The Bretton Woods Committee

No. 1 • OCTOBER 2025

The Bretton Woods Committee Working Group on Artificial Intelligence in Finance, co-chaired by William C. Dudley and Sarah Hammer

Governing Artificial Intelligence Across the Global Financial System

LAYING THE GROUNDWORK: STRATEGIC CONTEXT AND CORE CONCEPTS FOR AI IN FINANCE

Framing the Challenge

Generative artificial intelligence (AI) is no longer a theoretical possibility at the margins of financial services—it is rapidly becoming a core driver of transformation across the sector. From AI-driven credit decisions and algorithmic trading to infrastructure optimization, generative AI is impacting how financial institutions operate, compete, and scale. The pace of innovation is testing the limits of existing frameworks, creating a widening gap between technological capacity and regulatory readiness. This white paper series is designed to equip policy makers, industry leaders, and regulators with a strategic understanding of both the potential and the perils of this shift. It begins by establishing a shared conceptual foundation for the technologies reshaping finance and concludes by proposing oversight approaches that are not only responsive to today's use cases but also resilient to tomorrow's challenges.

A Road Map for Governing AI in Finance

AI is poised to reshape global finance. From front-line customer engagement to back-end operational infrastructure, generative AI is already changing how institutions manage risk, allocate capital, streamline operations, and serve clients. The consequences of these changes are likely to alter market behavior, reshape financial intermediation, and even redefine the architecture of financial services. To navigate this transformation effectively, policy makers, industry leaders, and technologists must share a common understanding of generative AI's benefits and its risks.

In a series of four papers, the Bretton Woods Committee will examine the applications, risks, regulatory challenges, and policy opportunities associated with generative AI in finance. This work is intended to help shape a forward-looking regulatory framework that fosters innovation while protecting consumers, ensuring market integrity, and safeguarding the stability of the financial system.

The first paper establishes foundational concepts, including the distinction between traditional machine learning and generative AI, and maps the adoption of

both of these innovations across the financial sector. It surveys current applications—from personalized financial advice and algorithmic trading to credit modeling and compliance tools—and evaluates the capabilities and vulnerabilities that make AI both transformative and potentially destabilizing. It also introduces key considerations such as explainability, governance, and systemic risk.

The second paper will explore real-world use cases in greater depth, highlighting operational benefits, business opportunities, and new sources of risk. The third will assess the regulatory landscape across major jurisdictions, identifying where existing frameworks fall short in addressing generative AI–specific challenges. The fourth and final paper will offer concrete policy proposals, including recommendations for regulatory coordination and proposed technical standards for model risk management.

These four papers follow a deliberate sequence, moving from a shared understanding of AI's role in finance to actionable proposals for managing its risks. By structuring the series in this way, we aim to equip decision makers with the clarity and insