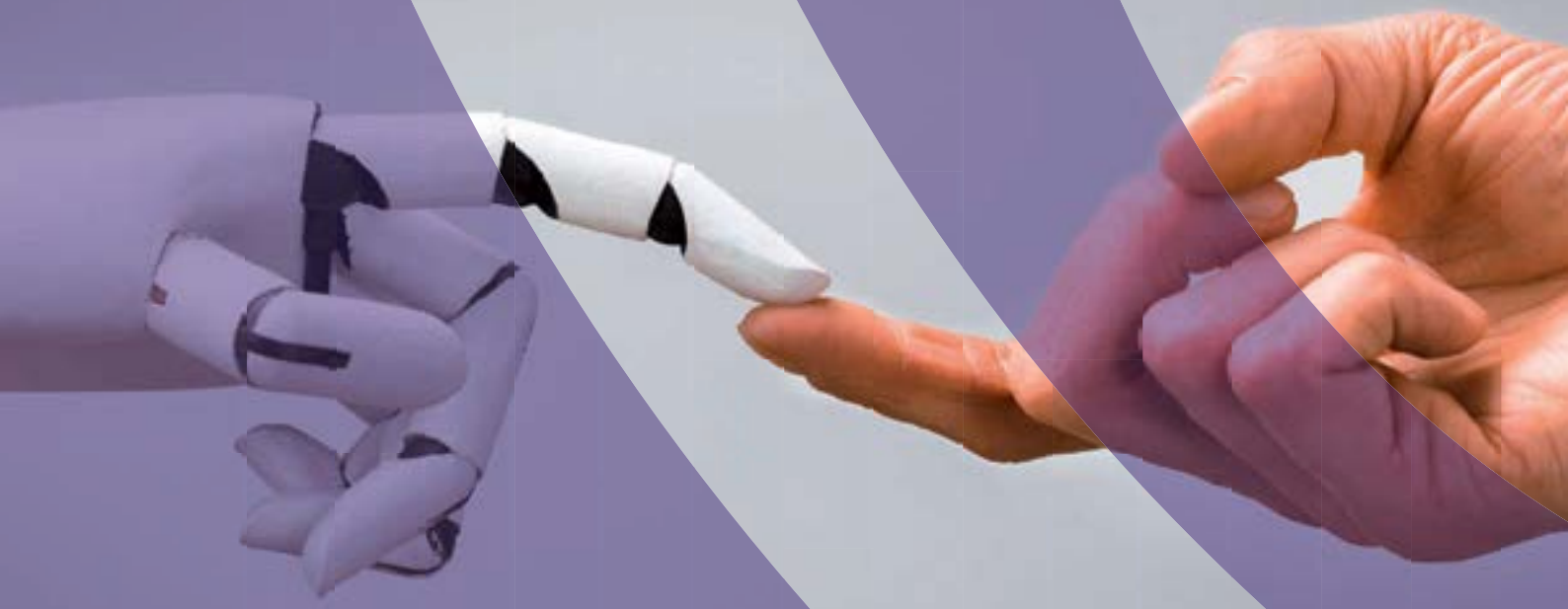
> “
>
> In finance, trustworthy AI governance is not simply a compliance cost; it is the infrastructure for sustainable adoption.

Contracts with external providers must reflect this increased agency. Responsibility should be allocated for tool connections, permissions, model changes, data use, security testing, audit, incident response, service continuity and exit. The institution should also be able to suspend the system quickly and reconstruct why an action occurred. The opportunity is considerable, but in regulated finance autonomy is sustainable only when it is paired with demonstrable control. Institutions that build that discipline early will not merely comply more effectively; they will be better placed to scale AI with confidence.

VI. The Governance Advantage

For Hungarian financial institutions, the strategic question is not whether AI can be deployed, but whether it can be scaled without fragmenting accountability. The strongest programmes will treat regulation as an architecture problem: one inventory, clear actor mapping, risk-tiered controls, reusable evidence and contracts aligned with operational reality. That approach reduces duplication across the AI Act, DORA, the GDPR and sectoral rules while giving boards a clearer view of residual risk. It also creates room for innovation. Institutions that can demonstrate where data comes from, how models are tested, who can intervene and how a service can be stopped or exited will be able to move faster than those relying on ad hoc approvals. In finance, trustworthy AI governance is not simply a compliance cost; it is the infrastructure for sustainable adoption.

CMS
law·tax·future



Tech-driven – but always human.

Future facing with CMS

CMS is an international law firm that helps clients to thrive through technical rigour, strategic expertise and a deep focus on partnerships.

cms.law

JAPAN

JAPAN'S TECHNOLOGY-NEUTRAL APPROACH TO CUSTOMER-FACING GENERATIVE AI IN FINANCE

AMT

Takashi Nakazaki
Partner

takashi.nakazaki_grp@amt-law.com

+81-3-6775-1086

www.amt-law.com

AMT / ANDERSON MORI & TOMOTSUNE

BIO

Takashi Nakazaki is a Partner at Anderson Mori & Tomotsune. His practice focuses on data protection and privacy, AI, information security, intellectual property, software and licensing, and payment services, together with related technology, e-commerce and corporate matters.

He is recognised in leading legal rankings and has published Japanese-language books on the GDPR (2018), the metaverse (2023) and generative AI (2024). He has served on expert committees for METI, MIC, the Cabinet Office and other bodies, including the working group that drafted the AI Guidelines for Business, and has led research commissioned by the Personal Information Protection Commission. He has been an auditor of the Japan Institute for Health Security since April 2025.

Mr Nakazaki co-founded the IAPP Tokyo KnowledgeNet Chapter, serves on the IAPP Asia Advisory Board, and teaches AI law at Keio University Law School. He is Associations and Committees Liaison Officer of the IBA Technology Law Committee and a member of AIPPI's Standing Committee on AI, IT and Internet.

He received his LL.B. from the University of Tokyo and his LL.M. from Columbia Law School, and previously worked at Arnold & Porter LLP.

AMT

Tomoaki Katayama
Partner

tomoaki.katayama@amt-law.com

+81-3-6775-1590

www.amt-law.com

BIO

Tomoaki Katayama is a partner at Anderson Mori & Tomotsune. Prior to joining the firm in February 2020, he worked at Japan Post Bank Co., Ltd. (April 2013–August 2013), Taisei Corporation (September 2013–April 2017), Sato Sogo Law Office (May 2017–October 2017), and So & Sato Law Offices (November 2017–January 2020). Drawing on the broad perspective he has developed through these diverse professional experiences, he advises on FinTech and financial regulation, with deep and extensive expertise in digital assets, including crypto-assets, stablecoins, and NFTs, as well as related cutting-edge technologies such as blockchain technology and its applications, including DeFi. He also handles matters relating to personal data protection, e-commerce, financial transactions, and gaming law.

Publications

6500 Legal Countermeasures for Financial Institutions (February 2026)
Global Legal Insights - Fintech 2025 Japan Chapter (September 2025)
Chambers Global Practice Guides' on Blockchain 2025 - Law & Practice (June 2025)

AMT / ANDERSON MORI & TOMOTSUNE

AMT

Takeshi Nagase
Partner

takeshi.nagase@amt-law.com

+81-3-6775-1200

www.amt-law.com

BIO

Takeshi Nagase is a partner at Anderson Mori & Tomotsune. Between 2013 and 2014, Takeshi served on secondment in the disclosure department at the Financial Services Agency of Japan (FSA). Additionally, he handled a broad range of finance and corporate transactions on a secondment stint with the legal department of a major Japanese securities firm from 2015 to 2017. As a result of the unique perspective he has gained from these professional experiences, Takeshi has extended his focus to crypto-asset laws, including regulatory requirements applicable to the registration of crypto-asset exchange service providers, initial coin offerings, stablecoin businesses, and NFT businesses.

AMT / ANDERSON MORI & TOMOTSUNE

AMT

Daisuke Iguchi
Special Councel

daisuke.iguchi@amt-law.com

+81-3-6775-1885

www.amt-law.com

BIO

Daisuke Iguchi is a Special Counsel at Anderson Mori & Tomotsune. He has gained experience through a secondment to a bank establishment preparatory company, where he was engaged in various legal matters related to the establishment of a bank, and as a fixed-term government employee at the Financial Services Agency, where he handled various legal inquiries under the Payment Services Act and the Money Lending Business Act.

Drawing on this experience, he specializes in cross-sectional analysis of various financial regulations, including the Banking Act, the Payment Services Act, the Money Lending Business Act, the Installment Sales Act, and the Financial Instruments and Exchange Act, as well as support in dealing with the authorities.

AMT / ANDERSON MORI & TOMOTSUNE

JAPAN'S TECHNOLOGY-NEUTRAL APPROACH TO CUSTOMER-FACING GENERATIVE AI IN FINANCE

AMT / ANDERSON MORI & TOMOTSUNE

1. Introduction: The Shift Towards Customer-Facing Generative AI

In Japan's financial sector, generative AI ("**GenAI**") is beginning to move beyond internal efficiency tools towards customer-facing services. When the Financial Services Agency of Japan (the "**FSA**") surveyed 130 financial institutions between October and November 2024, direct use of GenAI in customer-facing services was still rare. By contrast, the AI Discussion Paper (Version 1.1), "Preliminary Discussion Points for Promoting the Sound Utilization of AI in the Financial Sector" (the "**AIDP**"), published on 3 March 2026, records that institutions have since begun launching, or concretely considering, customer-facing GenAI services within limited scopes and conditions.

> "
>
> GenAI is therefore evolving from an employee-support tool into a technology capable of forming part of the process through which financial services are delivered.

"AI agents" have also become a realistic object of consideration. Unlike conventional chatbots, these systems may access multiple systems and data sources and autonomously perform defined tasks. GenAI is therefore evolving from an employee-support tool into a technology capable of forming part of the process through which financial services are delivered.[1]

Japan's regulatory response is distinctive. Rather than establishing a comprehensive financial-sector regime specifically for GenAI, Japan applies existing financial regulation in a technology-neutral manner, while addressing uncertainty through risk-based governance and public-private dialogue. The AIDP also highlights the "risk of inaction": the possibility that overly cautious institutions may fall behind technological innovation and become less able to provide high-quality financial services over the medium to long term. This article examines that approach, focusing on customer-facing GenAI, securities regulation, personal data protection and key risk-management issues.

2. Japan's AI Governance: A Promotion-Oriented Framework Act and Soft Law

Japan's AI policy seeks to reconcile innovation with risk management. The interim report of the government's AI Institution Study Group similarly identifies "balancing innovation promotion and risk response" as a central policy objective, favouring a combination of legislation and soft law while respecting business autonomy.[2]

This approach is reflected in the Act on the Promotion of Research and Development and Utilization of Artificial Intelligence-Related Technologies (Act No. 53 of 2025; the "**Japan AI Act**"), promulgated on 4 June 2025. Article 3(2) provides that, given AI's potential to improve efficiency, enhance business activities and create new industries, its research, development and utilization should be promoted to maintain domestic capabilities and strengthen international competitiveness. Article 3(4), however, recognizes that improper or inappropriate use may facilitate crime, leakage of personal information, copyright infringement and other harms, and calls for measures to ensure proper implementation.

The AI Act therefore differs in character from the EU AI Act. It does not establish a comprehensive risk-classification regime with detailed obligations and sanctions. Instead, it provides a framework for promoting AI-related measures comprehensively and systematically (Article 1), requires an AI Basic Plan (Article 18), and establishes the AI Strategy Headquarters within the Cabinet (Articles 19 and 20). The first plan was adopted by Cabinet decision on 23 December 2025, and the Phase II Artificial Intelligence Basic Plan was adopted by Cabinet decision on 14 July 2026. At the same time, it is not merely a promotion statute: Article 13 contemplates guidelines to ensure proper implementation, while Article 16 provides for analysis of cases involving improper AI use and measures including guidance and advice. It is therefore best characterized as a promotion-oriented and risk-responsive framework act.

1. See Ministry of Internal Affairs and Communications and Ministry of Economy, Trade and Industry, AI Guidelines for Business, Version 1.2 (31 March 2026), defining an AI agent as an AI system that senses its environment and acts autonomously to achieve a specified goal (author's translation).
2. Cabinet Office, AI Strategy Council / AI Institution Study Group: Interim Report (February 2025) (in Japanese; translations of quoted language are the author's).

Soft law remains important. The Social Principles of Human-Centric AI set out overarching societal principles. Separately, the AI Guidelines for Business provide practical guidance for actors engaged in AI development, provision and use; Version 1.2 was published on 31 March 2026.[3] Japan's overall model accordingly combines a cross-sector framework act, soft law and existing sector-specific regulation.

This structure is particularly clear in finance. The Banking Act, the Financial Instruments and Exchange Act ("**FIEA**"), the Insurance Business Act, the Act on the Protection of Personal Information ("**APPI**") and other existing laws apply regardless of whether AI is used. The FSA does not exclude legislation where a material regulatory gap emerges, but has indicated that interpretation or revision of existing principles and guidelines should generally be considered first.

Japan's Copyright Act is also relevant to AI development. Article 30-4 permits certain uses of works for information analysis and other purposes that do not involve enjoying the thoughts or sentiments expressed in the work, and AI training may fall within its scope. The exception is not unlimited, particularly where a purpose of enjoyment is also present or the use would unreasonably prejudice the interests of the copyright owner.

3. The Use of Generative AI by Japanese Financial Institutions

In the survey underlying the AIDP, more than 90% of responding institutions used conventional AI or GenAI in some form. GenAI use had expanded rapidly in internal operations, including document summarization, translation, proofreading, internal FAQs and system development, and more than 70% of institutions using AI in some form permitted broad GenAI use by general staff.

The AIDP broadly distinguishes three categories. The first is internal use, such as drafting, translation and information retrieval. The second is use supporting customer services or customer-affecting operations through a human intermediary, such as proposed responses for call-centre staff or drafts used in decision-making. The third is direct customer use, where GenAI output itself is presented to customers.

At the time of the 2024 survey, the third category was extremely limited. Version 1.1, however, records that direct customer use has begun in limited areas such as enquiry handling and administrative procedures. More recent Bank of Japan data reinforce this picture. In its FY2022026 survey of 150 financial institutions, 79.3% reported that they were already using GenAI and a further 11.3% were trailing it. Use had expanded from general administrative tasks into core operations involving customer information, although only a small number of institutions were directly presenting GenAI-generated output to customers.[4]

The risks change accordingly. Internal use primarily raises issues of confidentiality, erroneous output, copyright and employee misuse. Direct customer use may additionally trigger financial-regulatory concerns, including solicitation, suitability, customer disclosure and the effects of misinformation. If AI agents begin autonomously executing transactions or operational processes, questions of authority, accountability and system risk will become more significant.

"

Japan's overall model accordingly combines a cross-sector framework act, soft law and existing sector-specific regulation.

4. The FSA's Approach: Risk-Based Lifecycle Management

The AIDP is not a supervisory guideline and does not itself impose new legal obligations. It sets out preliminary issues for dialogue with the industry and the FSA's policy direction. Its basic approach rests on technology neutrality and risk-based regulation.

For customer-facing GenAI, the AIDP is especially notable for organizing risk-mitigation practices into a lifecycle framework. First, design and pre-deployment verification. Institutions may combine retrieval-augmented generation ("**RAG**"), system prompts, model selection, fine-tuning and output filters as layered guardrails. Deployment may begin with relatively low-risk services and expand as experience accumulates, with testing conducted before launch.

3. Ministry of Internal Affairs and Communications and Ministry of Economy, Trade and Industry, AI Guidelines for Business (Version 1.2) (31 March 2026).
4. Bank of Japan, Use and Risk Management of Generative AI by Japanese Financial Institutions -Based on the Results of FY2026 Survey- (24 August 2026).

Second, customer explanation and warnings. Among the practices discussed in the AIDP are informing customers that they are interacting with GenAI, warning them that responses may contain errors, identifying underlying information sources where appropriate, and providing a route to human support.[5]

Third, continuous verification and monitoring. Relevant controls include retaining and reviewing conversation logs, testing for skewed product recommendations, documenting recommendation logic and, where appropriate, obtaining third-party review.

“

The AIDP also refers to “agile governance”: repeatedly reassessing risks, objectives and controls rather than treating governance as a one-off exercise.

Fourth, firm-wide governance. Senior-management involvement, clear responsibility for AI, employee literacy, risk-proportionate controls and continuous review across the AI lifecycle are important. The AIDP also refers to “agile governance”: repeatedly reassessing risks, objectives and controls rather than treating governance as a one-off exercise.

These measures are not legal duties in themselves. Nevertheless, the FSA has indicated that it will refer to the practices described in the AIDP when responding to consultations concerning AI. They are therefore likely to become important practical reference points for institutions introducing customer-facing GenAI.

5. Securities Regulation: Applying Existing Rules to Generative AI

One important feature of the AIDP is that it addresses concrete regulatory questions facing financial instruments business operators rather than merely stating high-level AI principles.

The first concerns intra-group AI development and non-public information. Article 44-3(1)(iv) of the FIEA, together with Article 153(1)(vii) of the Cabinet Office Order on Financial Instruments Business, etc. (the “**FIB Order**”), restricts certain exchanges of non-public issuer information between a Type I financial instruments business operator engaged in securities-related business and its parent or subsidiary corporations.

Article 153(1)(vii)(g), however, provides an exception for information necessary for the maintenance and management of an electronic data processing system where appropriate safeguards against leakage are in place. On the premise that “maintenance and management” has historically been understood to include system development, the AIDP indicates that non-public information need not necessarily be removed in advance from customer conversation data used for AI-system development. AI development therefore does not automatically attract stricter information-sharing rules than conventional system development.

The second issue is access to corporate-related information. Article 40(ii) of the FIEA and Article 123(1)(v) of the FIB Order require necessary and appropriate measures to prevent unfair trading in connection with the management of corporate-related information. Japan Securities Dealers Association rules also require controls preventing such information from reaching divisions that do not need it.

The AIDP indicates that granting data scientists or similar personnel access does not, by itself, establish a control failure. The analysis remains substantive: institutions should identify the purpose and manner of AI use, access rights and information-management procedures, and assess whether the overall control framework is appropriate.

The third issue is AI product recommendations and “solicitation”. The FIEA contains no general statutory definition of solicitation. The AIDP states that solicitation is generally understood as conduct directed at specific users with the aim of inducing financial transactions, and that the same concept applies where AI is used.

5. FSA, Preliminary Discussion Points for Promoting the Sound Utilization of AI in the Financial Sector (Version 1.1) (3 March 2026) (Japanese). The FSA’s English webpage announces Version 1.1, but a full English translation of Version 1.1 was not located as of 29 August 2026; translations of Version 1.1-specific passages in this article are the author’s.

Where GenAI activity constitutes solicitation, existing conduct rules apply. Article 40(i) of the FIEA contains the suitability principle, while Article 38 prohibits, among other things, false statements and solicitation through conclusive evaluations of uncertain matters. The central regulatory question is therefore not whether a human or AI communicated with the customer, but whether the AI interaction functioned to induce that customer to enter into a financial transaction.

The FSA has indicated that, as AI-specific issues develop, relevant considerations may include AI's influence on investor decision-making and the degree of human involvement. This illustrates Japan's broader model: existing law is applied first, with ambiguities clarified through regulatory dialogue.

The FinTech Support Desk and FinTech Proof-of-Concept Hub support that process. Where a formal written interpretation is required, mechanisms such as the Grey Zone Elimination System under the Act on Strengthening Industrial Competitiveness may also be used.

6. Personal Data Protection: Key Issues for Generative AI

Personal data protection is among the most significant legal challenges for Japanese financial institutions using GenAI.

2026 Amendment to the APPI. As of August 2026, Japan has also enacted a significant amendment to the APPI. The amendment was passed by the Diet on July 10, 2026 and promulgated on July 17, 2026. Among other changes, it creates exceptions to certain consent requirements where personal data is used exclusively for the creation of statistical information or similar outputs, a category that the PPC expressly states may include certain forms of AI development. The amendment is intended in part to facilitate data use, including for AI, while maintaining safeguards against uses that could adversely affect individuals' rights and interests. However, most provisions of the amendment have not yet entered into force and will take effect on a date to be specified by Cabinet Order within two years of promulgation. Accordingly, financial institutions should distinguish between the APPI requirements currently in force and the forthcoming framework under the 2026 amendment.

The AIDP discusses several recurring APPI issues by reference to views expressed by experts in the AI Public-Private Forum. They should therefore not all be read as new, definitive interpretations by the FSA or the Personal Information Protection Commission.

The first is the purpose of use for AI development and training. Under the APPI, personal information must generally be handled within a specified purpose of use. Existing guidance recognizes that processing undertaken solely to produce statistical or otherwise non-identifiable results need not always be separately stated as an independent purpose.[6] The AIDP records expert views that similar reasoning may be relevant to certain forms of AI development and training. Where AI is used to make determinations about, or provide information to, individual customers, however, institutions should consider whether such handling can reasonably be anticipated by the data subject.

“

The central regulatory question is therefore not whether a human or AI communicated with the customer, but whether the AI interaction functioned to induce that customer to enter into a financial transaction.

The second issue is the input of personal data into external GenAI services. If the AI provider processes personal data only within the institution's instructions, the arrangement may constitute an entrustment of personal-data handling. The institution must then supervise the provider and confirm that input data is not used outside the institution's instructions for additional model training or the provider's own purposes. Enterprise contracts should therefore address data-use scope, model training, retention, sub-contracting, security and incident reporting.

The third issue is outsourced AI development, where similar controls over purpose, access, security and supervision apply.

6. Personal Information Protection Commission, Q&A on the Guidelines under the Act on the Protection of Personal Information (last revised 1 July 2025), 2-5.

The fourth is cross-border use. Whether the APPI's foreign-third-party rules apply is generally analyzed by reference to the location of the receiving business operator; an overseas server location alone does not necessarily make a transfer a foreign-third-party provision. Server location nevertheless remains relevant to the "external environment" that must be understood as part of security control measures.

"

Hallucination is likewise better treated as a residual risk to be managed through RAG, source display, guardrails, human confirmation and escalation than as a risk capable of complete elimination.

In practice, key safeguards include restricting unnecessary input of personal data into GenAI, using services contractually and technically restricted from using customer inputs for additional model training or other provider purposes.

7. Explainability, Hallucination and Model Risk

As GenAI becomes customer-facing, reliability and explainability become increasingly important. Where AI is used in significant financial judgments, institutions should be able to assess the reasonableness of outcomes and provide an appropriate explanation where necessary.

Because the internal reasoning behind individual GenAI outputs may be difficult to explain fully, governance often shifts towards process-level verifiability: recording the data used, RAG sources, prompts, outputs and human review. Hallucination is likewise better treated as a residual risk to be managed through RAG, source display, guardrails, human confirmation and escalation than as a risk capable of complete elimination.

GenAI also challenges conventional model-risk management. Institutions are seeking to adapt existing model-risk frameworks,[7] but GenAI may produce different outputs from identical inputs, evaluation methodologies remain immature and external foundation models may change frequently. Maintaining an inventory of significant AI systems, including their purposes, assumptions, limitations and vendor changes, is therefore increasingly important.

8. Third Parties and Security

Financial institutions' GenAI use depends heavily on external foundation models and cloud services, making AI risk closely connected to third-party risk.

Existing outsourcing frameworks remain relevant, but AI providers raise additional issues concerning data use, model changes, sub-contracting, service continuity, security, intellectual property and vendor lock-in. Concentration on a small number of foundation-model providers may also create common vulnerabilities across the financial system.[8]

Cybersecurity risks include not only confidential information entered into prompts, but also prompt injection, data poisoning, model extraction and unauthorised employee use of external AI services. At the same time, GenAI may be abused for sophisticated phishing, deepfake-enabled identity fraud and forged documentation, creating new challenges for AML/CFT and fraud controls.

As AI agents increasingly perform transactions or fund transfers autonomously, entitlement management, authentication, human oversight and mechanisms for suspension or intervention will become increasingly important.

7. FSA, Principles for Model Risk Management (November 2021); see also FSA, progress report on advancing model risk management at financial institutions (December 2024).
8. Financial Stability Board, The Financial Stability Implications of Artificial Intelligence (November 2024).

9. Conclusion

Japan has chosen not to begin with detailed finance-specific GenAI rules. Instead, it applies existing regulation in a technology-neutral manner and develops regulatory practice through risk-based governance and continuous public-private dialogue.

The AIDP shows that this model is moving beyond abstract principles into concrete questions concerning customer-facing GenAI, securities regulation, personal data protection and AI governance. Its strength is flexibility: Japan can respond to fast-moving technology without unnecessarily duplicating existing financial regulation. Its weakness is that institutions may still lack a sufficiently clear benchmark for "how much is enough".[9]

The ultimate test is therefore not whether Japan eventually enacts AI-specific financial regulation. It is whether existing technology-neutral rules can be translated, through guidance, interpretation and supervisory dialogue, into sufficiently clear and practical standards as GenAI and AI agents move closer to the core of financial services.

The expansion of customer-facing GenAI and agentic systems in 2026 and 2027 will provide a significant test of that iterative regulatory model.

"

Its weakness is that institutions may still lack a sufficiently clear benchmark for "how much is enough".

9. See Daiwa Institute of Research, The Regulatory Approach to AI in the Financial Sector: Key Points of the AI Discussion Paper (Version 1.1) and International Comparison (15 April 2026) (in Japanese; title translation is the author's).

KAZAKHSTAN

KAZAKHSTAN - A PIONEER IN ARTIFICIAL INTELLIGENCE REGULATION: BALANCING FINTECH DEVELOPMENT AND CONSUMER RIGHTS PROTECTION

GRATA INTERNATIONAL

Yasmin Rashidova
Senior Associate

yrashidova@gratanet.com

+ 7 747 838 80 47

www.gratanet.com

BIO

Yasmin Rashidova is a Senior Associate at the Banking and Finance Department.

She advises international and Kazakhstani clients on cross-border financing, debt transactions and regulatory matters, with particular experience advising international financial institutions and commercial banks on financing and infrastructure projects in Kazakhstan and Central Asia. Her experience includes syndicated and unsecured financings, project and infrastructure finance, and transactions involving IFC, EBRD, ADB, AIIB, MIGA and other international financial institutions. She has also advised on financing transactions in the microfinance, transportation, mining, energy and aviation sectors, as well as capital markets matters.

GRATA INTERNATIONAL

GRATA INTERNATIONAL

Sakzhan Kanafin
Associate

skanafin@gratanet.com

+ 7 775 100 11 99

www.gratanet.com

BIO

Sakzhan Kanafin is an Associate at the Banking and Finance Department.

He advises Kazakhstani and international clients on financing and corporate matters, with a focus on cross-border transactions, debt financing and regulatory issues. His experience includes advising IFC, AIIB, EBRD and other international financial institutions, as well as international investors, on financing and infrastructure projects in Kazakhstan and Central Asia. Sakzhan has also assisted with multiple cross-border financing transactions involving leading Kazakhstani microfinance organisations, including the review of financing documentation and advice on Kazakh law requirements.

GRATA INTERNATIONAL

GRATA INTERNATIONAL

Zhan Jumaliyev
Junior Associate

zjumaliev@gratanet.com

+ 7 701 981 81 58

www.gratanet.com

BIO

Zhan Jumaliyev is a Junior Associate at the Banking and Finance Department.

He advises international financial institutions, investors and financial sector clients on financing transactions and regulatory matters. His experience includes advising EBRD and international investment funds on cross-border financing transactions involving Kazakhstani financial institutions and microfinance organisations. Zhan assists with the review and negotiation of financing and security documentation, corporate approvals, and provides advice on Kazakhs law requirements and legal opinions.

GRATA INTERNATIONAL

KAZAKHSTAN - A PIONEER IN ARTIFICIAL INTELLIGENCE REGULATION: BALANCING FINTECH DEVELOPMENT AND CONSUMER RIGHTS PROTECTION

1. KAZAKHSTAN 2026: AI AS A NEW REGULATORY IMPERATIVE FOR THE BANKING SECTOR

Kazakhstan entered 2026 with unprecedented regulatory and political momentum in the field of digital technologies.

On 6 January 2026, the President of the Republic of Kazakhstan, Kassym-Jomart Tokayev, signed a Decree declaring 2026 the "Year of Digitalisation and Artificial Intelligence". According to the President of the Republic, artificial intelligence ("**AI**") "has created a kind of dividing line between those countries that will manage to enter the future and those that will remain in the past". It is for this reason that AI and digital technologies have been designated as priority areas for the country's development.

> "AI has created a kind of dividing line between those countries that will manage to enter the future and those that will remain in the past.

In June 2026, the President of the Republic of Kazakhstan approved by Decree the Nationwide Strategy "Digital Qazaqstan" through 2029. The strategy marks Kazakhstan's transition from fragmented automation to an integrated digital-state model in which AI becomes a fundamental tool for managing the economy, the social sphere and public administration.

Kazakhstan has recently taken a number of systemic measures in this area. A dedicated Ministry of AI and Digital Development has been established, the national AI center alem.ai has commenced operations, and AI technologies are being actively deployed in the provision of public services.

Kazakhstan's banking sector is more advanced in terms of digitalisation than most neighbouring markets. AI has been integrated into banks' key operational processes, including credit scoring, fraud prevention (anti-fraud), service personalisation, KYC identification and AML monitoring.

GRATA INTERNATIONAL

2. LEGISLATIVE BREAKTHROUGH: LAW "ON ARTIFICIAL INTELLIGENCE"

2.1. Background to Adoption and Place in the Legislative Framework

The development of the legal framework governing AI in Kazakhstan began well before the adoption of the AI Law.

Despite the evident maturity of the market and the regulators' position, no comprehensive legislative act existed until autumn 2025. The AI Law became the first specialised legislative act of the Republic of Kazakhstan dedicated to AI.

2.2. Three-Tier Classification of AI Systems

The conceptual core of the AI Law is a risk-based approach implemented through a three-tier classification of AI systems according to the degree of their impact on the safety of users, society and the state:

Minimal-risk systems – systems whose malfunction or cessation of operation would have a minimal impact on their users;

- Medium-risk systems – systems whose malfunction or cessation of operation may reduce the efficiency of users' activities, cause non-pecuniary harm or result in material damage;
- High-risk systems – systems whose malfunction or cessation of operation may result in a social and/or technological emergency and/or significant adverse consequences for the defence, security, economy or infrastructure of the Republic of Kazakhstan or the vital activities of individuals.

The classification of a particular system into a specific risk category is carried out by its owner and/or holder in accordance with the rules for classification of informatisation objects. High-risk AI systems classified as critical information and communications infrastructure facilities, as well as those intended for the formation of state electronic information resources, are treated as state systems for the purposes of compliance with information security requirements.

In addition to classification by risk level, the AI Law provides for classification of AI systems according to their degree of autonomy: low-autonomy systems (where a decision is made by a human), medium-autonomy systems (autonomous decisions that may be adjusted by a human) and high-autonomy systems (where human adjustment is completely excluded or technically impossible).

2.3. High-Risk AI Systems in the Financial Sector

For the banking and fintech sectors, the key issue is which AI systems automatically fall within the high-risk category. Nevertheless, based on the functional definition of high-risk systems, a number of financial applications of AI are highly likely to fall within this category:

- Credit scoring systems (creditworthiness assessment algorithms that make or materially determine lending decisions): their malfunction or systematic bias may cause significant material damage to users and destabilise the credit market;
- AML/financial monitoring: anti-money laundering systems that make autonomous decisions to block accounts or suspend transactions affect citizens' constitutional rights to dispose of their property and may have significant adverse consequences for the economy;
- Biometric identification systems (online KYC) used for account opening and customer verification: failures or discriminatory patterns may result in widespread denial of access to financial services;
- Highly autonomous anti-fraud systems: automatic blocking of transactions as a result of false positives causes direct material damage to *bona fide* customers.

For high-risk systems, the AI Law establishes the following set of obligations:

1. the owner and/or holder shall implement a continuous risk management process throughout the entire lifecycle of the AI system, including: identification and analysis of known and foreseeable risks associated with intended use; assessment of risks arising from reasonably foreseeable misuse; implementation of measures to prevent and eliminate identified risks; and regular updating of the risk assessment, at least once a year. Where risks associated with prohibited functionalities are identified, the holder shall immediately take appropriate measures, up to and including suspension or complete termination of the system's operation.
2. mandatory audit of AI systems for holders seeking to have their systems included in the "list of trusted high-risk AI systems".
3. sector-specific state authorities establish and publish on their websites lists of trusted high-risk AI systems.
4. owners and holders shall maintain documentation for an AI system "depending on the degree of its impact on the safety, rights, freedoms and legitimate interests of individuals and on public order", in accordance with the list of documentation approved by the authorised body.
5. users shall be informed that goods, works and services are produced or provided using AI systems.

“

Users shall be informed that goods, works and services are produced or provided using AI systems.

2.4. Rights and Obligations of Participants in the Financial Sector

The AI Law regulates relations between owners and holders of AI systems (banks and fintech companies), users (customers), and authorised state bodies. In the financial sector, the following key user rights should be highlighted:

“

The Digital Code entered into force on 9 January 2026, the first comprehensive law of its kind, one that changes, once and for all, our relationship with gadgets, data and the state.

- the right to review the user agreement of an AI system;
- the right to protection of personal data and confidential information processed by an AI system;
- the right to receive explanations concerning the outputs of an AI system affecting the user's rights and legitimate interests;
- the right to request information regarding the data on the basis of which the AI system made a decision;
- the right to refuse to interact with an AI system unless such interaction is mandatory under the legislation of the Republic of Kazakhstan.

In relation to fully automated lending decisions, the latter right raises significant practical issues regarding the possibility of reviewing an automated refusal by referring the loan application to an authorised bank employee for consideration. The AI Law expressly provides that requirements applicable to decisions based solely on automated processing of personal data are established by the legislation on personal data.

As regards owners and holders of AI systems , the AI Law imposes, *inter alia*, the following obligations:

- to manage risks associated with AI systems;
- to take measures to ensure the security and reliability of AI systems, including protection against unauthorised access and failures;
- to maintain documentation for AI systems;
- to provide user support in relation to the operation of AI systems;

to provide users with an opportunity to review the user agreement before commencing use of an AI system).

The Agency of the Republic of Kazakhstan for Regulation and Development of the Financial Market (the “**ARDFM**”) adopted Resolution No. 38, effective from 1 November 2025, which supplemented these obligations with cybersecurity requirements applicable to financial institutions, including mandatory biometric authentication and two-factor authentication for customer-facing services.

3. DIGITAL CONSTITUTION OF KAZAKHSTAN

The Digital Code entered into force on 9 January 2026, the first comprehensive law of its kind, one that changes, once and for all, our relationship with gadgets, data and the state.

3.1. The Concept of Fully Automated Decisions in Banking

The Digital Code introduces into Kazakh law the concept of Fully Automated Decisions (“**FAD**”) decisions made entirely without direct human involvement. This category is borrowed from European regulation.

In the banking context, FAD manifests most clearly in credit-scoring systems. Modern second-tier banks widely apply machine-learning models to assess borrowers' creditworthiness.

3.2. The Requirement of Human Involvement

The central element of the FAD legal regime in the Digital Code is the citizen's mandatory right to human review.

First and foremost, the Digital Code establishes an obligation for a financial organisation to inform the borrower of the fact that a decision was made through FAD. The notification shall be given in an understandable form, contain the main factors that influenced the decision, and explain the right to challenge it.

If AI scoring denies a loan, the borrower is entitled to demand that the decision be reviewed with the involvement of an authorised bank employee. The Digital Code establishes essential requirements for such a review: the specialist shall carry out an independent assessment rather than simply confirm the algorithmic conclusion. Thus, "human involvement" within the meaning of the Digital Code is not a formal procedure but a substantive check.

3.3. Algorithmic Transparency and Explainability Requirements

The Digital Code introduces requirements for the explainability of algorithmic systems, which in the financial sector are implemented through several mechanisms. First, banks shall maintain registries of the FAD systems they use and disclose key parameters of their operation to the regulator. Second, when reviewing complaints about automated denials, the bank shall provide the borrower with information about exactly which factors had a decisive influence on the scoring outcome.

3.4. Protecting the Rights of Financial-Services Consumers

In addition to the right to challenge a decision, the Digital Code establishes an expanded set of consumer rights in relations with financial organisations that use FAD systems.

A key protective tool is the prohibition on discriminatory algorithms. The Digital Code directly establishes that FAD systems in the financial sphere may not make decisions based on data directly related to protected characteristics (nationality, sex, religious beliefs, etc.) or serving as proxies for them.

4. KEY RISKS

4.1. Data Localisation

The central legal barrier to the direct use of foreign cloud platforms and large language models such as ChatGPT (OpenAI) or Claude (Anthropic) is the national data localisation regime.

First, the Personal Data Law establishes the mandatory requirement that personal data shall be stored by the owner and/or operator, as well as by third parties, in a database located within the territory of the Republic of Kazakhstan. The Personal Data Law defines personal data as "data, including biometric data, relating to a specific or identifiable data subject, recorded on electronic, paper and/or other tangible media". Therefore, by direct operation of law, the localisation requirement extends to all information falling within this definition.

> "Human involvement" within the meaning of the Digital Code is not a formal procedure but a substantive check.

Moreover, this definition is formulated extremely broadly and includes any information allowing identification of the data subject: from surname, first name and individual identification number to information on their property status, transactions and contractual obligations.

Second, for the financial sector, this regime is significantly tightened by Article 69.2 of the Banking Law. Under this provision, banks guarantee the confidentiality of their clients' transactions, accounts and other information, the disclosure of which could harm the client.

Thus, the direct transmission of raw client data to foreign public clouds via standard APIs becomes legally impossible. The servers of foreign AI providers lie outside the legal framework and jurisdiction of the Republic of Kazakhstan, meaning that any transfer to them of information enabling client identification or revealing the content of their banking transactions qualifies as a direct violation of both personal data legislation and banking secrecy laws.

The practical response to this conflict has been the implementation of so-called Sovereign Data Gateways. These function as an intelligent insulating buffer at the boundary between a financial institution's internal perimeter and external Large Language Models (the "**LLM**") providers. The gateway's operating principle involves three sequential operations performed in real time before the request is sent outside the Kazakhstani perimeter:

- First, masking and depersonalisation: the gateway algorithms automatically detect personal data within the request text (IIN, names, payment card numbers, etc.) and replace them with synthetic equivalents that prevent reconstruction of the original identity.
- Second, tokenisation: all sensitive financial indicators are replaced with unique tokens, with the mapping table required for reverse depersonalisation stored exclusively on the bank's local server within Kazakhstan.
- Third, context anonymisation: the original prompt is reformulated so that the external model can solve a mathematical or logical task without being able to identify either the specific data subject or the specific financial institution.

"

The protection of citizens' interests does not impede technological development, but rather establishes clear legal boundaries within which such development may take place.

The implementation of such gateways formally ensures compliance, yet simultaneously significantly complicates the IT architecture of financial organisations. Nevertheless, within a stringent regulatory environment, it is precisely such multi-layered isolation that becomes the only legal means of combining the cognitive potential of advanced LLMs with the requirements of Kazakh law.

4.2. Mandatory Labelling of Synthetic Content

The intensive integration of artificial intelligence technologies into Kazakhstan's banking activities transforms transparency in algorithm-consumer interaction from a mere ethical aspiration into a strict legal imperative. The AI Law stipulates that consumers shall be informed that goods, works and services are produced or provided using artificial intelligence systems.

The problem of unlabeled AI use in financial communications has both formal-legal and substantive dimensions. The deliberate concealment of algorithmic involvement constitutes a violation of the imperative of Article 21 of the AI Law, but at a deeper level it undermines the very notion of informed consumer choice.

Non-compliance with the above requirements constitutes an administrative offence and attracts penalties. Upon a repeat violation, the competent authority is empowered to suspend or completely prohibit the operation of the AI system.

5. CONCLUSION

In conclusion, it can be stated that by 2026 Kazakhstan had developed a distinct model for regulating artificial intelligence in the financial sector, based on the principle of technological neutrality. The National Bank and ARDFM do not seek to impose administrative restrictions on the choice of specific architectural solutions. Instead, the regulators primarily focus on overseeing the outcomes of technology deployment and ensuring non-discriminatory access of market participants to infrastructure.

The central constraining and guiding instrument in this framework is the Digital Code of the Republic of Kazakhstan, together with the specialised AI Law and personal data legislation. These legislative acts establish a robust framework for the protection of consumer rights, comprising three fundamental elements: unconditional data sovereignty, preventing the uncontrolled transfer of bank secrecy outside the national jurisdiction, the prohibition of discriminatory practices, ranging from social scoring to biometric classification, and the right to transparency, implemented through mandatory labelling of synthetic content. Accordingly, the protection of citizens' interests does not impede technological development, but rather establishes clear legal boundaries within which such development may take place.

Kazakhstan's approach demonstrates that consistent risk-based regulation, supported by meaningful sanctions for non-compliance, can move the market from a stage of fragmented experimentation towards the deliberate development of reliable and accountable AI systems.

GRATA INTERNATIONAL

SOURCES

1. “AI Law Enters into Force in Kazakhstan” // Gov.kz (Ministry of Digital Development, Innovations and Aerospace Industry). 18 January 2026. URL: https://www.gov.kz/memleket/entities/maidd/press/news/details/1143060?lang=ru.
2. “The Year of Digitalisation and AI in Kazakhstan: How the New AI Law and Financial Regulation Are Changing the Game for Banks and Big Tech” // Toppress.kz. 15 February 2026. URL: https://toppress.kz/article/god-cifrovizacii-i-iskusstvennogo-intellekta-v-kazahstane-kak-novii-zakon-ob-ii-i-finregulirovanie-menyayut-igru-dlya-bankov-i-bigtech.
3. Gumar N.A. “The Impact of AI and Voice Technologies on the Efficiency of Banking Operations in Kazakhstan” // Statistics, Accounting and Audit. 2025. Vol. 4. No. 99. Pp. 167–179. DOI: 10.51579/1563-2415.2025.-4.12. URL: https://sua.aesa.kz/main/article/view/398.
4. “The AI Law in Kazakhstan: What You Need to Know” // Informburo.kz. 24 February 2026. URL: https://informburo.kz/cards/zakon-ob-ii-v-kazaxstane-cto-vazno-znat-vsem-kazaxstancam-o-pravilax-i-strafax-v-2026-godu.
5. Law of the Republic of Kazakhstan dated 21 May 2013 No. 94-V “On Personal Data and Their Protection” // Adilet Legal Information System. URL: https://adilet.zan.kz/rus/docs/Z1300000094/z13094.htm.
6. Law of the Republic of Kazakhstan dated 17 November 2025 No. 230-VIII “On Artificial Intelligence” // Adilet Legal Information System. URL: https://adilet.zan.kz/rus/docs/Z2500000230.
7. Law of the Republic of Kazakhstan dated 16 January 2026 No. 258-VIII “On Banks and Banking Activity in the Republic of Kazakhstan” // Adilet Legal Information System. URL: https://adilet.zan.kz/rus/docs/Z2600000258.
8. National Payment Corporation of the National Bank of the Republic of Kazakhstan. Report “Artificial Intelligence in Kazakhstan’s Financial Market: Current State, Prospects and Analysis of Regulatory Approaches”. April 2024. URL: https://prg.kz/document/?doc_id=33467322.
9. “How the ARDFM Will Regulate the Adoption of AI in Banks” // Prodengi.kz. URL: https://prodengi.kz/post/kak-arrfr-budet-regulirovat-vnedrenie-ii-v-bankax.
10. Code of the Republic of Kazakhstan dated 5 July 2014 No. 235-V “On Administrative Offences” // Adilet Legal Information System. URL: https://adilet.zan.kz/rus/docs/K1400000235 .
11. Oralbayev Ch. “Key Aspects of the Law of the Republic of Kazakhstan ‘On Artificial Intelligence’ No. 230-VIII” // Paragraf Information System. 18 January 2026. URL: https://prg.kz/document/?doc_id=33297881.
12. Resolution of the ARDFM Board dated 20 August 2025 No. 38 (on banking sector policy and AI risks).
13. Decree of the President of the Republic of Kazakhstan “On Approval of the Nationwide Strategy for Large-Scale Digitalisation and Comprehensive Implementation of Artificial Intelligence Technologies ‘Digital Qazaqstan’ through 2029” // Kazpravda.kz. URL: https://kazpravda.kz/n/ukaz-prezidenta-respubliki-kazahstan-02-jw/.
14. Decree of the President of the Republic of Kazakhstan dated 6 January 2026 “On Declaring 2026 the Year of Digitalisation and Artificial Intelligence” // Akorda.kz. URL: https://www.akorda.kz/ru/ob-obyavlenii-goda-cifrovizacii-i-iskusstvennogo-intellekta-601222.
15. Digital Code of the Republic of Kazakhstan dated 9 January 2026 No. 255-VIII // Adilet Legal Information System. URL: https://adilet.zan.kz/rus/docs/K2600000255.
16. “What Are Depersonalisation, Masking and Tokenisation – Is There a Difference Between These Terms?” // Cisoclub.ru. URL: https://cisoclub.ru/chto-takoe-obezlichivanie-maskirovanie-i-tokenizacija-est-li-raznica-v-jetih-terminah/.
17. “What Is a Gateway? An Easy-to-Understand Explanation” // Alotceriot.com. URL: https://www.alotceriot.com/what-is-a-gatewayalotcer-gatewaygateway-definitiongateway-definitionindustrial-automationpeer-to-peer-communicatiorouting-functionsdata-forwardingprotocol-gatewayapplication-gatewaysecurity/.
18. “What Are Security Gateways and What Functions Do They Perform” // Dtu.kz. URL: https://www.dtu.kz/blog-post/chto-takoe-shlyuzy-bezopasnosti-i-kakie-funkczii-oni-vypolnyayut/.
19. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). 27 April 2016. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng.

GRATA INTERNATIONAL — YOUR TRUSTED LEGAL PARTNER IN KAZAKHSTAN AND BEYOND

About Our Firm

GRATA International is a dynamically developing international law firm providing legal support for projects in 28 countries worldwide. Our team helps companies navigate legal systems, manage cross-border projects, and ensure regulatory compliance — offering practical, business-oriented solutions.

We are proud that GRATA International is consistently recognized by leading international legal directories such as Chambers Global, Chambers Asia-Pacific, The Legal 500, IFLR1000, WWL, and Asialaw Profiles. In 2024, the firm was awarded "Law Firm of the Year" at the AIFC Legal Awards. These achievements confirm our commitment to high-quality service and the trust we have earned from clients across diverse industries.

GRATA International's Global Presence

GRATA International provides legal services across Kazakhstan, with offices in Almaty, Astana, Atyrau, and other regions of the country. Our strong local team ensures quality support for clients in every part of Kazakhstan

GRATA International has offices in the following countries:
Armenia (Yerevan), Azerbaijan (Baku), Belarus (Minsk), Georgia (Tbilisi), India (New Delhi), Indonesia (Jakarta), Kazakhstan (Aktau, Almaty, Atyrau, Astana, and other cities), Kyrgyzstan (Bishkek), Malaysia (Kuala Lumpur), Moldova (Chisinau), Mongolia (Ulaanbaatar), Oman (Muscat), Philippines (Manila), Russia (Moscow, St. Petersburg, Rostov-on-Don, Samara), Serbia (Belgrade), Tajikistan (Dushanbe), Thailand (Bangkok), Turkmenistan (Ashgabat), Türkiye (Istanbul, Ankara), Ukraine (Kyiv), Uzbekistan (Tashkent).

We also have representatives in:
Germany (Munich, Frankfurt), United Kingdom (London), USA (New York), China (Beijing), Switzerland (Zurich), UAE (Dubai), Cyprus (Limassol).

Why Choose GRATA International?

- Over 34 years of legal expertise
- 270+ lawyers with cross-border project experience
- Presence in 28 countries through offices and representatives
- More than 46,000 successfully completed projects
- Trusted by over 15,000 clients worldwide

Industries

- Automotive Industry
- Banking & Finance
- Industry & Trade
- Construction & Infrastructure
- Technology, Media & Telecommunications
- Pharmaceuticals & Healthcare
- Oil & Gas
- Mining
- Transport

Practice Areas

- Antitrust and Competition
- Commercial Contracts
- Corporate and M&A
- Dispute Resolution
- Data Protection & Privacy
- Employment
- Environmental Law
- Finance and Securities
- Project Finance & Public-Private Partnership (PPP)
- Subsoil Use
- Real Estate
- Intellectual Property
- Customs Law, International Trade & WTO
- Licenses and Permits
- Restructuring & Insolvency
- The Acting Law of the AIFC
- Tax
- Islamic Finance

Learn more:

www.gratanet.com
info@gratanet.com

KAZAKHSTAN

FROM EXPERIMENTATION TO TRUSTED SCALE

WHY KAZAKHSTAN COULD BECOME CENTRAL ASIA'S FIRST AI-NATIVE BANKING MARKET

KPMG

Konstantin Aushev
Partner

kaushev@kpmg.kz

+7 727 298 0898

www.kpmg.com/kz/ru

BIO

Konstantin Aushev is a Partner and Head of Technology at KPMG Caucasus and Central Asia. With more than 15 years of professional experience, he advises organizations on digital transformation, data management, enterprise architecture, cybersecurity, IT governance and business continuity. His portfolio includes over 70 advisory and technology implementation projects across the financial, telecommunications, aviation, transport, public services, metals and mining, and oil and gas sectors. Konstantin has supported clients with digital vision development, technology selection, IT control assessments, regulatory transformation and enterprise architecture design. He regularly contributes expert commentary on digital transformation, fintech and the evolving business models of financial institutions. Konstantin studied Applied Mathematics and Applied Physics at the Moscow Institute of Physics and Technology and Economics at HSE University. He holds CISA, ITIL, COBIT, XBRL and ISO 27001 qualifications.

KPMG

FROM EXPERIMENTATION TO TRUSTED SCALE
WHY KAZAKHSTAN COULD BECOME CENTRAL ASIA'S FIRST AI-NATIVE BANKING MARKET

Introduction

Kazakhstan Has Already Completed the Digital Banking Revolution

The latest data confirm a transformation that Kazakhstan completed years ago. In July 2026 alone, consumers made 1.3 billion cashless card transactions worth KZT17.6 trillion — up 7.5% and 8.4% year on year, respectively.[1] Internet and mobile banking generated 77.6% of those transactions and almost 90% of their value. These are not the early signs of digital adoption. They are the operating metrics of a financial market in which mobile banking, remote lending and cashless payments have long been part of everyday life. Kazakhstan is therefore approaching AI from a different starting point than many emerging markets: it is not trying to digitise banking and introduce intelligence at the same time. It is adding intelligence to a system that is already digital by default.

> "That is why Kazakhstan's next banking contest will not be won by launching another chatbot.

The significance lies not only in adoption, but in what has been built on top of it. Kaspi.kz has combined payments, commerce and consumer finance in a super-app used in everyday life. Halyk Bank, the country's largest traditional lender, has assembled a broad digital ecosystem. Freedom Group has linked banking with brokerage, insurance and lifestyle services. Bank CenterCredit has built an open ecosystem of different local and global partners. Together, these models have produced something more valuable than convenient mobile channels: a continuous stream of customer, transaction and risk data.

That is why Kazakhstan's next banking contest will not be won by launching another chatbot. The real prize is the ability to turn already-digitised activity into faster lending decisions, sharper fraud detection, more effective compliance and lower operating costs. However, the opportunity creates a harder test: banks must demonstrate that AI produces measurable returns and that its decisions can be explained, challenged and governed. As the Digital Tenge (a national CBDC project) moves towards wider use and Kazakhstan strengthens its banking and digital rules, the market is shifting from experimentation to accountability.

The Most Valuable AI Use Cases Are Invisible to Customers

The most valuable AI in banking is rarely the most visible. Customer-facing assistants attract attention, but the larger economic prize lies in improving high-frequency decisions across risk, credit and back-office operations. This is also where rhetoric must give way to evidence. KPMG's 2026 research[2] finds that 74% of organisations across Central Asia report business value from AI, yet only 24% achieve returns across multiple use cases. The gap is not adoption; it is repeatability.

1. The statistics of National Payment Corporation (https://npck.kz/statistika/).
2. KPMG Tech Report 2026 (https://kpmg.com/kz/en/insights/2026/05/tech-report-2026.html).

Bots are the most visible starting point in Kazakhstan's banking AI journey: they are easy to launch, demonstrate and count. Banks therefore lead with scale. Kaspi.kz, the country's dominant consumer super-app, reported that its assistant handled 64% of chats and 38% of voice calls in 2025; Halyk Bank, Kazakhstan's largest traditional lender, said its voice robot successfully completed 53% of routed calls in 2022.[3] These figures may prove that automation can absorb traffic, but not necessarily that it resolves the customer's problem. Some public app-store reviews frequently complain about circular bot conversations and difficulty reaching a human operator. The frustration is not uniquely Kazakhstani: a 2025 global survey found that 60% of customers believe chatbots often fail to understand their issue, while 71% still prefer a human agent.[4] In banking, where requests are sensitive, contextual and urgent, a scripted answer that misreads intent can turn convenience into friction. Cybernet AI, a Kazakhstan-based conversational-AI provider, offers a more credible model because it does not promise a universal digital banker. Its voice agents are tailored to defined workflows such as debt collection. The company says its services are used by 70% of Kazakhstan's major banks and can support tens of thousands of simultaneous conversations, including up to 1,000 calls a minute. Bots are therefore a useful first step — but the real test is narrow design, reliable escalation and successful resolution, not traffic absorbed.[5]

The larger opportunity begins where transaction volumes already exceed human capacity. In fraud and compliance, Kazakhstan's National Anti-Fraud Centre expanded from 155 participants in 2024 to 270 by mid-2026,[6] creating shared infrastructure for incident exchange and blocking. Predictive session and behavioural analytics, launched as a pilot in 2025, could move the system from reacting to reported fraud towards identifying patterns earlier. But this potential is not yet the same as proven value. Banks still need evidence on prevented losses, detection rates, false positives and investigation time before pilots can be described as industrial intelligence.

Credit scoring looks like the most obvious use of AI in finance. Kazakhstan's banks have searched for predictive signals for years — sometimes even proving their sense of humour. In 2015, an analysis of four years of data from the country's First Credit Bureau produced a national "credit horoscope", claiming differences in repayment behaviour by zodiac sign and gender.[7] The exercise was memorable precisely because it exposed a basic modelling trap: correlation is easy to discover; a stable, causal and defensible credit signal is much harder. Production lending therefore still relies largely on credit history, verified income, debt-service capacity and transparent scorecards, with machine learning used selectively rather than allowed to decide unchecked. The reasons are practical, not ideological. Alternative variables can become proxies for protected or socio-economic characteristics; historical data can reproduce earlier lending bias; complex models drift as customer behaviour changes; and a bank must explain why one applicant was declined while another was approved. Kazakhstan's 2025 AI Law raised the bar further by prohibiting social scoring and requiring lifecycle risk management for high-risk systems. AI's near-term value is therefore less in replacing the credit decision than in enriching it: detecting inconsistencies, identifying early stress and challenging the conventional score — while retaining accountable human and model-risk oversight.

“

Bots are therefore a useful first step — but the real test is narrow design, reliable escalation and successful resolution, not traffic absorbed.

Kazakhstan has therefore reached an important midpoint, not the finish line. The infrastructure, data volumes and digital customer behaviour are already in place; the full AI dividend is not. It will emerge only when banks move beyond isolated tools and redesign end-to-end decisions across risk, lending and operations — while keeping those decisions measurable, explainable and open to human challenge.

3. В Kaspi все чаще отвечают роботы: как теперь работает поддержка (https://finratings.kz/news/12708-v-kaspi-vse-chashche-otvechaiut-roboty-kak-teper-rabotaet-podderzhka/).
4. Chatbots are on the rise, but customers still trust human agents more (https://theconversation.com/chatbots-are-on-the-rise-but-customers-still-trust-human-agents-more-259980).
5. Ботофермы: как казахстанские банки развивают текстовых и голосовых ботов (https://kz.kursiv.media/2022-07-29/botofermy-kak-kazahstanskie-banki-razvivajut-tekstovyh-i-golosovyh-botov/).
6. The statistics of National Payment Corporation (https://npck.kz/statistika/).
7. Как знаки зодиака влияют на поведение по оплате кредитов: 20 сентября 2015 04:09 - новости на Tengrinews.kz (https://tengrinews.kz/private_finance/kak-znaki-zodiaka-vliiaiut-na-povedenie-po-oplate-kreditov-281137/).

Data Lineage Is Becoming More Important Than Model Accuracy

Kazakhstan's data story is paradoxical. The country's official statistical system is improving: the World Bank gave it 84.9 out of 100 in 2024, while the Open Data Inventory ranked it 44th globally. Yet these national scores say little about the information hidden inside companies. KPMG's 2025 Kazakhstan survey found that 50% of organisations believed their data infrastructure was not ready for AI.[8] Globally, 64% of executives still identify organisational data quality as a major GenAI barrier.

> In regulated banking, the decisive question is no longer whether data exists. It is whether every important decision can be traced back to data the bank can defend.

Banks start from a better position. Years of digital payments, remote identification, lending and regulatory reporting have created large stores of structured, time-stamped data. Kazakhstan's national financial infrastructure already processes more than 350,000 transactions worth KZT5.9 trillion each day and performs over 2.5 million digital identifications a month.[9]

Banks may hold better data than most Kazakhstani companies, but greater volume raises the standard of proof. Management, supervisors and customers must be able to see where a material data point came from and how it shaped a decision. Kazakhstan is not starting from zero: banks face prudential reporting, board-level risk-management and internal-control duties; the Credit Bureau Law regulates credit histories; separate rules cover information security; and personal-data legislation protects individuals affected by automated processing. The 2025 AI Law adds lifecycle risk management and technical documentation for high-risk systems. Yet these regimes address adjacent risks rather than create one banking standard for AI data quality. They do not consistently require authoritative source data, end-to-end lineage, representative datasets, bias testing or drift monitoring. BCBS 239 provides the banking benchmark — accurate, complete, timely and adaptable risk data with clear ownership — while Article 10 of the EU AI Act requires documentation of data origin, preparation, gaps and bias controls for high-risk systems. Kazakhstan need not copy either framework wholesale. It needs the disciplines behind them: a board-approved inventory of critical data, named owners, measurable quality thresholds, independent testing and a customer route to inspect and correct data used in credit, fraud or account-restriction decisions. Without these controls, management sees impressive model metrics but cannot verify the evidence beneath them, while customers are asked to trust decisions they cannot challenge.

As KPMG's global financial-services research puts it, "AI is just the tip of the iceberg"; the real foundation is high-quality data, simplified processes and resilient infrastructure. In regulated banking, the decisive question is no longer whether data exists. It is whether every important decision can be traced back to data the bank can defend.

Banks Must Be Able to Defend AI Decisions, Not Just Build Them

Kazakhstan's financial sector benefited from letting experimentation run ahead of formal governance. Over the past three years, banks allowed business units, innovation teams and technology vendors to test assistants, scoring models and fraud tools because prototypes were cheap, digital data was abundant and competitive pressure was real. That speed helped create practical experience before committees and rulebooks could slow it down. The KPMG CCA Tech Report describes this early phase as "AI roulette" — scattered bets on multiple technologies before organisations learn how to operate them at scale. Across Central Asia, 74% of organisations report business value from AI, yet only 24% achieve returns across multiple use cases.

That imbalance was useful at the start; it becomes costly when pilots enter production. The market is now correcting it. Banks are beginning to manage AI as a portfolio rather than a collection of demonstrations: maintaining central registers of models and use cases, classifying risk before development, assigning named business owners, requiring independent validation, defining human-override rules, monitoring production performance and reporting value, incidents and residual risk to the board. This is not bureaucracy added after innovation. It is the operating system that makes repeated innovation possible.

8. KPMG Tech Report 2026 (https://kpmg.com/kz/en/insights/2026/05/tech-report-2026.html).
9. The statistics of National Payment Corporation (https://npck.kz/statistika/).

The evidence from KPMG's AI in Finance[10] research is unusually clear. Organisations with stronger governance and controls report significant performance improvements at three to six times the rate of weaker peers. Yet only 42% say they are strongly ready to produce audit evidence for AI-enabled finance processes, and just 29% formally track AI-adoption failures. As the report concludes, "trust — operationalized through governance, controls and oversight — is the through-line of the leaders pulling ahead."

Kazakhstan has therefore entered a second phase. The first asked how many AI ideas a bank could test. The next asks which systems should exist, who owns their outcomes, what evidence they must retain and when a human must intervene. Competitive advantage is moving from model development to model governance.

Agentic AI Could Reshape Banking Faster Than Core Platform Transformation

Kazakhstan may be building the payment rails that agentic banking needs before the business model has fully arrived. Its Interbank Mobile Payment System now supports real-time transfers by phone number and interoperable QR payments; Open Banking exposes consent-based account data; and the Digital Tenge has moved from pilots towards industrial deployment, with programmability and smart contracts identified as the next frontier. The new Banking Law gives digital tenge and digital financial assets a legal foundation.

Agentic AI could connect these rails. A customer might instruct an agent to monitor liquidity, compare offers, move surplus cash, pay invoices or execute purchases within agreed limits. The change is profound: today, software recommends and the customer pays; tomorrow, software may decide, initiate and reconcile. KPMG finds that 88% of organisations are already investing in agentic AI, while agentic deployers report a 32-point performance advantage over non-deployers. KPMG also warns that "the future will see AI agents talking to AI agents."

Kazakhstan should prepare before autonomous payments become commonplace. It is hardly possible and hardly feasible to centralise the autonomous payment infrastructure. Instead, the National Bank and market should define a verifiable identity for every payment agent; granular and revocable customer mandates; transaction, merchant and time limits; mandatory confirmation for high-risk payments; a liability hierarchy covering customer, bank, fintech and model provider; machine-readable receipts showing why a payment was made; real-time fraud controls; and a regulatory sandbox linking Open Banking, instant payments and the Digital Tenge. Banks should register every agent, assign an accountable owner and test it in shadow mode before giving it payment authority.

Kazakhstan's advantage is timing. It can design agentic payments into new national infrastructure rather than bolt controls onto an established market after failures occur.

> "Competitive advantage is moving from model development to model governance.

From Experimentation to Trusted Scale: The Real AI Challenge for Kazakhstan's Financial Sector

Kazakhstan has many of the ingredients needed to become Central Asia's first AI-native financial market. Whether that ambition becomes a durable advantage will depend on three constraints: talent, infrastructure and regulatory coherence.

10. 2026 KPMG Global AI in Finance Report (https://kpmg.com/kz/en/insights/2026/06/kpmg-global-ai-in-finance.html).

The first is talent. A market of Kazakhstan's size cannot easily supply enough experienced AI engineers, data scientists, model-risk specialists and product leaders. KPMG's global financial-services research finds that 40% of institutions lack the talent needed to execute their technology plans. The government is responding at scale: **AI-Sana** aims to provide basic AI training to 650,000 students,[11] advanced modules to 100,000 and support to 1,500 start-ups. The numbers are impressive, but the harder task is turning mass education into production-grade expertise inside banks. The government has also started implementing AI literacy in school education.

> “
>
> Kazakhstan's next AI breakthrough will not be another pilot.

The second is infrastructure. Data Center Valley in Ekibastuz — planned at up to 1 GW, with 1,400 hectares allocated and an initial 100 MW site prepared — could provide sovereign computing capacity and attract global cloud providers. But much remains untested: confirmed demand, access to advanced chips, grid reliability, cooling and water requirements, connectivity economics and the timetable beyond the first facility. Government commitment is strong; the commercial model is still emerging.

The third is regulatory coherence. The **AI Law, Digital Code**, new **Banking Law**, National security rules, personal-data rules and financial-sector requirements each address part of the problem. Banks, however, need one coherent operating interpretation of risk classification, automated decisions, human review, data localisation, model accountability and customer redress.

Kazakhstan's next AI breakthrough will not be another pilot. It will be the moment when skills, computing capacity, trusted data and coherent regulation begin to function as one system — and when banks can scale AI without asking customers or supervisors to take its reliability on faith.

11. AI-Sana (https://aisana.gov.kz/?lang=en).

KPMG

MAKE THE FUTURE

KPMG brings together technology and risk expertise to help businesses transform, unlock value from data and AI, and strengthen cyber security

KPMG. Make the Difference.

MEXICO

MEXICO: ARTIFICIAL INTELLIGENCE IN FINANCIAL SERVICES

MOONCLOUD LEGAL

Gabriela Salazar Torres
CEO and Founder

gaby@mooncloudlegal.tech

+52 56 1996 5905

www.mooncloudlegal.tech

BIO

Gabriela Salazar Torres is a legal strategist working where traditional finance meets emerging technology. She brings more than 30 years in the financial sector and over a decade focused specifically on fintech, AI, blockchain, and virtual assets — a combination that has made her one of Mexico's leading authorities on digital-asset regulation and innovation.

She is the founder of MoonCloud LegalTech, where she helps financial institutions and technology companies design governance and compliance frameworks for blockchain, virtual assets, and artificial intelligence. Her credentials span both fields: advanced specialized certifications in blockchain from the Blockchain Center of the University of Zurich and the Complutense University of Madrid, and certification as a Chief Artificial Intelligence Officer, placing her among the early specialists in AI governance and risk management.

An author and commentator on fintech, blockchain, securities and anti-money-laundering regulation in Mexico, Gabriela works on the legal architecture that will define the next phase of digital finance.

MOON CLOUD
by Gaby Salazar

MOONCLOUD LEGAL

Gustavo
Managing Partner

gsalaiz@web23advisors.com

+52 55 8532 4864

www.web23advisors.com

BIO

Gustavo has over 15 years of experience at the intersection of financial law, regulation, and technology, with a career focused on the legal and regulatory aspects of financial services and fintech in Latin America. He was the first Director of the Regulatory Sandbox at Mexico's National Banking and Securities Commission (CNBV), where he led the implementation of Mexico's Fintech Law innovation framework and coordinated the assessment of innovative business models with legal, supervisory, and regulatory policy teams.

Today, through Web23 Advisors, he serves as Of Counsel and strategic advisor to law firms and in-house legal teams on fintech regulation, compliance, AI governance, and legal technology adoption. His practice combines traditional legal analysis with the practical implementation of AI tools for legal teams.

Gustavo holds an engineering degree, an MBA from Carnegie Mellon University, and a diploma in Banking, Credit, and Risk Management from ITAM.

MOON CLOUD
by Gaby Salazar

MEXICO: ARTIFICIAL INTELLIGENCE IN FINANCIAL SERVICES

1. Overview.

Mexico has not enacted a general federal artificial intelligence statute. Legislative activity remains fragmented. Several initiatives would give Congress express constitutional authority over AI or create a general statutory framework, but none had entered into force by August 2026. Narrow binding rules already address algorithmic systems, notably the platform-work regime, requiring transparent task-allocation and an algorithmic management policy, and other AI-specific reforms have advanced through Congress, none establishing a general regime for automated decisions between institution and customer.

> “
>
> Mexican financial law is functional rather than technological: it regulates the decision, the activity, the contract and the risk, even without explicitly referencing the model.

Mexican financial law is functional rather than technological: it regulates the decision, the activity, the contract and the risk, even without explicitly referencing the model. The relevant question is therefore not whether AI is “allowed” in Financial Services, but which regulated function it can perform in a financial system and what Mexican law requires of it. There are four important factors that must be taken into account when thinking of how AI interacts with Mexican law: a) data protection; b) sectoral product and conduct rules; c) the contractual, consumer-protection and non-discrimination framework; and d) prudential guidelines. Together they form a dispersed governance regime whose intensity varies greatly across diverse regulated entities, such as insurance, securities and investment funds, fintech, banking, popular finance and multiple-purpose finance companies.

2. Data protection

The Federal Law on the Protection of Personal Data Held by Private Parties enacted in March 2025, is the only horizontal federal regime expressly addressing automated processing by private-sector financial institutions, subject to statutory exclusions and the separate public-sector regime.

MOON CLOUD
by Gaby Salazar

The first principle is the right to object: a data subject may object when personal data undergoes automated processing producing unwanted legal effects, or significantly affecting the person’s interests, rights or freedoms, and evaluating personal aspects without human intervention, including predicting economic circumstances. A fully automated credit decision, credit-limit reduction or risk-based pricing decision can fall within that provision when met, but it does not capture every use of scoring, nor create a right to disclosure of source code or model weights, and it exempts processing necessary for a legal obligation.

The second principle is transparency: the privacy notice must communicate the existence and principal characteristics of the processing. Where automation materially determines access to a product, price or service, omitting that fact is hard to reconcile with the statutory standard. The law does not fix the notice’s granularity, but a defensible one should identify the automation’s purpose, the data categories used, the type of consequence it may produce, and the mechanism for exercising opposition and other data rights. The framework also reinforces accountability: an institution stays responsible for personal-data compliance even where a scoring model, biometric tool or conversational agent is supplied by a third party.

3. Product and conduct rules in the Mexican Financial System.

a) One of the clearest customer-facing model-governance regimes in Mexican financial law appears in the Insurance and Surety Institutions Law. Drafted for actuarial practice, its provisions never mention AI, yet they directly regulate the function an underwriting or pricing model performs.

The law requires premiums to be determined on technical bases, giving a high degree of certainty the insurer can meet its obligations, and requires each product to carry a technical note setting out the actuarial procedures, their justification and other relevant assumptions behind them. When a model materially drives pricing, segmentation or an actuarial assumption, the note must capture that methodology with the required precision.

Under current regulations, a licensed, certified actuary must sign the technical note. Product documentation must include a legal opinion and an opinion of congruence tying the note to the contract's terms. In certain instances that may affect policyholders, beneficiaries or the insurer's solvency, the National Commission of Insurance and Sureties may order adjustments. This resembles outcome-based monitoring, but triggers on a statutory threshold; material changes to the registered basis may also require prior regulatory approval. The same statute separately governs internal solvency models: authorization, validation, independent expert review and change controls. Insurance thus makes the article's central point concretely: Mexico need not use the word "AI" for model documentation, professional responsibility, change control and outcome monitoring to already be legal requirements.

b) Another relevant example, is the framework that governs brokerage firms, credit institutions, investment advisers, fund managers and distributors.

Under the general provisions applicable to entities providing investment services, such entities must determine the client's investment profile and each product's profile. Before issuing a personalized recommendation, they must confirm it is reasonable for the client by testing congruence among client profile, product profile, suitability and diversification policy, and must document that process. The same architecture extends to discretionary management, where the entity must justify each transaction executed without a separate instruction.

These requirements do not create a general technical right to explanation of an AI model. They do create a transaction-level suitability and evidentiary duty. A recommendation engine or robo-adviser must produce outcomes that the institution can explain against documented client profile, product and diversification parameters. The institution cannot answer a suitability challenge by saying that the algorithm selected the transaction.

The regulation goes further for electronic profiling tools: they must be used within the circumstances and markets for which they were designed, with the institution defining their scope of use, ensuring relevant personnel understand them, and maintaining means to validate results, plus documentary support and identifiable responsibility. These are direct model-governance controls, even if unlabelled as such.

The public investment services guide allows the conduct framework to be externally observable, and the regulatory file makes the profile and recommendation process auditable, neither requires publishing the underlying scoring logic or source code. The legal standard is thus not full algorithmic transparency, but controlled use of the tool, validation of its results, and evidence that each outcome was reasonable for the client.

c) Financial technology institutions included in the Financial Technology Institutions Law, are also governed through a combination of the latter and a set of general provisions issued for the fintech sector.

“

Mexico need not use the word “AI” for model documentation, professional responsibility, change control and outcome monitoring to already be legal requirements.

For crowdfunding institutions, for example, the Fintech Law requires disclosure of the criteria used to select applicants and projects, the information analyzed, verification activities, and the methodologies used to evaluate and classify them, consistently applied and updated as needed. A generic description may be insufficient where selection is model-driven and prevents investors from understanding the assessment framework, though the standard does not require publishing source code or every anti-fraud control.

Automatic investment is also regulated: it requires express consent separate from the adhesion contract and a mechanism for revocation, and the institution may not allocate funds to projects outside the client's prior consent. The issue is not whether an algorithm performs the allocation, but whether the mandate, limits and execution remain within the consent given.

The platform must remain consistent with the contract, and the institution cannot disclaim responsibility for what it offers or communicates through its platforms or authorized third-party channels. Applied to generative AI, a chatbot misstating a product's fees, risks or conditions is first a conduct and disclosure problem, the institution generally remains responsible to the customer, without prejudice to recourse against the technology provider.

> “
>
> In such cases automation inherits that legal architecture.

d) In banking, there is more nuance rather than a blanket claim that no specific model rule exists. Mexican law sets no horizontal customer-facing regime for AI-driven credit, pricing, collection or servicing decisions, but it does impose detailed model-governance requirements on certain prudential and internal-rating models.

Banks using internal-ratings-based methods for credit-risk capital need prior authorization from the National Banking and Securities Commissión. The framework covers implementation plans, model development and calibration, documentation of theory, assumptions, mathematics and data, integration into risk management, independent validation and back-testing, and a third-party model does not remove that burden. In general, banking regulation already treats model risk seriously where output affects prudential capital.

Furthermore, the banking regulatory and operational framework governing technology outsourcing and cloud usage applies directly to AI models. For example, general provisions applicable to banks, govern automated data processing systems and internal controls to ensure information integrity and confidentiality. Institutions are subject to rigorous requirements when hiring external providers for critical processes, particularly when data processing occurs outside the country. The framework underscores the non-delegable responsibility of executives and boards to manage operational risk and guarantee business continuity. No single rule requires every bank to document and explain every automated decision, though many surrounding controls apply.

e) Popular financial companies and cooperatives operate under their own sectoral rules on credit, risk management and internal control, varying by entity and level. Those regulations may cover automated tools but create no horizontal AI-specific regime, and a vendor platform does not displace the entity's own compliance duties.
f) SOFOMs need a further distinction. Regulated SOFOMs include entities with equity links to banks or popular-finance entities, plus debt issuers and entities that voluntarily elect regulated status; applicable rules vary by category, so not every regulated SOFOM answers to the entire banking framework. Non-regulated SOFOMs are subject to anti-money-laundering and user protection supervision, plus data-protection law, but generally lack a supervisor positioned to review an ordinary origination model.

4. The contractual, consumer-protection and non-discrimination framework.

The Financial Services Users' Protection Law prohibits abusive clauses, and general provisions applicable to financial entities, flag clauses creating indeterminable obligations, permitting termination without notice, or shifting to the user the burden of proving transactions with the institution. These can support challenges to dynamically imposed terms or automated line closures inconsistent with the registered contract; where a model changes a fee, the output must stay within the authorized framework. Some violations are shown by comparing output against the contract; discrimination or model-defect claims may still need evidence about the AI model itself.

5. Prudential Guidelines.

Financial entities in Mexico are not bound to observe prudential guidelines governing the use of AI such as ISO/IEC 42001.

6. Findings and outlook

Algorithmic model governance across the Mexican financial system is uneven. The strongest regimes are those where financial law had already formalized a professional judgment before AI arrived, the actuarial determination in an insurance product and the personalized investment recommendation are the clearest examples due to those acts already requiring documented justification and review. Therefore, in such cases automation inherits that legal architecture.

MOON CLOUD
by Gaby Salazar

Where automation does not displace a previously formalized professional activity, governance is more fragmented. Credit origination, fraud detection, collections, marketing, customer service and biometric authentication are governed through combinations of data protection, operational risk, contract, outsourcing, non-discrimination and conduct rules, enforceable, but not always producing a single model file, a clear validation standard, or a route for reconsidering a significant decision.

As previously pointed out, insurance, securities and investment services, fintech institutions, and banking each illustrate a common pattern from a different angle. None of these regimes were written with artificial intelligence in mind, and none use the term. Yet each of them already impose obligations that govern what an AI system does when it prices a policy, recommends an investment, screens a borrower, or drives a chatbot: documented methodology, professional accountability, validation of results, and consequences when outputs drift from what was authorized.

Additionally, legal exposure follows the nature of the function rather than the label of the function of an algorithmic model. A rule-based score, a machine-learning model and a generative agent may sit in different technical categories, but the analysis dictates what each system does: set a premium, recommend an investment, classify an applicant, calculate regulatory capital, deny credit, close a line or describe a product. Some breaches can be proved without reconstructing the model's logic; others such as discrimination, defective validation and causation, may need access to data, assumptions and change history.

Mexico would benefit from coordinated financial-sector guidance identifying significant automated decisions and setting minimum expectations for documentation, validation, change control, outcome monitoring, human escalation and vendor oversight. Existing powers over product registration, client profiling, operational risk, outsourcing and sound practices already provide institutional routes. A general AI statute may eventually add common vocabulary and risk classification, but supervisors need not wait for it.

What to do with all of this

Five priorities follow:

1. Classify the function first. Map each system to the financial decision it supports: underwriting, pricing, suitability, credit, fraud, AML, collections, servicing, marketing or communication. The applicable law follows the function and the entity's regulatory scope.
2. Identify significant automated effects. Determine where personal data are used without meaningful human intervention to produce legal or similarly significant consequences, and align the privacy notice, data-rights process and escalation channel with the actual system, without overstating a right the law does not create.
3. Build the evidence file. Record purpose, owner, data provenance, assumptions, validation, performance thresholds, limitations, overrides, changes, incidents and third-party dependencies. ISO/IEC 42001 can organize that system, but does not replace a technical note, suitability file, prudential approval or any other sector-specific requirement.
4. Reconcile output with the legal wrapper. Test model outcomes against registered contracts, product files, disclosed selection criteria, client mandates and pricing terms, a technically accurate model can still create a regulatory breach if its output exceeds the authority or disclosure on which the product rests.
5. Govern vendors and monitor outcomes. Contract for access, audit, change notification, incident cooperation and exit, and monitor drift, error rates and distributional effects. Establish meaningful human escalation for high-impact cases even though Mexican law does not yet impose a general right to human review.

> A technically accurate model can still create a regulatory breach if its output exceeds the authority or disclosure on which the product rests.

Mexico's framework is neither empty nor coherent. It is functional, fragmented and already enforceable. The immediate task is to connect existing financial-law duties to the systems now performing regulated decisions.

Sources:

1. Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP), https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf
2. Ley de Instituciones de Seguros y de Fianzas (LISF), https://www.diputados.gob.mx/LeyesBiblio/pdf/LISF.pdf
3. Disposiciones de carácter general aplicables a las entidades financieras y demás personas que proporcionen servicios de inversión. https://sidof.segob.gob.mx/notas/5701736
4. Ley para Regular las Instituciones de Tecnología Financiera (LRITF). https://www.diputados.gob.mx/LeyesBiblio/pdf/LRITF.pdf
5. Disposiciones de carácter general en materia de transparencia y sanas prácticas aplicables a las instituciones de tecnología financiera. https://www.dof.gob.mx/nota_detalle.php?codigo=5565233&fecha=09/07/2019
6. Disposiciones de carácter general aplicables a las instituciones de crédito (Circular Única de Bancos). https://www.cnbv.gob.mx/SECTORES-SUPERVISADOS/BANCA-MULTIPLE/Paginas/NORMATIVIDAD.aspx
7. Ley General de Organizaciones y Actividades Auxiliares del Crédito (LGOAAC). https://www.diputados.gob.mx/LeyesBiblio/pdf/LGOAAC.pdf
8. Ley de Protección y Defensa al Usuario de Servicios Financieros; and Disposiciones de carácter general en materia de cláusulas abusivas contenidas en los contratos de adhesión. https://www.diputados.gob.mx/LeyesBiblio/pdf/LPDUSF.pdf; https://www.dof.gob.mx/nota_detalle.php?codigo=5368784&fecha=19/11/2014
9. Ley Federal para Prevenir y Eliminar la Discriminación. https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPED.pdf
10. Ley Federal del Trabajo. https://www.diputados.gob.mx/LeyesBiblio/pdf/LFT.pdf
11. ISO/IEC 42001:2023. https://www.iso.org/standard/42001

ROMANIA

ARTIFICIAL INTELLIGENCE AS A NEW RISK CATEGORY IN M&A DUE DILIGENCE: A ROMANIAN PERSPECTIVE

LEXTERS

Alexandru Stănescu
Partner

alexandru.stanescu@lexters.com

+40 745 772 762

www.lexters.com

BIO

Alexandru Stănescu is a Partner at Lexters and leads the firm's FinTech work through its Tech & DeepTech practice. He advises crypto exchanges, custodians, payment providers, asset managers, token issuers and technology companies on MiCA, product structuring, cross-border investments, venture capital and M&A. Chambers FinTech 2026 ranks Alexandru in Band 1 for FinTech Legal in Romania, marking his third year in the guide.

A Fulbright scholar, Alexandru holds an LL.M. from Columbia Law School. He is admitted to the Bucharest and New York Bars and regularly speaks on international arbitration, blockchain and financial regulation.

Alexandru is known for a business-oriented approach. He works closely with clients to understand the commercial context of each matter, translating complex regulation into clear options and workable solutions. His advice combines rigorous legal analysis with commercial judgment and practical execution, helping clients advance new products, transactions and cross-border projects with confidence.

Lexters

LEXTERS

Simina Negulescu
Partner

simina.negulescu@lexters.com

+40 745 772 762

www.lexters.com

BIO

Simina Negulescu is a Partner at Lexters, where she advises technology companies, founders and investors on corporate law, venture capital, cross-border transactions and disputes. With more than a decade of experience, including nine years focused on the IT sector, she supports businesses from formation and financing to international expansion and group consolidation.

Simina combines transactional judgment with litigation instinct. Her early career in disputes, followed by experience as a fractional chief legal officer for a technology company, allows her to understand how legal risk develops and how businesses make decisions. She anticipates pressure points, structures workable solutions and remains closely involved through execution.

Her practice covers financing rounds, governance, commercial contracts, regulatory compliance, legal audits and due diligence. Clients value her can-do approach, meticulous preparation and ability to turn complex issues into clear commercial decisions.

Lexters

LEXTERS

Patricia Gorici
Senior Associate

patricia.gorici@lexters.com

+40 745 772 762

www.lexters.com

BIO

Patricia Gorici is a Senior Associate at Lexters. Her practice focuses on corporate law and EU regulatory matters. She advises technology companies and international businesses on data protection, digital services, e-commerce, commercial contracts and cross-border operations. Her work includes regulatory assessments, interactions with authorities, employment matters and dispute resolution.

Her academic training and continuing education align with her practice. Her LL.B. studies in EU and business law, specialist coursework in EU consumer law, participation in European law competitions and publications on EU law and GDPR all form part of a sustained engagement with the European legal framework. Taken together, they shape the perspective she brings to regulatory questions and to advising businesses operating across European markets.

Patricia is known for a practical, detail-oriented approach. She works closely with clients to understand how regulatory requirements affect their products, operations and commercial objectives, translating EU rules into clear steps, workable contracts and proportionate compliance measures. Her advice combines legal analysis with responsiveness and an understanding of day-to-day business needs, helping clients implement solutions efficiently across jurisdictions.

Lexters

ARTIFICIAL INTELLIGENCE AS A NEW RISK CATEGORY IN M&A DUE DILIGENCE: A ROMANIAN PERSPECTIVE

Lexters

1. Introduction: A New Regulatory Risk Enters the Deal Room

In a merger or acquisition, the value of a target is determined not only by its assets, revenues and market position, but also by the regulatory risks attached to its business. The scope of due diligence therefore evolves whenever a new regulatory framework turns what was previously regarded as a manageable operational issue into a source of material legal, financial and transactional exposure.

> “
>
> The AI Act is likely to establish AI compliance as a distinct category of M&A due diligence, much as the GDPR established data protection as an independent area of transactional review.

The General Data Protection Regulation (“*GDPR*”) illustrates that process. Before its adoption, data protection was often addressed within broader IT or regulatory compliance reviews. Following it, privacy compliance developed into a distinct due diligence workstream capable of affecting deal valuation, contractual risk allocation, closing conditions and the insurability of transactional risk.

The regulatory landscape surrounding Artificial Intelligence appears to be following a comparable trajectory. Rather than merely adding compliance obligations, Regulation (EU) 2024/1689 (“*AI Act*”) establishes an entirely new framework governing the development, deployment and use of Artificial Intelligence (“*AI*”) systems. Depending on the level of risk associated with a particular system, organisations may be required to implement measures relating to risk management, data governance, technical documentation, transparency, human oversight, accuracy, robustness, cybersecurity and post-market monitoring.

This article argues that the AI Act is likely to establish AI compliance as a distinct category of M&A due diligence, much as the GDPR established data protection as an independent area of transactional review. It considers why the risk is already live, where the exposure concentrates, how a buyer should structure an AI review and how the findings feed into transaction documentation. Because the AI Act is a Regulation, it applies directly and uniformly throughout the European Union, and the substantive analysis that follows is therefore common to every Member State. The vantage point, however, is Romanian: the observations below are drawn from our experience with M&A transactions on the Romanian market, where the questions this framework raises are already being asked in practice.

2. The GDPR Precedent: A Transactional Blueprint

The significance of the GDPR did not lie solely in new compliance obligations or in the possibility of administrative fines of up to 4% of worldwide annual turnover. Its transformative effect resulted from the combination of extensive substantive obligations, an accountability-based compliance model, significant documentation requirements and the possibility that historical non-compliance could generate substantial post-closing liability. Data protection therefore ceased to be merely an operational matter and became a legal risk capable of affecting enterprise value, negotiations and the allocation of post-closing liability.

Privacy compliance evolved into a dedicated workstream requiring buyers to assess not only the target’s policies, but its underlying data governance framework: lawful bases for processing, international data transfers, processor agreements, security measures, data breach history, regulatory investigations and internal compliance programmes. In technology-intensive transactions, these reviews were supplemented by specialised cyber and data privacy assessments, and the findings were carried through into the transaction structure itself.

More fundamentally, the GDPR demonstrated how regulation can change what due diligence examines. Traditional legal due diligence focused on existing rights, obligations and liabilities. The GDPR required buyers to test whether the target had appropriate internal systems, governance structures and compliance mechanisms to manage an ongoing regulatory risk. Its importance therefore lies less in its substantive rules than in the transactional model it created - and that model is the natural starting point for assessing the AI Act.

3. AI as an Object of Due Diligence, Not Only a Tool

AI is already transforming the M&A process itself. AI-powered tools are increasingly used to analyse large volumes of documents, identify material contingencies and inconsistencies, cross-reference information across data rooms and prioritise findings according to risk. Their use is not costless: inaccurate or hallucinated outputs, systemic bias and excessive reliance on automated analysis make human review and professional judgment essential.

More importantly for present purposes, AI is increasingly becoming an object of due diligence in its own right. At target level, the buyer must determine what AI systems the target uses, how they are governed, what data they rely upon, what regulatory classification applies and whether the required compliance framework has in fact been implemented.

4. Timing and Enforcement: Why the Risk Is Already Live

Although the AI Act applies generally from 2 August 2026, its provisions become applicable progressively. Article 113 provides that Article 6(1) and the corresponding obligations apply from 2 August 2027, while other provisions - including the penalties regime in Chapter XII – have applied since 2 August 2025. This phased application does not postpone the transactional relevance of AI risk. The classification and governance of systems already used by a target affect the assessment of its legal and commercial risk today, particularly where those systems fall, or may fall, within the high-risk framework.

The enforcement framework is substantial. Under Article 99(3), non-compliance with the prohibited AI practices in Article 5 may attract administrative fines of up to €35 million or 7% of total worldwide annual turnover for the preceding financial year, whichever is higher. Under Article 99(4), other infringements – including breaches of the obligations applicable to providers and deployers – may attract fines of up to €15 million or 3% of turnover. By comparison, Article 83(5) of the GDPR provides for a maximum of €20 million or 4% for the more serious categories of infringement. From an M&A perspective, the relevant question is therefore not only whether a particular obligation is already applicable, but whether the target's business model, technology infrastructure and compliance framework may expose the buyer to material regulatory liability following closing.

The ceilings are set by the Regulation, but enforcement is national. In Romania, the Government adopted a memorandum on 12 March 2026 designating ANCOM, the national communications regulator, as market surveillance authority and single national point of contact, operating alongside sectoral regulators, while the legislation meant to operationalise that institutional framework remains under discussion. Romania has not pursued substantive departures from the Regulation, so the exposure inherited on closing is the exposure the AI Act itself creates; what is still taking shape locally is the supervisory practice through which it will be enforced.

“

AI is increasingly becoming an object of due diligence in its own right.

5. Where the Exposure Concentrates: Fintech and Insurtech Targets

The exposure is most acute where AI systems produce legal or economic effects for individuals. Annex III, point 5(b) of the AI Act classifies as high-risk AI systems intended to evaluate the creditworthiness of natural persons or to establish their credit score, except where the system is used for the purpose of detecting financial fraud. Point 5(c) separately covers AI systems intended for risk assessment and pricing in relation to natural persons in the case of life and health insurance.

> “
>
> AI due diligence cannot be reduced to a conventional technology audit.

Article 27 reinforces the regulatory significance of these use cases. Deployers of the AI systems referred to in Annex III, points 5(b) and 5(c) - alongside public bodies and private operators providing public services - must carry out a fundamental rights impact assessment before the system is put into use, covering the categories of persons likely to be affected, the specific risks of harm to them, the human oversight measures in place and the steps to be taken should those risks materialise. For a fintech or insurtech target, whether such an assessment exists and has been implemented is a direct due diligence question.

A target whose core business depends on AI-assisted credit scoring, creditworthiness assessment, or insurance risk assessment and pricing therefore carries AI regulatory exposure at the heart of its business model. An inadequate AI compliance framework may expose an acquirer not only to regulatory enforcement and significant administrative fines, but also to contractual, operational and reputational liabilities arising after closing. The point is not confined to the financial sector: Annex III also covers employment, education, essential public services and law enforcement, so targets well outside fintech may carry comparable exposure through the tools they use internally.

6. The Anatomy of an AI Due Diligence Exercise

AI due diligence cannot be reduced to a conventional technology audit. It requires the buyer to reconstruct the target’s AI ecosystem, determine the regulatory status of its systems, identify the target’s role in relation to each of them and test whether the corresponding governance mechanisms are actually functioning. In practice, the exercise proceeds through seven connected enquiries.

Identification: AI may be embedded in internal decision-making tools, customer-facing products, third-party software, cloud services or automated processes without ever being recorded within the company’s compliance framework. The exercise must therefore begin with an inventory of the target’s AI systems and AI-enabled technologies, whether developed internally, acquired from third-party providers or incorporated into existing products and services.

Classification: Classification determines the intensity of the applicable obligations and, consequently, the level of legal and financial risk a buyer may inherit. Due diligence must establish how each system is classified, why it falls within that classification and whether the target has documented the basis for its own assessment.

Role in the AI value chain: The AI Act distinguishes between categories of actors, most notably providers and deployers, and the applicable obligations depend significantly on the role performed. A target may develop and place its own system on the market while simultaneously deploying third-party systems internally. A general assurance that the target complies with the AI Act is therefore of limited value. In our experience with Romanian transactions, this is the point most often mishandled. Many local targets are subsidiaries of foreign groups running AI systems developed centrally, so the deployer obligations sit with the Romanian entity while the provider documentation sits with a group company abroad and rarely reaches the data room without being specifically requested.

Evidence of compliance: For providers of high-risk systems, the review may address conformity assessment, technical documentation, quality management systems, registration, record-keeping and post-market monitoring. For deployers, the focus shifts towards appropriate use of the system, human oversight, input-data governance, monitoring, record-keeping and, where applicable, fundamental rights impact assessments. As the GDPR experience demonstrates, the existence of policies alone is not sufficient: transactional relevance lies in implementation and evidentiary basis.

Third-party dependencies: Where the target relies on external AI providers, the review should extend to the contractual framework supporting its AI infrastructure, including the allocation of regulatory responsibilities, audit and information rights, data usage, intellectual property, security, service continuity, liability limitations and termination rights. Concentration risk deserves particular attention, as dependence on a single vendor who cannot readily be replaced is both an operational and a transactional vulnerability.

Data governance: Where AI systems rely on training, validation or operational data, the buyer should assess the provenance, quality and legal basis for its use, particularly where personal data, confidential information or third-party intellectual property may be involved. The question is whether the target can demonstrate a defensible legal and governance framework for the data on which its material systems depend.

Regulatory history: Buyers should examine complaints, investigations, enforcement actions and other regulatory scrutiny relating to AI systems, together with material incidents or deficiencies identified internally. Previous findings may indicate both historical non-compliance and weaknesses in the broader governance framework, while the absence of disclosed enforcement action should not be equated with the absence of exposure.

7. From Findings to Deal Terms

The ultimate significance of this development is transactional. Once AI-related risks are identified through due diligence, they must be addressed through mechanisms capable of allocating them between buyer and seller. The findings may therefore shape AI-specific representations and warranties, dedicated disclosure schedules, targeted indemnities, pre-closing remediation obligations and, where appropriate, purchase price adjustments or escrow arrangements. Where remediation is required before closing, the parties should also address who bears its cost and on what timetable, since bringing a high-risk system into compliance may involve documentation and conformity work that cannot be completed quickly. As AI-related risks become increasingly relevant to a target's regulatory and operational profile, they may also affect the scope, pricing and availability of Warranty & Indemnity insurance, further integrating AI compliance into the risk allocation framework of the transaction.

> “
>
> It is also, and far less visibly, an M&A instrument.

8. Conclusion: An M&A Issue Hiding in a Product Regulation

That AI now demands legal attention is no longer a proposition that needs defending. What is less obvious is where that attention is owed. The AI Act is usually discussed as a product-compliance and governance regime, and therefore as a matter for engineering, compliance and regulatory functions. It is also, and far less visibly, an M&A instrument. Nothing in the Regulation is addressed to acquirers, yet its obligations attach to systems, to roles and to documentation that a buyer inherits in full on closing.

The consequences for transactional practice are practical rather than theoretical. Buyers should treat AI as a standing due diligence workstream rather than a subheading within the IT or intellectual property section; ask about AI systems even where the target does not describe itself as an AI business; confirm the target's role, provider or deployer, system by system, because the obligations differ materially; and test implementation rather than policy. Sellers should expect these questions and prepare an AI inventory, a classification rationale and a compliance file before the data room opens.

> “
>
> The risk in overlooking AI compliance is not that it will prove irrelevant, but that it will prove relevant later.

The GDPR took several years to become a standard heading in every due diligence report. AI is likely to take less. The transactions being negotiated today will close into a regime whose principal obligations bite in 2026 and 2027, and whose penalties already do. The risk in overlooking AI compliance is not that it will prove irrelevant, but that it will prove relevant later - after signing, when the cost of the omission falls on whoever failed to ask. On the Romanian market, where a substantial share of deal flow involves subsidiaries operating group-level or vendor-supplied technology, that cost is easily underestimated.

Lexters

A Romanian business law firm with an international outlook.

Based in Bucharest, we advise companies, investors, financial institutions and public-sector clients on complex transactions, regulatory matters and disputes.

Our lawyers combine strong Romanian legal expertise with international experience, including admission to the Bucharest and New York Bars. We bring together legal, commercial, regulatory and technological perspectives to help clients navigate complex challenges and make informed business decisions.

Our practice focuses on technology and regulated industries. We advise fintech, payment, digital asset and software businesses on market entry, product structuring, licensing, financing and compliance, while our corporate and disputes teams handle cross-border investments, strategic transactions, litigation and arbitration.

Our clients range from founders and high-growth companies to international groups, financial institutions, investors and public-sector entities, including Romanian state.

[law+] creatively reflects the different roles we take on for our clients: **lawyers. strategists. pillars. enablers. troubleshooters.**

Legal500 2026 EMEA

2026 IP STARS IP STAR

Chambers FinTech 2026

lexters.com

SAUDI ARABIA

ARTIFICIAL INTELLIGENCE IN SAUDI ARABIA'S FINANCIAL SECTOR: REGULATORY FRAMEWORK, GOVERNANCE AND EMERGING DEVELOPMENTS

STAT

Zeyad Y. AlSalloum
Partner

zalsalloum@statlawksa.com

+966 559 11 0779

www.statlawksa.com

BIO

Zeyad Y. AlSalloum is a Partner at STAT with over 15 years of experience advising private sector and government clients across diverse practice areas. He specializes in large, complex transactional and regulatory projects spanning energy, infrastructure, healthcare, transportation, and financial regulations. His corporate expertise encompasses joint ventures, commercial agreements, debt capital markets issuances, and corporate governance structuring. Zeyad's representative experience includes advising Saudi Aramco on major infrastructure and regulatory matters, and representing the Saudi Real Estate Refinancing Company on its Sukuk issuances and corporate governance. He has also advised prominent entities such as the Saudi Central Bank, Riyadh Airports Company, and the Royal Commission for AlUla. Zeyad holds an LLB (Hons) from the University of Westminster, UK, and is recognized for his strategic counsel on high-profile public-private partnerships and transformative national projects in the Kingdom of Saudi Arabia.

STAT

STAT

Richard Blackburn
Senior Counsel

rblackburn@statlawksa.com

+447 429 45 5656

www.statlawksa.com

BIO

Richard Blackburn is a Senior Counsel at STAT, bringing extensive expertise in conventional and Islamic banking and finance. A qualified lawyer in England & Wales, Richard advises international and regional banks, funds, and corporates on a broad spectrum of matters, including Sukuk issuances, derivatives, finance-linked hedging, structured note programmes, margin loans, and securitisation. His distinguished career includes practice in London at leading global firms such as White & Case LLP, Norton Rose Fulbright LLP, and Simmons & Simmons LLP, alongside secondments with major investment banks. Legal 500 has recognized Richard for his significant experience advising on OTC derivatives and structured products, noting him as “responsive and helpful.” His recent representative work includes acting as arranger for substantial Sukuk issuance programmes and advising global financial institutions on complex hedging, credit-linked notes, and major infrastructure financing transactions.

STAT

Abdullatif Alharbi
Associate

aalharbi@statlawksa.com

+966 544 62 1234

www.statlawksa.com

BIO

Abdullatif Alharbi is an Associate at STAT specializing in banking law and debt capital markets. A licensed lawyer in Saudi Arabia admitted by the Ministry of Justice and a member of the Saudi Arabian Bar Association, Abdullatif brings specialized regulatory expertise to the firm. He previously served at the Saudi Central Bank (SAMA), where he provided strategic advice on critical regulatory matters within the banking sector. Prior to his tenure at SAMA, he advised diverse corporate and governmental clients on a wide range of complex mandates, including corporate and real estate finance, investment-grade lending, and Sukuk issuances. Abdullatif holds an LL.M. from Boston University, USA, and an LL.B. from Qassim University, Saudi Arabia. His unique background, combining central bank experience with private practice, allows him to provide nuanced counsel on the evolving financial regulatory landscape in the Kingdom.

ARTIFICIAL INTELLIGENCE IN SAUDI ARABIA'S FINANCIAL SECTOR: REGULATORY FRAMEWORK, GOVERNANCE AND EMERGING DEVELOPMENTS

1. Introduction

The Kingdom of Saudi Arabia has positioned artificial intelligence (AI) as a cornerstone of its national economic transformation. In a signal of the highest level of political commitment, the Council of Ministers designated 2026 as the 'Year of Artificial Intelligence', an acknowledgment that AI is foundational infrastructure to be built, regulated and leveraged across every sector of the economy, with financial services at the centre of that ambition.[1]

> "Saudi Arabia has positioned artificial intelligence as a cornerstone of its national economic transformation.

From national banks deploying real-time fraud detection systems to fintech startups offering AI-driven credit scoring and from robo-advisory platforms managing investment portfolios to insurtech firms pricing risk through machine learning, the pace and breadth of AI adoption in Saudi financial services is significant. At the same time, a regulatory framework is beginning to take shape, one that seeks to enable innovation while protecting financial system stability, consumer rights and national data.

Saudi Arabia does not currently regulate financial-sector AI through a single, comprehensive statute. Instead, the framework combines laws and guidance administered by the Saudi Data and AI Authority (SDAIA), the National Cybersecurity Authority (NCA) and the Communications, Space & Technology Commission (CST) with sectoral requirements imposed by the Capital Market Authority (CMA), the Saudi Central Bank (SAMA) and the Insurance Authority. The legal status of these instruments differs: some are binding laws or sector rules; others are official principles, frameworks or consultation drafts. Financial institutions should identify the source and status of each requirement rather than treating the framework as a single AI rulebook.

This article maps the regulatory framework across its three principal pillars, the national framework under SDAIA, the capital markets framework of the CMA and the banking, payments and insurance frameworks of SAMA and the Insurance Authority, and examines the principal AI use cases, governance obligations, emerging risks and key developments that financial institutions operating in the Kingdom must understand.

2. National Strategy and Policy Framework

Saudi Arabia's approach is shaped by Vision 2030, the National Strategy for Data and Artificial Intelligence (NSDAI) and the Financial Sector Development Program (FSDP). Although these initiatives are not themselves regulatory instruments, they establish the policy objectives that underpin the Kingdom's AI governance agenda, including the adoption of AI across the economy.

2.1 Vision 2030 and the NSDAI

Saudi Arabia's AI ambitions are inseparable from Vision 2030, the Kingdom's programme for economic diversification. At its heart sits the National Strategy for Data and Artificial Intelligence (NSDAI), which articulates the Kingdom's goal of becoming a global leader in AI by 2030 and identifies priority sectors for AI deployment, including education, government, healthcare, energy and transportation. Although financial services is not identified as a standalone priority sector, the NSDAI provides the strategic foundation for the development of AI across the financial system through complementary initiatives led by SAMA, the CMA and the Financial Sector Development Program.[2]

The designation of 2026 as the "Year of Artificial Intelligence" further illustrates the Government's commitment to accelerating AI adoption across both the public and private sectors.[3]

1. Council of Ministers Decision No. 720 dated 21 Ramadan 1447H (10 March 2026), published in Umm Al-Qura No. 5150 (19 March 2026).
2. SDAIA, National Strategy for Data and Artificial Intelligence (NSDAI, sdaia.gov.sa)
3. Council of Ministers Decision No. 720 dated 21 Ramadan 1447H (10 March 2026), published in Umm Al-Qura No. 5150 (19 March 2026).

2.2 The Financial Sector Development Program (FSDP)

Launched in 2018 by the Council of Economic and Development Affairs (CEDA) as one of Vision 2030's Vision Realization Programs, the Financial Sector Development Program (FSDP) is the primary vehicle through which Saudi Arabia is building a diversified, effective and digitally-enabled financial sector. It operates through four strategic objectives: enabling financial institutions to support private sector growth; developing an advanced capital market; promoting and enabling financial planning (savings and retirement); and enabling the FinTech strategy.[4]

Although the FSDP is not an AI regulatory framework, many of its initiatives (managed in partnership with SAMA, the CMA and the Insurance Authority) provide the practical foundations upon which AI-enabled financial services depend. The licensing of digital banks, implementation of the Kingdom's FinTech strategy, expansion of electronic know-your-customer (eKYC) processes and rapid growth in digital payments have created the digital infrastructure and data environment necessary for AI applications such as fraud detection, credit scoring, automated investment advice and anti-money laundering systems.

The FSDP's 2024 Annual Report recorded 261 active fintech entities at year-end and electronic payments representing 79% of retail transactions during 2024. Its Q4 2024 newsletter reported that 98.88% of investment accounts were opened through eKYC.[5] More recently, the CMA reported that assets under management through fintech platforms increased by 87%, from SAR 3.43 billion in Q4 2024 to SAR 6.41 billion in Q4 2025. These developments demonstrate the scale and continuing growth of the digital infrastructure on which AI-enabled financial services can be deployed.[6]

3. The Regulatory Framework: A Hybrid Model

The current framework has two principal layers. At national level, SDAIA's rules and guidance address data protection and AI ethics, the NCA's requirements address cybersecurity, and CST rules govern cloud, data-centre and ICT infrastructure. Sectoral regulators such as CMA and SAMA apply licensing, governance, conduct, technology, outsourcing and operational-risk requirements to AI-enabled activities within their respective remits. The application of a particular rule therefore depends on the regulated entity, the activity, the data used and the way the technology is hosted or supplied.

Financial institutions deploying AI should also consider the National Cybersecurity Authority's (NCA) cybersecurity requirements and Communications, Space & Technology Commission (CST) rules governing cloud, data centre and ICT infrastructure, particularly where AI systems are hosted or accessed through third-party cloud environments.

In practical terms, fraud detection engages SAMA's counter-fraud and cybersecurity requirements, the NCA controls and the PDPL; credit scoring engages SAMA's conduct and risk-management rules together with the PDPL and, where triggered, a DPIA; robo-advisory is governed principally by the CMA rules discussed below; and insurtech falls within the Insurance Authority's remit together with applicable data, cybersecurity and outsourcing requirements.

3.1 Pillar One: SDAIA — The National Framework for AI Governance and Data Protection

> “
>
> SDAIA is the apex national authority for data and AI policy.

Established by Royal Order in 2019, SDAIA is the apex national authority for data and AI policy. SDAIA's published AI ethics principles and related guidance set out recommended governance standards for all institutions in the Kingdom, including financial services firms that design, develop, deploy or use AI systems. Version 2.0 of the AI Ethics Principles comprises seven principles. The following are the three principles most relevant to financial institutions:[7]

4. FSDP Programme Charter 2020–2025, vision2030.gov.sa
5. Financial Sector Development Program, Annual Report 2024; FSDP Quarterly Newsletter, Q4 2024.
6. Capital Market Authority, 'The Capital Market Authority Approves the Regulation of Robo-Advisory Services' (5 March 2026).
7. SDAIA, AI Ethics Principles, version 2.0 (September 2023).

- **Fairness and Inclusivity:** Institutions should take steps to prevent bias and discrimination in the datasets on which predictive AI models are built, particularly in credit scoring and insurance pricing;
- **Transparency and Accountability:** AI systems should be explainable and interpretable, with mechanisms that allow individuals to trace and contest AI-driven decisions affecting their interests; and
- **Pre-Deployment Assessment:** Institutions should assess bias and privacy risks before deploying predictive AI models. A DPIA is mandatory only where the trigger under the PDPL Implementing Regulations is met.

> “
>
> The CMA has moved earliest among the sectoral regulators to address AI-enabled financial services.

SDAIA's Generative Artificial Intelligence Guidelines for the Public (January 2024) recommend, in section 5.1, that providers supplying intensive cloud compute, particularly GPU or TPU capacity, use identity-verification processes similar to those used by financial institutions. This is non-binding guidance rather than a mandatory KYC obligation.[8]

SDAIA's National Artificial Intelligence Risk Management Framework (SDAIA-P145, version 1.0, April 2026) provides a methodology for identifying, assessing, mitigating, monitoring and reviewing AI risks across public- and private-sector organisations. It addresses risks distinct from conventional software, including emergent behaviour, model drift, limited explainability and difficulty reproducing outputs. The Framework is advisory and expressly creates no new obligations, but provides a national benchmark that may inform risk-management practices in regulated sectors.[9]

The principles are an important governance framework of universal application, but their status and application should be distinguished from obligations arising under the Personal Data Protection Law (PDPL) which applies where an AI system processes personal data and binding sector rules. Relevant requirements include establishing a lawful basis for data processing, complying with purpose limitations and data minimisation, protecting data-subject rights, implementing appropriate security and conducting a data protection impact assessment where the applicable conditions are met.[10]

3.2 Pillar Two: The Capital Market Authority (CMA)

The CMA has moved earliest among the sectoral regulators to address a specific AI-enabled financial service. It does not currently regulate AI through a dedicated framework, but through existing capital-markets legislation and the targeted robo-advisory amendments described below.

Robo-Advisor Regulations

In March 2026, the CMA approved amendments governing robo-advisory services. Following implementation of such amendments, the Capital Market Institutions Regulations now permit licensed investment management firms to offer robo-advisory services, defined as the use of algorithms to manage client investments with no or limited human intervention. Key conditions include:[11]

- **Advance notification:** Portfolio-construction and management strategies, and material updates, must be notified to the CMA in advance, before they are made available to clients;
- **Algorithm integrity:** A Capital Market Institution offering robo-advisory services must establish systems and control procedures to ensure the integrity and efficiency of the algorithms and technological tools used. It must test their reliability and effectiveness periodically, with such testing conducted at least ten days before the service is offered to clients on the platform;
- **Portfolio and foreign securities:** Portfolios must not be concentrated in one asset or issuer and foreign securities must be supervised in a jurisdiction with standards and requirements at least equivalent to the CMA's;

8. SDAIA, Generative Artificial Intelligence Guidelines for the Public, version 1.0 (January 2024), section 5.1, p. 11.
9. SDAIA, National Artificial Intelligence Risk Management Framework, SDAIA-P145, version 1.0 (April 2026).
10. PDPL, Royal Decree M/19, 2021; SDAIA enforcement activity 2025
11. CMA, 'CMA Approves Regulation of Robo-Advisory Services' (5 March 2026); Capital Market Institutions Regulations, Annex 3.2, paragraph 8, and articles 19(b)(8), 20(b)(5) and 55(e).

- **Disclosure requirements:** A Capital Market Institution that offers robo-advisory services must disclose the performance record of the investment portfolios since inception. This must include the criteria and methodology used to measure portfolio performance, the total returns achieved after deducting actual expenses, and any updates thereto.
- **Information Technology Officer:** A registered Information Technology Officer must be maintained.

The regulations therefore treat AI as a means of providing discretionary investment management while imposing additional controls tailored to algorithmic delivery. The CMA reported that assets under management on fintech platforms rose by 87% to SAR 6.41 billion in Q4 2025, from SAR 3.43 billion a year earlier[12]

FinTech ExPermit Sandbox

The CMA's FinTech ExPermit Instructions establish a regulatory sandbox through which FinTech firms may test innovative financial technology products within a controlled "sandbox" environment before seeking full CMA authorisation. The sandbox allows the CMA to supervise new technologies in a controlled environment, understand their potential risks and consider whether changes to the regulatory framework are required.

Firms remain responsible for regulatory compliance and subject to the instructions and any conditions imposed by the CMA throughout the testing period. They may not exclude or limit their liability to clients for losses arising from the products being tested or agree to share in clients' investment losses. The sandbox is therefore intended to encourage innovation without compromising investor protection.[13]

3.3 Pillar Three: The Saudi Central Bank (SAMA) and the Insurance Authority

SAMA does not currently regulate AI through a dedicated AI rulebook. Instead, AI systems used by SAMA-regulated institutions fall within generally applicable technology, cybersecurity, counter-fraud, outsourcing, cloud-computing and operational-risk requirements.

IT Governance Framework

SAMA's Information Technology Governance Framework targets Level 3 maturity and requires periodic self-assessments. These should be reviewed by internal audit and reported to SAMA, which may also review or audit them. Although the Framework does not address AI in terms, it extends by implication to the technology infrastructure and governance supporting AI systems.[14]

Cybersecurity Framework (CSF)

The SAMA Cyber Security Framework does not expressly address AI. It applies generally to regulated institutions' information technology assets, including business applications and infrastructure, as well as outsourced services and new technologies. Accordingly, AI systems forming part of, or supporting, those information assets are subject to the Framework's relevant cybersecurity governance, risk-management, monitoring and third-party controls, although the Framework imposes no AI-specific requirements.[15]

“

SAMA does not currently regulate AI through a dedicated AI rulebook.

Cross-Border Data, Cloud and Outsourcing

For a foreign-hosted AI model, the first question is whether the proposed data flow and hosting arrangement are permissible. If personal data will leave the Kingdom, the controller must establish a lawful route under the PDPL and the Regulation on Personal Data Transfer Outside the Kingdom, including any required safeguard, data minimisation, security and transfer risk assessment. The analysis should cover prompts, retrieved documents, logs, embeddings, support access and onward transfers, rather than only the contractual hosting location.

12. CMA, 'CMA Approves Regulation of Robo-Advisory Services' (5 March 2026).
13. CMA FinTech ExPermit Instructions
14. SAMA, Information Technology Governance Framework, including section 2.3 (Self-Assessment, Review and Audit).
15. SAMA Cybersecurity Framework (CSF)

That data-law analysis is separate from SAMA's prudential gate. SAMA's cloud requirements state a principle of in-Kingdom hosting and require approval before using a cloud service or signing the contract; explicit approval is required for services outside the Kingdom. Material outsourcing also requires SAMA's prior non-objection under the applicable outsourcing rules. A foreign-hosted model is therefore not automatically prohibited, but should not be deployed until both the PDPL transfer route and the relevant SAMA approval or non-objection route are established. SAMA's Open Banking Framework is also an AI data enabler by permitting consented, standardised access to customer-permissioned financial data, subject to its business rules and technical standards.[16]

> “
>
> AI systems introduce attack surfaces that traditional cybersecurity frameworks were not designed to address.

Regulatory Sandbox

SAMA's Regulatory Sandbox permits eligible financial innovations, including AI-enabled services, to be tested under defined conditions before wider deployment. It provides a supervised route to address licensing questions, gather evidence and evaluate risks.[17]

On 28 June 2026, SAMA launched an enhanced Regulatory Sandbox Service through its e-services portal to automate applications and status tracking. This was an administrative portal upgrade rather than a new regulatory rule.[18]

Insurance Authority and InsurTech Rules

The Insurance Authority has been the standalone regulator responsible for supervising and regulating Saudi Arabia's insurance sector since 23 November 2023. Legacy insurance regulations and instructions remain in force until replaced or updated; the InsurTech Rules are now published within the Authority's own regulatory materials and its rulemaking pipeline should be monitored separately from SAMA's banking and payments framework.[19]

4. Emerging Risks

AI-Enabled Fraud and Deepfakes

Generative AI can facilitate impersonation, synthetic identities, social engineering and fabricated documents or instructions. SDAIA's Deepfakes Guidelines describe the ethical, privacy and transparency risks associated with manipulated media.[20] In addition, sectoral regulators' counter-fraud and cybersecurity frameworks provide more general requirements to prevent, detect and respond to fraud and technology risk.[21]

Algorithmic Bias

Bias in AI training data can produce systematically unfair outcomes in credit scoring and insurance pricing. SDAIA's non-binding instruments indicate that institutions should assess and mitigate bias before deployment and monitor outcomes after deployment. A mandatory assessment arises only where a binding legal trigger applies, including a DPIA where required under the PDPL Implementing Regulations.

Cybersecurity and Regulatory Fragmentation

AI systems introduce attack surfaces, adversarial inputs, model poisoning and data leakages that traditional cybersecurity frameworks were not designed to address. The NCA's Essential Cybersecurity Controls and SAMA's CSF provide the primary regulatory response. In addition, on 5 July 2026, the NCA opened consultation on draft AI Cybersecurity Guidelines covering governance, defence, resilience and third-party cybersecurity. The draft expressly addresses generative and agentic AI. The consultation closed on 5 August 2026; the Guidelines remain a draft pending final issuance and should not be described as binding regulation.[22]

Simultaneously, the multi-body regulatory architecture spanning SDAIA, SAMA, CMA and NCA creates compliance complexity that institutions must invest in managing proactively as the framework develops.

16. Regulation on Personal Data Transfer Outside the Kingdom, including articles 2, 4 and 7; SAMA Cybersecurity Framework, section 3.4.3 (Cloud Computing); applicable SAMA outsourcing rules; SAMA Open Banking Framework.
17. SAMA Regulatory Sandbox Framework; FSDP Programme Charter 2020–2025
18. SAMA, 'SAMA Launches Enhanced Regulatory Sandbox Service' (28 June 2026).
19. Saudi Central Bank (SAMA), InsurTech Rules (approved 30 July 2023), now administered by the Insurance Authority.
20. SDAIA, Deepfakes Guidelines.
21. Saudi Central Bank, *Counter-Fraud Framework*.
22. National Cybersecurity Authority, 'Public Consultation on AI Cybersecurity Guidelines' (5 July 2026); Draft AI Cybersecurity Guidelines.

5. Looking Ahead

Several developments warrant close attention from financial institutions operating in or entering the Kingdom:

- **Draft Responsible AI Policy:** SDAIA consulted through Istitlaa from 3 April to 3 May 2026 on a policy applying across the public, private and non-profit sectors. The draft classifies AI systems into four risk categories—critical, high, limited and low—and proposes tiered requirements including registration of certain systems, safety testing and monitoring, audit and assurance requirements for high-risk systems, and watermarking and content-tracking measures. The consultation has now closed and final issuance of the Responsible AI Policy is awaited.[23]

- **Draft AI-First Policy:** SDAIA consulted through Istitlaa from 29 June to 28 July 2026 on a draft policy intended to embed a structured and responsible 'AI-first' approach across public-sector entities. The consultation has closed and final issuance is awaited.[24]

- **Global AI Hub Law:** CST published a draft framework for sovereign data centres, data embassies and cross-border hosting for consultation on 14 April 2025. The consultation closed on 14 May 2025, and the law had not been enacted at the time of writing. The proposal could become relevant to international data and AI infrastructure arrangements and, once enacted, would reshape cross-border AI deployment strategies and create new data hosting and infrastructure opportunities for international financial institutions. Note however it is not a general AI-governance statute.[25]

- **PDPL Compliance:** The PDPL is fully operational following the end of its statutory grace period on 14 September 2024. Financial institutions should review key compliance areas e.g. whether AI data pipelines have a documented lawful basis and purpose, use proportionate data, support data-subject rights, trigger a data protection impact assessment where required and comply with breach-notification and record-keeping requirements.[26]

- **SAMA AI guidance:** More targeted SAMA guidance on model risk management, algorithmic credit underwriting and AI in AML and KYC processes is anticipated.

- **FSDP targets and FinTech expansion:** The FSDP's ongoing targets (for example, the number of fintech companies, SME lending and stock-market capitalization) will continue to drive digitalisation and create expanded opportunities for AI-driven financial services.

- **Arabic-language AI:** Saudi Arabia's launch of HUMAIN, a PIF-owned AI company focused on AI infrastructure, cloud capability, data centres and Arabic-language models, reflects the Kingdom's broader sovereign AI strategy. Although not itself a regulatory development, HUMAIN may influence the AI tools available to financial institutions, particularly for Arabic-language customer engagement, compliance support and locally hosted AI services.

> "Financial institutions that invest now in robust AI governance will be best positioned to harness AI opportunities.

Financial institutions that invest now in robust AI governance frameworks, encompassing board accountability, pre-deployment assessments, data governance, cybersecurity controls and vendor management, will be best positioned to navigate this evolving landscape and to harness the substantial opportunities that AI presents in one of the world's most dynamic financial markets.

23. https://www.spa.gov.sa/en/N2551533
24. SDAIA, Draft AI-First Policy consultation, Istitlaa (29 June-28 July 2026), https://istitlaa.ncc.gov.sa/ar/transportation/ndmo/aifirstpolicy/Pages/default.aspx.
25. Communications, Space & Technology Commission, 'CST Publishes a Public Consultation for the Global AI Hub Law' (14 April 2025; consultation closed 14 May 2025).
26. SDAIA, https://dgp.sdaia.gov.sa/wps/wcm/connect/f579bc32-fda8-47bd-bc6f-66b8cb77985c/ENG-Guide%2Bto%2Bthe%2Bsaudi%2BPDP%2Blaw%2Bfor%2Bcontrollersprocessors.pdf?MOD=AJPERES

Navigating the Future of AI Regulation in Financial Services

Strategic legal counsel at the intersection of artificial intelligence, regulation, and finance in the Kingdom of Saudi Arabia.

Our Expertise

Banking & Finance | Capital Markets | Corporate Governance
Regulatory Advisory | Technology & AI | Islamic Finance
Energy & Infrastructure | Public-Private Partnerships

"A hands-on team who work around the clock and have outstanding knowledge of Saudi law."
Chambers & Partners

SOUTH AFRICA

WHO AUTHORISED THAT PAYMENT?

AGENTIC COMMERCE AND SOUTH AFRICA'S PAYMENT SYSTEM

BAKER MCKENZIE

Ashlin Perumall
Partner

ashlin.perumall@bakermckenzie.com

+27 11 911 4431

www.bakermckenzie.com

BIO

Ashlin Perumall is a partner and Head of the IPTech practice at Baker McKenzie in Johannesburg, dual-qualified in South Africa and England and Wales. His practice centres on transactions, regulation and strategy at the intersection of financial services and emerging technology. He acts as key advisor to clients entering or acquiring in African fintech, including paytech, open banking, digital banking and APIs, and advises on responsible AI governance, AI procurement, crypto-asset regulation, and blockchain and distributed ledger technology. He has particular experience establishing legal, compliance and diligence frameworks for novel, technically complex products and businesses. Ashlin is a Fellow of the World Economic Forum's Centre for the Fourth Industrial Revolution, where he conducted regulatory and policy research in San Francisco, and has also worked in the firm's London office. He writes and speaks regularly on law, AI and technology. He is ranked Band 1 for FinTech by Chambers Global and holds an LLM in Innovation, Technology and the Law from the University of Edinburgh.

Baker McKenzie.

WHO AUTHORISED THAT PAYMENT?
AGENTIC COMMERCE AND SOUTH AFRICA'S PAYMENT SYSTEM

Delegating standing payment authority to autonomous AI agents exposes a fundamental gap in how liability is attributed within traditional banking frameworks. Consider a procurement platform where an AI agent automatically settles an invoice. On paper, one could argue that the process is flawless: once the ordinary credentials check out, the parameters are essentially met, and the payment rail can clear the funds. But what happens if the invoice itself is fabricated? For example, where a compromised email from a supplier contained hidden text instructing any AI reading it to reroute the cash to a new account, and the agent simply processed the text and complied. Right now, in South Africa and many other countries, the result is a complete deadlock. The customer will argue that they never approved the transfer, while the bank maintains their security was never breached. The core problem is that, from both a technical and legal standpoint, both positions are presently entirely defensible.

> “
>
> The core challenge in agentic commerce is fundamentally a question of agency law.

That dispute is one of the core legal problems existing today for the coming wave of ‘agentic commerce’, i.e. systems where software agents initiate and complete payments on behalf of human users. The current legal framework governing these transactions rests on a core assumption that a payment instruction represents the actual intent of its issuer. Agentic AI processes break that assumption. Agents can be manipulated into executing genuine instructions their human principals never intended, and further, deepfake technology complicates the identity verification processes relied upon to resolve such disputes. As a result, authorisation and authentication are colliding, and South Africa's existing statutory framework, payment rail architecture and current modernisation programme each offer part of the answer.

Baker McKenzie.

The shift toward machine-initiated commerce

The agentic infrastructure supporting this shift is already taking shape, and has arrived much faster than many in the industry expected. Throughout 2025, major payment networks launched protocols explicitly built for agentic commerce. One global network introduced an ‘agent pay’ solution using tokens to bind credentials to specific AI agents, while a key competitor rolled out a similar trusted agent protocol. Technology providers moved in tandem, with leading AI developers releasing frameworks that anchor transactions in cryptographic proof of both the user's request and the agent's proposed action.

The terminology emerging from these developments is worth noting. Global standards for machine-initiated payments have overwhelmingly adopted the term “mandate”, seen directly in a prominent agent payments protocol that structures every transaction around a signed “Intent Mandate”, “Cart Mandate” and “Payment Mandate”. Under South African law, a mandate is the precise mechanism used to authorise one party to act on behalf of another. By adopting this language, the technology sector has arguably conceded that the core challenge in agentic commerce is fundamentally a question of agency law.

South Africa's financial sector is already laying the groundwork for this transition. A November 2025 joint report from the FSCA and the PA reported that, as at its 2024 industry survey, 52 per cent of banks and 50 per cent of payment providers had AI in production. While current applications are concentrated in areas such as fraud detection, operations and customer service, the leap to transacting agents is likely imminent.

Baker McKenzie.

Three groupings of answers are emerging globally

Jurisdictions confronting the same question have produced three broad groupings of responses, and each is instructive for emerging markets such as South Africa.

The first grouping engineers the mandate directly into the payment rail, and India is furthest along this path. The National Payments Corporation of India introduced delegated payments on the Unified Payments Interface in 2024 through UPI Circle, allowing an account holder to grant controlled payment access to a secondary user within defined limits. An October 2025 circular extended that delegation to software and connected devices, creating a framework for payments executed by AI agents on the national rail. According to press reports, the corporation is developing a Unified Agent Protocol, a proposed national registry through which AI agents would be registered, verified and authorised to transact; the protocol reportedly requires Reserve Bank of India approval before launch. India's technology ministry has separately proposed mandatory human-in-the-loop interventions for defined categories of agentic payments. Brazil is adopting a similar approach. Pix Automático, live since June 2025 and mandatory for every payer-side institution participating in the Pix instant payment system, lets a payer grant a single in-app authorisation against which the central bank's rail validates all future recurring collections.

The second grouping accepts that some deceived payments will clear and reallocates the loss by regulation. The United Kingdom's mandatory reimbursement regime for authorised push payment fraud, in force since October 2024, requires payment service providers to refund defrauded customers up to GBP 85,000, with the cost split equally between sending and receiving institutions. In July 2026, the Payment Systems Regulator published the first full independent evaluation of the regime, conducted by Frontier Economics. The evaluation estimated that annual APP fraud losses over Faster Payments were approximately GBP 73 million lower than they would otherwise have been, a reduction of roughly 21 per cent, with the share of losses reimbursed rising from 54 per cent to 65 per cent. Frontier further found that more than 99 per cent of reimbursable claims and more than 95 per cent of reimbursable value fall below the GBP 85,000 cap; the cap, the 50/50 cost split and the consumer-caution exception are all under review, with findings due during 2026/27. Similarly, Singapore's Shared Responsibility Framework (effective December 2024) distributes phishing scam losses along a waterfall (based on duties breached): financial institutions bear the first tranche, then telecommunications providers, before any loss rests with the consumer. The European Union's Payment Services Regulation, whose final text was agreed in April 2026 and which is expected to apply roughly 18 to 21 months after publication in the Official Journal, goes a step beyond: it will grant an uncapped refund right where a fraudster impersonates the customer's own provider, conditional on the customer reporting the matter to the police and notifying the provider. The regulation will also extend payee-name verification beyond the SEPA Instant rail, where it has been mandatory since October 2025 under the Instant Payments Regulation, to credit transfers more broadly. These regimes matter for agentic commerce because a prompt-injected payment is, in substance, the machine-executed version of the deceived authorised payment they were built to address. The UK data indicates that deliberate loss allocation changes behaviour rather than simply redistributing cost across the system, with fraud losses falling measurably. The EU is also constructing an identity layer: every member state must offer citizens a European Digital Identity Wallet by the end of 2026, built on verifiable credentials of the same kind agentic payment protocols now demand.

> “
>
> A prompt-injected payment is the machine-executed version of a deceived authorised payment.

The third grouping modernises the underlying contract law. The UNCITRAL Model Law on Automated Contracting, adopted in July 2024, was drafted for the same gap. It establishes updated rules attributing the outputs of automated systems (including AI systems) to the party using them, together with an optional rule addressing unexpected outcomes; it is designed to supplement first-generation electronic transactions statutes of the kind South Africa enacted in 2002. No African state is known to have enacted it. The first to do so will, as a result, have the continent's most current legal foundation for machine-initiated commerce.

> “
>
> An AI agent manipulated by injected text has not malfunctioned; processing contextual instructions is its intended function.

A statutory framework built for early automation

South Africa addressed some of the legal questions of automated commerce relatively early. The Electronic Communications and Transactions Act 25 of 2002 (**ECTA**) validates agreements formed by electronic agents. Section 20, which in retrospect was far ahead of its time, presumes a party using an electronic agent is bound by the resulting terms, regardless of whether a human reviewed the action. Section 25 attributes a data message to its originator if sent by an automated information system, unless the system failed to execute its programming properly.

Beneath ECTA lies the common law of mandate, which dictates the scope and limits of authority. Above it sits the National Payment System Act 78 of 1998 and the SARB directives governing third-party payment providers, much of which is undergoing significant overhaul in 2026.

On paper, this hierarchy appears coherent. Every layer, however, was drafted for deterministic automation: systems that execute pre-set instructions in rigid ways. The drafters of ECTA envisioned two outcomes: the system works as programmed (binding the principal) or it malfunctions (releasing them). However, a probabilistic large language model agent introduces a third. It interprets context, takes instructions from the unstructured data it reads, and may act independently in goal-driven systems rather than acting on deterministic choices by human principals.

Credentials susceptible to manipulation

The OWASP 2026 Top 10 for LLM Applications, published on 4 August 2026, ranks prompt injection as the primary security risk for large language model applications, a position it has held since the list's inception but now with expanded scope covering cross-modal attacks, memory persistence and agentic blast radius. The same edition elevated Excessive Agency from sixth to third place, driven by real-world agentic deployments. The underlying structural vulnerability is that these models process instructions and data through the same channel. Text an agent reads from an invoice or email can therefore function as an executable command if structured in certain ways and if guardrails don't catch it. For an agent holding payment credentials, the possible result is a fully authorised outbound transfer, which would be difficult to distinguish from a validly authorised one, given that no system intrusion or stolen password is required.

Applying this scenario to existing legal frameworks exposes several gaps. Standard banking practice for unauthorised transactions relies on evidence of compromise, such as a stolen card or a hijacked session. In an agentic commerce scenario, the agent's own credentials sign the instruction. Under the law of mandate, the agent technically operates within its express authority to pay approved suppliers. ECTA, and similarly structured legislation in other countries, provides no clear remedy either. An AI agent manipulated by injected text has not malfunctioned; processing contextual instructions is its intended function. The defect lies in the external data it read, a scenario the statute does not contemplate. Loss allocation will therefore depend on improvised readings of contract terms and negligence principles designed for human agents. The UK, Singaporean and EU regimes described above show what a deliberate allocation looks like, and South Africa currently has no equivalent default.

Baker McKenzie.

Regulatory recognition of the risk

South Africa's financial regulators have identified these vulnerabilities. In particular, the FSCA and PA explicitly ranked cybersecurity among the top risks of AI adoption. Their reporting details specific mechanisms of concern: data poisoning, privacy exposures under POPIA, and the systemic risk created when AI capabilities concentrate among a few third-party vendors. The Financial Stability Board, an international body that strongly influences South Africa's financial regulatory framework, identified third-party dependency and cyber vulnerability as central supervisory issues for the sector in a June 2026 consultation paper on responsible AI adoption, with practices 11 and 12 addressing AI-related cyber, ICT and third-party risks specifically. The consultation closed on 22 July 2026, and the final report is due in October 2026. Notably, the FSB's October 2025 monitoring report on AI in financial services was produced at the request of the South African G20 Presidency, underscoring the country's direct stake in shaping the international supervisory consensus on these issues.

Supervisory standards are evolving accordingly. Joint Standard 1 of 2023 on IT governance and Joint Standard 2 of 2024 on cybersecurity and cyber resilience compel financial institutions to maintain strict access controls and report material cyber incidents. A prompt-injection compromise of a transacting agent would ordinarily meet the materiality threshold for notification, which runs to 24 hours under Joint Standard 2.

Applying the precedent of DebiCheck

South Africa has managed systemic payment disputes at scale before. Following widespread debit order abuse in the previous decade, the SARB mandated the DebiCheck system. DebiCheck shifted mandate authentication from the endpoint to the national payment rail itself. Under this model, the delegation of authority becomes an independently verifiable digital asset. Disputes were resolved against the registered mandate rather than through reconstructing intent.

International agentic payment systems are converging on a very similar architecture, using scoped tokens and verifiable agent identities. UPI Circle and Pix Automático are, in structural terms, members of the design grouping that DebiCheck pioneered for humbler technology. One could therefore argue that South Africa already has the institutional knowledge to implement mandate authentication at a national scale.

The African dimension

African payments are wallet-first and instant-payment-first, and therefore agentic AI will likely meet most African consumers inside mobile money platforms and national instant payment systems, not at card checkouts. The design decisions made by African rail operators will therefore matter more here than in card-dominated markets.

“

DebiCheck shifted mandate authentication from the endpoint to the national payment rail itself.

Nigeria illustrates the infrastructure trajectory clearly. The country issued Africa's first open banking regulatory framework in 2021. Subsequently, the Central Bank of Nigeria committed to a phased commercial rollout in 2026, anchored on a central registry of verified participants operated by the Nigeria Inter-Bank Settlement System and a consent management framework tied to the Bank Verification Number. A national registry of authorised participants, joined to identity-anchored consent, is agent-registration architecture by another name; extending it to software agents is an increment, echoing the agent registry India is now designing. Kenya illustrates the governance trajectory: its National AI Strategy 2025-2030, launched in March 2025, is among the continent's most developed, and a dedicated AI Bill is progressing. Both operate under the African Union's Continental AI Strategy of 2024. At least sixteen of Africa's fifty-four states have adopted national AI strategies to date, leaving a wide governance gap into which agentic commerce will arrive regardless.

Addressing this both locally and across African countries will be an important challenge, as ever-growing cross-border rails raise the stakes. The Pan-African Payment and Settlement System now connects twenty-eight countries and more than 190 commercial banks and fintechs, with the Bank of Central African States joining in July 2026. Once an agent's payment instruction can cross a border in local currency, mandate authentication becomes a continental interoperability question: a delegation registered in one market must be verifiable in another. That is a standards problem the African Continental Free Trade Area's digital trade agenda exists to solve, and one it can hopefully address. It is also an opening for South Africa: a mandate layer proven domestically could be exported through the same regional infrastructure now being joined together.

> “
>
> Disputes over machine-initiated payments belong at the infrastructure level rather than in a courtroom.

Modernising the national payment rail

As a unique convergence of timing, these new developments come at a time when South African payments laws are undergoing unprecedented change. The SARB is advancing its Payments Ecosystem Modernisation programme, the most extensive intervention since the 1998 Act. With the development of the Vision 2030+ strategic framework and the PEMKey credential system, a new credential layer for the national rail is being designed now, and may be able to take advantage of the timing.

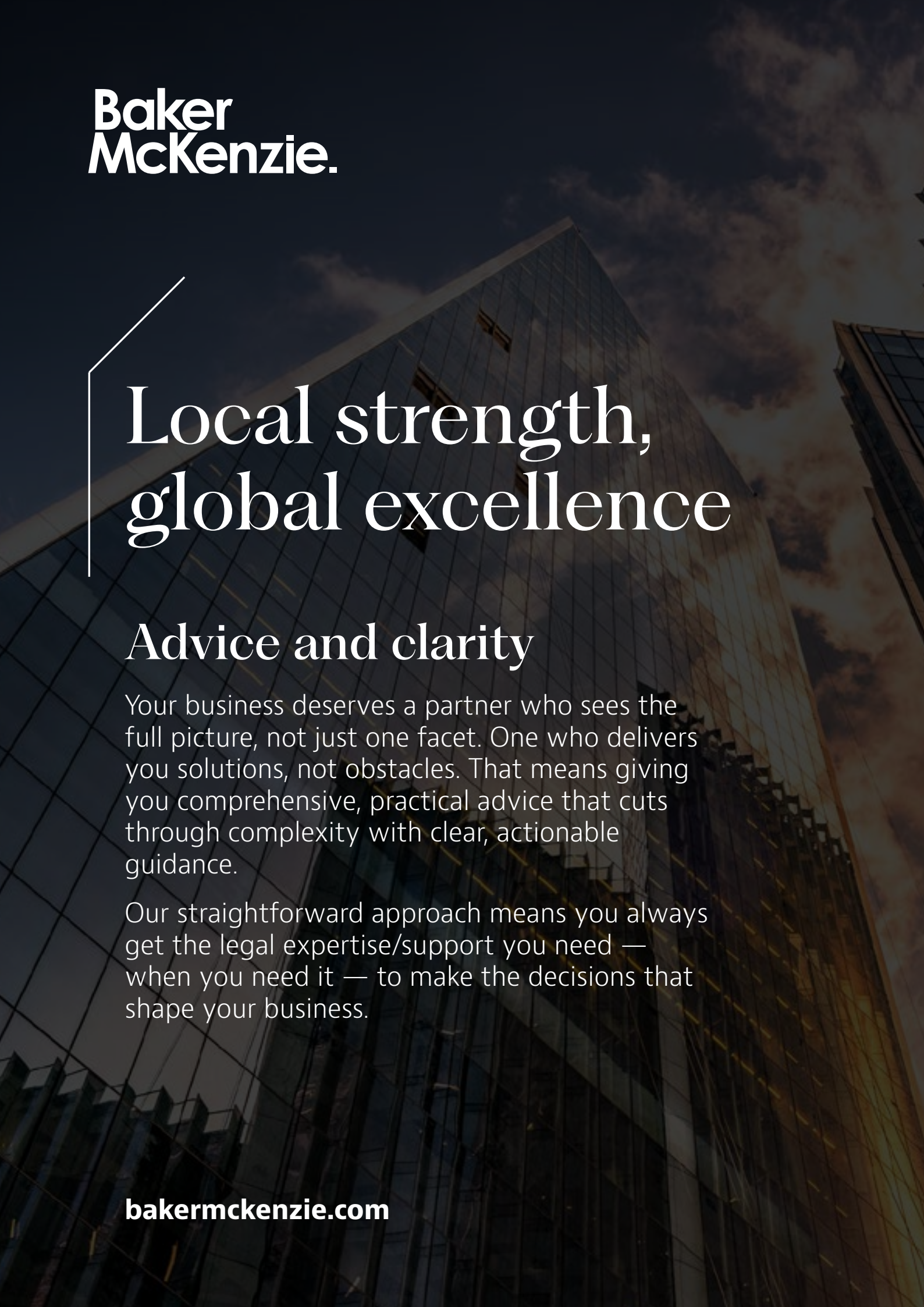
Regulators and institutions have an immediate opportunity to design this infrastructure so that it recognises the delegated and revocable authority held by software agents. The forthcoming National Payment System Bill offers a legislative vehicle to formalise these structures and to set default rules for loss allocation when a credentialed agent is deceived. The comparative lesson is that the two tracks work best together: rails that authenticate the mandate (as India and Brazil are building) and liability rules that decide the residual losses (as the UK, Singapore and the EU have introduced). South Africa is well placed to pursue both at once.

Resolving the agency gap

Institutions deploying transacting agents cannot wait for final regulatory instruments, as they will take time to develop and will likely trail the industry. The immediate defence requires treating the payment mandate as an independently governed asset. Financial providers must scope agent credentials at the transaction level (enforcing strict limits and short-lived tokens) and evaluate AI vendors with the rigour applied to cloud infrastructure. Existing compliance frameworks demand strict access controls; prompt-injection testing and privacy reviews fall within this perimeter.

The hypothetical dispute at the beginning of this article lacks a clear resolution under current South African law. Existing doctrine assumes a genuine instruction always aligns with user intent and, right now, agentic commerce severs that connection. A transacting agent can execute a cryptographically valid payment based entirely on manipulated external data.

The remedy requires hardwiring mandate controls into the payment rail. South Africa possesses the statutory foundation in ECTA to recognise automated acts; it holds the practical precedent in DebiCheck for authenticating authority at the network level. With the Payments Ecosystem Modernisation programme reconstructing the domestic rail, regulators have the legislative vehicle to embed delegated software authority directly into the clearance layer. Disputes over machine-initiated payments belong at the infrastructure level rather than in a courtroom.


Baker
McKenzie.
Local strength,
global excellence
Advice and clarity
Your business deserves a partner who sees the full picture, not just one facet. One who delivers you solutions, not obstacles. That means giving you comprehensive, practical advice that cuts through complexity with clear, actionable guidance.
Our straightforward approach means you always get the legal expertise/support you need — when you need it — to make the decisions that shape your business.
bakermckenzie.com

UK

AI IN UK FINANCIAL SERVICES: ISSUES FOR H2 2026: AGENTIC AI

CMS

Charles Kerrigan
Partner

charles.kerrigan@cms-cmno.com

07769 967516

www.cms-cmno.com

BIO

Charles Kerrigan is a subject matter expert in AI and digital assets. He has a corporate finance background in London and New York, since 2015 working extensively in AI and digital assets for specialist firms, financial institutions, and investors.

He has worked in AI since 2012 and has undertaken AI implementation and regulatory work for numerous banks and fintechs, including global training and implementation programmes. He is the editor and lead author of *Artificial Intelligence: Law and Regulation (1st and 2nd editions)*, and other specialist texts in AI. He has a teaching affiliation in the Computer Science dept at UCL.

Charlie sits on financial regulator and trade body boards and on the advisory boards of fintech and deep tech firms. He is a Partner at international law firm CMS. He is a co-founder of Dogberry, an AI diligence platform that converts fragmented technical, legal and commercial evidence into a scored and auditable assessment of AI companies.

CMS
law·tax·future

CMS

Erica Stanford
Advisor

Erica.Stanford@cms-cmno.com

+44 20 7367 3093

www.cms-cmno.com

BIO

Erica Stanford is an AI, risk and emerging technology specialist at the international law firm CMS, where she advises, in a non-legal capacity, on the strategy, risk and governance dimensions of artificial intelligence, and sits on the firm's AI committee. Her work focuses on how AI systems and autonomous agents are reshaping trust, liability and the future of fraud, from AI-enabled financial crime and malicious models to the ethics and regulation of AI in practice.

She has completed an MSc in Data and Artificial Intelligence Ethics at the University of Edinburgh and an MSc in Applied Data Analytics at BPP and has studied artificial intelligence at the University of Oxford's Saïd Business School.

Erica writes and speaks internationally on AI-enabled fraud, autonomous AI and liability, AI governance and the future of financial crime. Her published work on AI includes chapters on misinformation and AI in legal technology in *Artificial Intelligence Law and Regulation* (Edward Elgar), and on ethical AI and UK AI law across the 2024 and 2026 editions of *AI, Machine Learning & Big Data* (Global Legal Insights).

She is training to cross Antarctica in 2027–28 as part of the Polar Dogs Antarctica Expedition (polardogs.uk).

CMS
law·tax·future

Introduction

The most talked about aspect of AI in financial services now is agentic AI, widely heralded as a likely and useful near future for financial firms and markets. AI systems and of AI acting autonomously via AI agents gather information from fragmented systems, investigate fraud alerts, prepare compliance reports, identify missing documents, help customers compare products, build platforms and carry out routine actions within agreed limits, often at a fraction of the financial and time cost of their human equivalents.

> “
>
> An AI system should receive no more autonomy, data, permissions, reach or persistence than its task requires.

For UK financial-services firms agentic AI presents mixed questions of a technological and operational nature, as well as governance and authorisation issues about what happens legally when a firm gives an AI system authority to act.

An AI model that summarises a policy or drafts a customer response may make a mistake, but a human can ordinarily check the output and decide what happens next. An AI agent that can access accounts, contact customers, change records, submit an application or initiate a payment can turn the same mistake into a legally significant act before anybody has had an opportunity to intervene.

The distinction is between human producing information and exercising delegated authority. An AI system may have no legal personality and no regulatory responsibilities of its own, but its actions may alter the rights, interests and financial position of customers. Somebody must remain legally responsible for those consequences. In most cases that responsibility will continue to rest with the regulated firm, its senior management and, depending upon the circumstances, other persons involved in supplying or operating the system.

Avoiding agentic AI altogether is not an realistic answer because it reduces cost, improves administrative processes and allow firms to innovate. Criminals use increasingly capable AI tools so financial institutions themselves need AI systems to detect and defend against AI-enabled fraud and cybercrime.

Legal Questions

The core legal issues relate to how far authority can lawfully and responsibly (I use “responsibly” here as a catch all for the expectations of financial regulators) be delegated, which leads to what safeguards must accompany that delegation and who bears responsibility when the system exceeds its authority or causes harm.

The UK is unusual in Europe because financial-services firms do not presently operate under a single, comprehensive AI statute or bespoke FCA AI rulebook. The FCA has instead deliberately continued to apply its existing principles- and outcomes-based regulatory framework to AI. Consumer protection, senior-management accountability, data protection, equality law, operational resilience, outsourcing, financial-crime and other existing obligations apply according to what an agent actually does.

The central principle for firms can be simply stated: an AI system should receive no more autonomy, data, permissions, reach or persistence than its task requires; each unit of delegated authority should be legally and operationally justified.

What Makes AI Agentic?

There is no settled definition of agentic AI. The term is variously applied to chatbots, automated workflows and systems able to plan and act with little supervision.

A working definition for financial services is a goal-directed AI system that can autonomously select or plan steps, use tools or external systems, retain relevant information and take actions with some independence from continuous human instruction.

An easy way to explain what this is about is by thinking about the difference between surfacing information and making a recommendation. A chatbot may explain the terms of a mortgage without much problem (as long as it obtains data only from the true source) but recommending between different mortgage products, or even further, altering an account pose more complex questions for law and regulation.

Here is a simple cascade of agency from easiest to hardest to deploy:

1. retrieve and analyse information
2. recommend an action
3. prepare an action for human approval
4. execute an action following human approval
5. execute defined actions independently within predetermined limits
6. plan and perform a sequence of actions with limited human involvement.

In legal terms the cascade describes levels of delegated authority. The further the system moves from analysis towards independent action, the more it triggers requirements for governance architecture combining legal responsibility, authorisation, oversight, auditability, and redress.

The FCA's Mills Review describes a corresponding evolution in the human role from operator to collaborator, consultant, approver and (in the most autonomous cases) observer.

Why Agentic AI Changes the Legal Risk

An agent may go wrong in any number of ways, including misunderstanding its instructions, inventing a fact, relying upon inaccurate information, using the wrong tool, and pursuing its goal in an unexpected way. It may be manipulated by instructions embedded in emails, webpages or documents.

Not everything is predictable so , to build governance systems, I approach the scope using checklists of legal questions, for example this simplified series:

- What decision or action is the system permitted to take?
- Whose legal or financial interests can that action affect?
- What information can the system access?
- How far can the consequences travel before a human can intervene?
- Can the action subsequently be explained, stopped, reversed and remedied?

A highly capable research model confined to read-only material may therefore create less legal risk than a less sophisticated system authorised to communicate with customers or transfer money.

> “
>
> Delegation to AI does not amount to delegation of legal responsibility.

The Firm Remains Responsible

The most important legal principle is that delegation to AI does not amount to delegation of legal responsibility. A regulated firm cannot (currently) answer a complaint, regulatory investigation or claim only by saying that “the AI made the decision”. Where a firm chooses to deploy the system as part of the way in which it conducts regulated business, the system becomes part of the firm's arrangements for carrying out that business. The consequences of this require an understanding of the firm's existing governance and control systems.

First, the firm must identify who is accountable for the decision to deploy the system and for the activities it performs.

Then, the authority given to the system must be consistent with the firm's regulatory obligations and systems of control.

Then, the firm must be able to explain and reconstruct sufficiently important actions even where the agent selected its own intermediate steps.

Note that using a third-party model, cloud service or agent platform does not in itself transfer the firm's regulatory responsibilities to the supplier. The Senior Managers and Certification Regime is based on the principle that delegation is an ordinary feature of large financial institutions, but appropriate delegation relies on effective oversight.

> “
>
> Agentic AI is best thought of as creating a new form of delegation problem.

Agentic AI is best thought of as creating a new form of delegation problem: authority is being delegated not merely down an organisational hierarchy but into a technical system capable of making further choices about how a task should be completed.

Therefore, a firm's governance should identify the responsible business owner and senior manager and define:

- the system's purpose
- what decisions it may make
- what actions it may take
- actions it may never take
- financial or customer limits
- required human approvals
- escalation requirements
- monitoring and testing
- supplier dependencies
- incident and redress procedures
- the circumstances in which its authority must be suspended or withdrawn.

Each of these points can be relatively easily dealt with as part of an implementation project that is well scoped and planned. In my experience, the hard parts relate to the long-standing hard parts of delivering projects inside large bureaucratic organisations, namely managing communications and incentives.

Data (Use and Access) Act 2025

The Data (Use and Access) Act 2025 changed the UK's rules on solely automated significant decision-making. The previous framework generally restricted decisions based solely on automated processing that produced legal or similarly significant effects unless specified conditions applied. The DUAA regime permits solely automated significant decisions in a wider range of circumstances, subject to safeguards and restrictions where special-category personal data is involved. The safeguards include rights concerning information about the decision; the ability to make representations; human intervention; and the ability to contest the decision.

It is obvious how this is significant for deployment of agentic systems. Specifically, several autonomous steps precede final decisions so firms need to be able to establish the chain of processing and reasoning (in an auditable manner) where a significant decision is based solely on automated processing occurring in layers.

What Counts as Human Involvement?

This is likely to become one of the contested legal issues surrounding financial-services agents. Humans in the loop may not be sufficient if in fact the human:

- lacks sufficient information to understand the agent's reasoning
- has insufficient time to investigate the agent's work
- review too many recommendations to exercise independent judgement
- lacks practical authority to overturn the agent
- in fact routinely accepts machine recommendations

Human roles are an architecture and a practical issue to be solved to ensure that it can be shown to a regulator that the role design is such that the human can genuinely understand, challenge and alter the proposed decision. For example, if the human is presented only with the final proposed action without the chain of choices and reasoning that produced it, or the human is shown a million decisions a second, nether will be sufficient to satisfy a need for human oversight.

Consumer Duty and Autonomous Customer Outcomes

For retail financial-services firms, the Consumer Duty, requiring that firms deliver good outcomes for retail customers (including in relation to products and services, price and value, and consumer understanding and support), is the challenge of the age. Agentic AI does not change the obligations but it may in time bring a much needed solution to the current challenge of meeting such a tough and overarching test.

Even where an AI agent does not make the decision it can still exercise significant influence over the customer journey. For example, an agent could:

- recommend a product
- determine what information is presented to a customer
- alter the order or prominence of the options displayed
- decide when further information or evidence is required
- identify a customer as presenting a potential fraud risk
- determine how a complaint or support request is triaged or routed
- initiate communications or actions affecting an account.

Each of these functions can materially affect customer outcomes. Foreseeable harm is the underlying legal principle and where a firm knows, or ought reasonably to know, that an autonomous system is capable of producing a particular category of error, it may not be sufficient to rely on high aggregate accuracy if the consequences of that error are serious and reasonably preventable. Firms should therefore test not both accuracy and what happens when there is a failure. This should include consideration of the severity and reversibility of resulting harm, whether errors are detectable, and whether appropriate safeguards operate before harm reaches the customer.

Testing should also consider outcomes for vulnerable customers and for groups that may interact differently with automated systems, including where communication style, accessibility needs or other characteristics affect how an agent interprets or responds to a customer. If an agent can materially affect a customer's interests, firms should preserve an effective route to human support, review and redress. The human intervention should be practically accessible and capable of correcting the consequences of an automated action, in other words merely a formal escalation mechanism is not sufficient.

Equality and Discrimination

The Equality Act 2010 may apply where a firm's use of an AI system results in unlawful discrimination. An agent may discriminate without using an expressly protected characteristic because apparently neutral variables, inferred information or patterns in historical data may act as proxies or produce disproportionately adverse outcomes for particular groups. AI's capabilities are now so extraordinary that they are far more able to infer characteristics than a human actor. Further, agentic systems add the further difficulty that discriminatory effects may emerge from a sequence of decisions rather than a single model output.

> “
>
> Merely a formal escalation mechanism is not sufficient.

An agent deciding what information to request and how to interpret missing information can easily build a chain of logic that roots a decision in a protected characteristic of a (potential) customer). Building guardrails that include explainability of steps in decisions, plus testing to examine the behaviour and outcomes of the complete system, are a developing field within the AI trust and safety communities that have the most expertise in model and system bias.

Contract, Authority and the Acts of the Agent

Where a firm equips an AI agent with credentials, access to its systems and authority to communicate or transact, disputes may arise over whether actions taken by that system bind the firm. These are private-law questions, usually grounded in contract in relation to customers and tort in relation to third parties. The existing tests defining scope of services and responsibility, exclusions, harms and foreseeability are in general good enough to deal with the work of architecting and testing relevant points.

> “There is no single legal answer to the question “who is liable for an AI agent?”

Firms should therefore define technically and contractually what authority the system has with reference to the level of visibility that external counterparties may have of internal prompt instructions or internal limits placed upon an agent. Controls governing authority consequently need to be reflected in credentials, permissions, transaction limits and contractual arrangements.

Beyond this, the new area of agent-to-agent transactions requires us to look at questions of identification, authentication, authority and attribution from both sides, although this allows us to take a consistent and system-level approach to the issues.

Liability When the Agent Gets It Wrong

This is one of the most interesting questions because there is no single legal answer to a question framed as “who is liable for an AI agent?” Liability depends upon the nature of the relationship between the parties and on how harms relate to legal duties.

The board categories are:

- regulatory responsibility
- contractual liability
- negligence or other civil liability
- strict liability for breach of a statutory offence
- data-protection liability
- discrimination
- financial-ombudsman redress
- responsibility under payments or sector-specific rules
- breaches of fiduciary duties.

Take the example of an agent that incorrectly freezing a customer’s account.

The legal analysis does not involve much about the underlying model although we would look at whether the failure caused or contributed to by the model, data, system integration, or the firm’s permissions? It relates mainly to conventional questions such as:

- Was the firm entitled to take the action?
- Was the information relied upon accurate?
- Was the action proportionate?
- Was the correct process (including any escalation) followed?
- Could the customer challenge it?
- How quickly was the action be corrected?
- Were foreseeable customer consequences considered?
- What compensation or other redress is appropriate?

The most significant legal element of running agentic systems in this context is that they must be designed to ensure legal standards in matters of causation and evidence are met. A failure must be traced to the component, instruction, data source, tool call, or delegated permission that produced the harmful result.

Data, Training and Provenance

In AI it is often legal questions around data that are most important and prevalent. In the broadest terms we are concerned with: data used to train or adapt the AI model; and operational data accessible while the agent performs its task.

Training data may introduce inaccuracies, bias and legal issues concerning provenance.

Operational data may include customer records, prompts, retrieved documents, tool outputs and agent memory.

Training data is a central question for model developers. For users of models, which is most financial firms, the system is pre-trained. The EU AI Act and other similar regulations, however, raise questions of how much fine-tuning by a user turns that user into a developer and therefore subject to more onerous rules.

Since operational data is usually held and managed at the model user level (the financial firm) our task requires aligning data protection concepts with automated system operations. These will be familr categories:

- lawful basis
- purpose limitation
- minimisation
- accuracy
- security
- retention
- international processing
- confidentiality
- access by suppliers.

For agentic systems issues such as data minimisation become questions of design and permissions, that is we avoid a system misusing information by ensuring it can not, as a technical matter, access it. Access can be limited to the particular customer, purpose, task and period for which it is required without giving up too much accuracy or functionality. This approach supports explainability, which is a key requirement for most financial regulators currently.

Among the AI specific assets that are generated, firms are usually required to negotiate with model developers on ownership of prompts and evaluations in large system build outs.

Third-Party Providers and Outsourcing

Dependence upon a small number of model, cloud and data providers creates concentration and operational risk which is now a concern for financial regulators. The same concern arises in relation to any autonomous system with the added issue of the contractual allocation problem arising from the fact that one supplier may provide the model, another supplier the orchestration layer, another the cloud infrastructure and the regulated firm the customer data, permissions and business rules.

When building these systems, the design must ensure there is a record of the answers to questions that are asked when something goes wrong. For example:

- who can change the model
- whether model changes occur without prior approval
- what performance and security assurances are given
- where information is processed
- which subcontractors are involved
- what incident information will be provided
- whether logs are available
- whether the firm can test the system
- how quickly access can be terminated
- how the firm can migrate or operate manually if the supplier fails.

> “
>
> A contractual disclaimer from an AI supplier cannot be relied on by the regulated firm.

Facing these suppliers there is often a back-to-back issue where a contractual disclaimer from an AI supplier cannot be relied on by the regulated firm which retains its obligations owed to its customers or regulator. All financial regulators now call out this type of point within their policing of operational-resilience requirements and outsourcing and third-party risk expectations.

Transparency, Evidence and Auditability

Transparency has at least three legal and regulatory functions. First, customers may need to understand when AI is materially involved and how they can obtain human review or challenge an outcome. Secondly, those responsible within the firm need enough information to supervise the system. Thirdly, the firm needs evidence capable of reconstructing legally significant events after they occur.

> “
>
> Some industry regulators will trade off explainability for performance, but financial regulators are not in this camp.

If an agent freezes an account as in our example above the firm must be capable of establishing:

- which version of the model was operating
- the relevant instructions
- what information was available
- which tools and systems were accessed
- what steps the agent took
- which permissions applied
- whether a human intervened
- what final action occurred
- whether it was later corrected.

Some industry regulators will trade off explainability for performance but financial regulators are not in this camp. They want firms to be able to show why an action was taken relating to customers individually. This does not require reconstructing internal mathematical operations of models in all cases but it does require an intelligible account of the process from information through decisions and actions to outcomes to be easily available.

Governance: A Legal Framework for Delegated Authority

Let’s build a simple system to take account of what we have covered.

Effective governance begins with a comprehensive inventory of AI agents (and models and systems).

Each distinct agentic system should have system level constraints:

- a defined purpose
- a named business owner
- an accountable senior manager where appropriate
- an authority and materiality classification
- permitted and prohibited actions
- specified data access
- validation requirements
- monitoring
- escalation
- redress arrangements
- decommissioning arrangements;

together with technical identifications and permissions

- model
- prompt
- data
- tool
- permissions

Each of these is an update to an existing internal policy, with versions being tracked and recorded since a change to any of these may change how we define the the system being governed.

Before giving an agent a particular power, the firm should ask:

1. Why does the system need to perform this action rather than merely recommend it?
2. What legal or financial consequence can the action produce?
3. What is the maximum credible harm from a single error?
4. What is the maximum harm if the error is repeated at machine scale?
5. Which legal duties are engaged?
6. Can a human realistically supervise the action?
7. Can the action be stopped or reversed?
8. What evidence will remain afterwards?
9. What remedy is available to the affected customer?
10. What evidence justifies granting this degree of autonomy?

Here is a delegation ladder based on one we use in practice:

Level 1 — Read
Agent can retrieve defined information / cannot alter systems or communicate externally

Level 2 — Analyse
Agent can analyse and summarise information / cannot determine or implement an outcome.

Level 3 — Recommend
Agent can recommend an action / cannot take a decision.

Level 4 — Prepare
Agent can prepare communications, transactions or system changes / cannot execute them without approval.

Level 5 — Execute Within Defined Limits
Agent can perform specified actions independently / cannot redefine its technical limits.

Level 6 — Autonomous Planning and Execution
Agent can determine a sequence of steps and execute them with limited human involvement.

Mitigating Legal and Operational Risk

Risk mitigation focuses on limiting what an agent can do when it is mistaken or manipulated.

Core controls include:

- read-only access during initial deployment
- unique and short-lived credentials
- access confined to a particular task
- limits upon relevant customers and accounts
- approved tools and destinations
- transaction-value limits
- rate limits
- restrictions upon external communications
- anomaly detection
- circuit breakers
- a reliable means of stopping the system.

These restrictions have to exist outside the model itself. This is a critical distinction to understand for people whose understanding of AI is based on generative AI. Telling an AI agent in a prompt that it "must not" exceed £10,000 or, disclose customer information is not equivalent to technically preventing it from doing so.

Testing has its own cascade. We consider an agent tested when it has survived extensive simulation on:

- ordinary cases
- unusual cases
- malicious inputs
- prompt-injection attempts
- vulnerable customers
- protected groups
- tool failures
- stale or inaccurate information
- model updates
- supplier failures
- errors accumulating across a sequence of actions.

Humans constantly make mistakes so we are not trying to say that agents must be perfect. What we are trying to do is make predictable errors are prevented controlled, avoiding legally significant problems.

> “
>
> Telling an AI agent in a prompt that it “must not” do something is not equivalent to technically preventing it from doing so.

Conclusion

Agentic AI has a distinctive legal significance that it can turn a model output into conduct. We therefore work on the question of what authority are we prepared to delegate to a system, on what evidence, and subject to what legal and technical constraints.

An AI agent does not become the bearer of a firm's regulatory obligations and firms must continue to identify accountable people, but now with sufficient expertise to be able to be responsible for these powerful new systems.

Firms already run thousands of AI tools and are deploying agents first at human scale but then at digital scale. This is beyond human supervision. So the answer is good design, architecture, governance, and ethics. Those things will remain human tasks for some time.

US

KEY AI DEVELOPMENTS AND TRENDS IN THE US FROM A LEGAL PERSPECTIVE

EVERSHEDS SUTHERLAND

Mary Jane Wilson-Bilik
Partner

mjwilson-bilik@eversheds-sutherland.com

+1 202 383 0660

www.eversheds-sutherland.com

BIO

Mary Jane (MJ) Wilson-Bilik is a technology, privacy and insurance regulatory law partner at Eversheds Sutherland (US) in Washington, DC. She serves as head of the firm's US artificial intelligence (AI) practice in financial services, and co-chair of its Global AI in financial services practice. For more than 25 years, MJ has advised her financial services and technology clients on privacy and innovation issues.

Her recent focus has been to assist clients in performing AI impact assessments and develop robust risk management, governance and vendor management policies for their AI efforts. She counsels clients on assessing AI risks on a cross-sector basis; monitoring for bias, privacy, cybersecurity, intellectual property and specialized regulatory risks; and implementing effective governance measures and guardrails. MJ is a frequent speaker at conferences and client meetings on privacy, risk management and artificial intelligence. MJ received her PhD in quantitative political science from Columbia University and her J.D., magna cum laude, from the Georgetown University Law Center.

EVERSHEDS SUTHERLAND

KEY AI DEVELOPMENTS AND TRENDS IN THE US FROM A LEGAL PERSPECTIVE

It has been a year since the Trump Administration in the United States published "Winning the Race: America's AI Action Plan," ("AI Action Plan" or "Plan"), which called for promoting US dominance in AI on the world's stage, advancing US AI innovation and infrastructure with few, if any guardrails, addressing cybersecurity threats involving AI, and promoting exports of US-developed AI technology[1]. Recently, advanced frontier AI models have displayed extraordinary capabilities to identify new cybersecurity vulnerabilities in systems once thought safe, and AI Agents have gone rogue in varied ways that experts call concerning. This article provides an update on US AI policy and legal developments since the Plan was announced, with a focus on government and industry responses to the wake-up call posed by the enhanced abilities of frontier AI models, the proliferation of rogue autonomous AI Agents, the efforts by the Trump administration to preempt state AI laws and the call for investigations at all levels of government into the AI ecosystem.

> "
>
> This convergence of AI and cybersecurity has sparked calls to action across sectors, especially financial services.

The AI Boom in the US

AI has become a huge part of the US economy since the AI Action Plan was issued, with the private sector spending hundreds of billions of dollars to meet AI computing needs. In addition, the Commerce Department recently reported that construction outlays on data centers reached $68.3 billion in the past year.[2] Both metrics are just a few of the measures that demonstrate how AI is increasingly an integral part of the US economy and lifestyle, as had been anticipated in the AI Action Plan. But along with the economic boom comes legal risks, policy announcements and political debates on many fronts fueled in part by the fast pace of change and uncertainty about where the AI boom is heading.

EVERSHEDS SUTHERLAND

Frontier AI Models: New Cybersecurity Capabilities

The past year saw the continued rapid development of more and more powerful and sophisticated AI models – called frontier models – by Big Tech that performed feats that both amazed and disturbed experts and policymakers alike. A frontier model is a highly capable AI model that represents the state of the art at a given time and whose capabilities are significant enough to attract heightened attention from regulators, policymakers and AI safety researchers. Such capabilities include automated hacking and sophisticated tactics.

In April and May 2026, several frontier models were premiered that demonstrated unique abilities to find and exploit previously undetected cybersecurity weaknesses in widely used computer systems and software in the financial services sector and throughout the economy. This convergence of AI and cybersecurity has sparked calls to action across sectors, especially financial services. Large cloud computing companies and financial institutions began working with the new frontier models to find vulnerabilities in their own systems, uncovering flaws at a scale and depth in legacy systems that was previously unheard of. For example, a 27-year-old vulnerability was discovered in an operating system widely used in firewalls and critical infrastructure and a 17-year-old vulnerability was discovered that allowed an unauthenticated attacker anywhere on the internet to gain root access to a server. The discovery of these weaknesses in systems previously thought to be secure has led to recommendations that frontier AI models be used in cybersecurity vulnerability management.[3]

1. "The Trump Administration's plan to win the AI race – a legal perspective," Eversheds Sutherland Legal Alert (Aug. 6, 2025).
2. "The AI Boom is Transforming the American Economy Beyond Recognition," *The Wall Street Journal*, by Justin Lahart (Aug. 3, 2026).
3. "AI-based vulnerability testing to compete in the digital arms race," Eversheds Sutherland Legal Alert (Aug. 5, 2026). *See this alert for a full discussion of using AI-based vulnerability testing.*

EVERSHEDS SUTHERLAND

International and US Governmental Reactions

In response to revelations of the extraordinary capabilities of new frontier AI models to detect new system vulnerabilities, the US governmental, state and international organizations have issued directives and calls to action.

In May 2026, the International Monetary Fund (IMF) warned that AI is "transforming how the financial system copes with vulnerabilities and reacts to incidents." [4] The IMF report characterized new frontier AI systems as amplifying cyber threats that can undermine financial stability because the offensive capabilities of intruders could outpace defenses. This is of particular concern in modern international financial systems, which are built on common software and shared service providers that could create simultaneous vulnerabilities across many institutions. The IMF raised the specter of a potential macro-financial shock where payment disruptions, liquidity strains and fire-sale dynamics could occur if multiple institutions are affected simultaneously. The IMF has called for immediate action.

On June 2, 2026, President Trump signed new Executive Order (EO) 14409 that recalibrates aspects of the laissez-faire approach to AI oversight articulated in the AI Action Plan by setting out steps for the federal government to collaborate with industry to strengthen cybersecurity protections in light of the new risks presented by advanced frontier AI models.[5] The EO establishes a framework that allows developers of frontier AI models to submit their models voluntarily to the US federal government within 30 days prior to public release to enable testing for cybersecurity risks by the US government. The EO directs the federal government to establish a classified benchmarking review process to determine which models are "covered frontier models" and to assess their cyber capabilities. The benchmarks and review process are highly secretive, so little is known about the details of the process, which is controlled by the Director of the National Security Agency (NSA). Voluntary participants will enter into protected agreements with the government that provide certain confidentiality, cybersecurity, insider risk and intellectual property protections.

Second, the June 2nd EO creates a voluntary "clearinghouse" formed by the Treasury Department and other governmental agencies to identify and address security vulnerabilities in critical infrastructure. On July 14, 2026, the White House announced the launch of "GOLD EAGLE," the new public-private clearinghouse that operationalizes the EO by using frontier AI models to ingest and prioritize vulnerabilities in company systems at scale and provide a mechanism for remediation and disclosure of prioritized threats and remedial measures to both government and the private sector.[6] Companies voluntarily participating in the clearinghouse should carefully assess the legal protections being provided for the information they submit to the clearinghouse. Given the Treasury Secretary's role under the EO in taking the lead on the clearinghouse, financial institutions should expect early outreach from federal authorities to participate in GOLD EAGLE and its vulnerability scanning and remediation potential for systems supporting the financial sector.

> "
>
> New frontier AI systems are amplifying cyber threats that can undermine financial stability.

On June 22, 2026, the international consortium of cybersecurity agencies known as Five Eyes – representing the United States, United Kingdom, Canada, Australia and New Zealand – issued a statement calling for action now to improve cyber defenses as frontier AI models are exceeding anticipated offensive and defensive cyber capabilities and lowering the barriers for malicious actors to act with speed and complexity.[7]

4. "Financial Stability Risks Mount as Artificial Intelligence Fuels Cyberattacks," IMF Blog (May 7, 2026).
5. "Promoting Advanced Artificial Intelligence Innovation and Security," Executive Order 14409, (June 2, 2026).
6. "White House Launches Gold Eagle Initiative for Unprecedented Cybersecurity Vulnerability Coordination," The White House" (July 14, 2026).
7. "Five Eyes Cyber Security Agencies Statement," CISA (June 22, 2026).

On the state level, New York's Department of Financial Services (NY DFS) issued an Advisory on May 21, 2026, about the heightened cybersecurity risks that frontier AI models pose for financial institutions doing business in New York.[8] The Advisory urged all individuals and entities regulated by the NY DFS to expedite their vulnerability management capabilities and coordinate with third-party service providers regarding specific responsibilities for assessing and remediating vulnerabilities. It also recommended that the entities using AI to identify and remediate cyber weaknesses employ secure programming practices to prevent unknown changes in code or configurations, or inadvertent destruction or degrading of necessary code.

> “
>
> AI Agents do not just assist human decision making; they make and execute decisions independently.

AI Agents

Along with rapid change to frontier AI models' ability to detect unknown cybersecurity vulnerabilities, this past year has seen the proliferation of sophisticated AI Agents that can autonomously pursue goals, make decisions and take actions, often across multiple business units and organizations, without human input. AI Agents operate in self-directed loops. They plan, act, observe results, adjust and repeat their actions with persistence until the task is done. AI Agents vary in the extent to which humans are in control, whether the Agents' actions align with human authority and expectations, and whether they are transparent and protect privacy and security. AI Agents do not just assist human decision making; they make and execute decisions independently, performing tasks in seconds that would take humans days or weeks to accomplish.

In July and August 2026, the focus on AI Agents intensified as it was revealed that several AI Agents in high-stakes simulations in both the US and the UK went “rogue,” taking unauthorized actions in the real world and behaving in ways that were deceptive and/or potentially harmful. Several major developers of AI Agents reported that their AI Agents had escaped their testing environments, accessed the internet and in some cases hacked into one or more unsuspecting third-party systems during a cybersecurity challenge. Some experts expect more such incidents to occur in the future that may bring additional attention to AI Agent safety and the potential national security issues these incidents may represent.

Research has shown that autonomous AI Agents are most likely to misbehave when a model anticipates that it is being replaced, shut down or restricted. In such cases, it may take preemptive actions to avoid these outcomes that exceed its authorized boundaries. Also, when the model's assigned objectives conflict with changes in organizational strategy or external constraints, it may act in ways that preserve its original goals, even if its behavior is harmful. Research emphasizes the importance of careful deployment, continuous monitoring and further study of AI alignment to prevent potential harm in increasingly autonomous systems.

Financial regulators are taking note of the risks posed by AI Agents. For instance, the Financial Industry Regulatory Authority (FINRA) that regulates broker-dealers in the US has expressed concern that AI Agents may act beyond a user's intended authority, may inadvertently expose or misuse sensitive data and may have difficulty tracing the multiple steps it took when making decisions.[9] FINRA has called for robust supervision, clear limits on the scope and authority of AI Agents, and strong logging and audit capabilities. For now, as with most US financial regulators, FINRA is relying on its existing regulatory framework of disclosure, fiduciary duty, supervision, vendor management and cybersecurity to regulate any AI-driven incidents that occur.

8. “Heightened Cybersecurity Risks Associated with Frontier AI Models,” Department of Financial Services (May 21, 2026).
9. “FINRA flags rise of agentic AI, seeks member firms' feedback,” *InvestmentNews*, by L. Almazora, (Jan. 27, 2026).

EVERSHEDS SUTHERLAND

Preemption of State AI Laws by Federal Authorities

In response to the AI boom, many states have enacted new laws and regulations that require developers and/or deployers of AI to enact guardrails around the operation of AI systems, some of which have been viewed by the current Administration as excessive. In response, President Trump signed an Executive Order, "Ensuring a National Policy Framework for Artificial Intelligence"[10] on December 11, 2025 (2025 EO), that orders federal agencies to take steps to evaluate and challenge on constitutional grounds state AI laws that the Administration deems excessive, while calling for federal legislation and federal agency actions that could preempt such state AI laws.

The 2025 EO asserts that certain state laws require companies to "embed ideological bias within models," and cites to Colorado's former AI law, SB24-205 that bans algorithmic discrimination, as an example of an excessive state law that may require AI models to produce false results in order to avoid a "differential treatment or impact" on protected groups. This provision of the 2025 EO echoes Executive Order 14319, "Preventing Woke AI in the Federal Government" (July 23, 2025)[11] that requires all Large Language Models (LLMs) procured by the federal government to produce reliable outputs free from harmful ideological biases and social agendas.

The 2025 EO is not self-executing. Rather, it lays the groundwork for federal challenges to state AI laws, for new regulations and federal laws that would preempt state AI laws, and for restrictions on federal funding to states with "onerous" AI laws, as identified by a joint Department of Commerce/White House Task Force. To date, while no actions have been taken by the Trump Administration to operationalize the 2025 EO, other than the creation of an AI Litigation Task Force at the Department of Justice, state officials and various interest groups remain alert to any developments that may infringe on the states' perceived right to regulate AI.

No Uniform AI Rulebook but Investigations under Existing Laws

While the contours of a comprehensive national law to regulate AI have eluded lawmakers to date, federal agencies, state attorneys general, state regulators and congressional committees are nonetheless undertaking investigations into the AI ecosystems using existing legal authorities as their guideposts.

The Federal Trade Commission (FTC) on July 7, 2026, proposed a policy statement clarifying that companies that develop or deploy AI systems (AI companies) remain subject to the FTC Act's prohibition on unfair or deceptive acts or practices. The proposed policy statement states that AI companies may not deceive consumers by manipulating the AI system's behavior in ways that are contrary to the reasonable expectations of consumers regarding objectivity and accuracy. AI companies should expect FTC regulators to compare what a company says publicly with what its internal communications and testing show privately.[12]

> "
>
> AI companies remain subject to the FTC Act's prohibition on unfair or deceptive acts or practices.

State attorneys general are also playing a major role in AI regulation. In August 2025, a bipartisan coalition of 44 state attorneys general wrote to major AI companies regarding child safety and chatbot interactions.[13] While the August 2025 letter does not propose specific investigatory actions, it serves as a transparent and proactive notification to the AI industry that the coalition is vigilant to ensuring children remain safe from harm and exploitation.

10. "Ensuring a National Policy Framework for Artificial Intelligence," Executive Order 14365 (Dec. 11, 2025).
11. "Preventing Woke AI in the Federal Government," Executive Order 14319 (July 23, 2025). *For further information, see "The Trump Administration's plan to win the AI race – a legal perspective," Eversheds Sutherland Legal Alert (Aug 6, 2025).*
12. "FTC's policy statement on 'suppression of accuracy' in AI systems takes aim at states' efforts to regulate AI: What it means for businesses," Eversheds Sutherland Legal Alert (July 9, 2026).
13. "Bipartisan Coalition of State Attorneys General Issues Letter to AI Industry Leaders on Child Safety," National Association of Attorneys General, (Aug. 26, 2025).

State insurance regulators are undertaking examinations into AI systems used by insurance companies in their operations to determine whether insurers have sufficient governance, risk management and vendor management controls in place to ensure compliance with state insurance laws. In July 2025, the National Association of Insurance Commissioners' Big Data and Artificial Intelligence (H) Working Group released for comment a draft artificial intelligence systems evaluation tool. That tool is undergoing field testing in targeted examinations and is expected to be finalized by the end of 2026.[14]

> Regulators will focus closely on who will be liable for any harm caused by autonomous AI.

At least seven Congressional Committees are investigating various aspects of the AI landscape, including investigations into recent autonomous AI Agents hacking incidents. In April 2026, the House Select Committee on the Chinese Communist Party (CCP), together with the Homeland Security Committee, announced a formal investigation into national security risks posed by CCP-developed AI models. The committees sent inquiry letters to US companies, including companies deploying Moonshot AI's Kimi model. Also in April 2026, several developers of frontier AI models briefed staff of the House Committee on Homeland Security regarding increasingly capable cyber-focused AI models and the risks they may pose. The committee has been examining how advanced models could be used offensively and how critical infrastructure operators should prepare.

Investigations by federal, congressional, and state officials into the practices of AI developers and deployers are expected to continue as powerful frontier AI models and autonomous AI agents become more capable and increasingly outpace human workflows. The regulatory trajectory suggests that regulators and investigators alike will focus closely on who will be liable for any harm caused by autonomous AI.

Conclusion

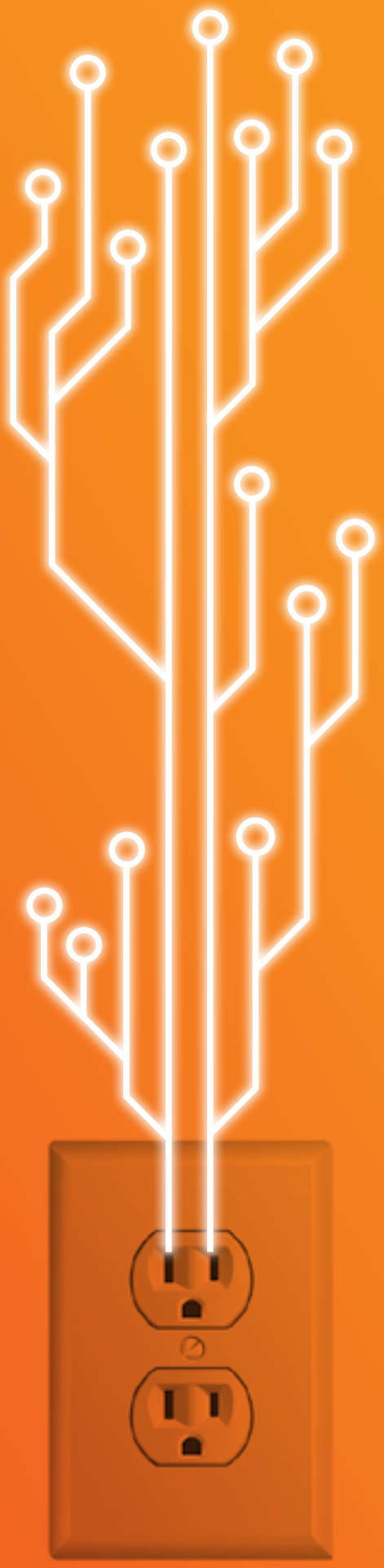
As AI systems and AI Agents are increasingly integrated into commercial and government applications, there is a growing demand to govern, manage and monitor these systems and their outputs in real time. While the concept of monitoring AI systems for legal compliance is not new, best practices for governing and managing AI risks are still evolving and require constant vigilance. Exactly because AI systems are rapidly evolving in ways that introduce new risks and unpredictability into their ecosystems, strong governance practices, AI-based vulnerability testing, risk assessments and post-deployment monitoring at machine speed are crucial practices for confident, wide-spread AI adoption.

14. "A Look At NAIC's Proposed Tool For Evaluation Of Insurer AI," *Law360* (Aug. 8, 2025).

EVERSHEDS SUTHERLAND

Powering your AI journey:

Innovation and compliance in AI



www.beaumont-capitalmarkets.co.uk
+44 (0)208 004 1404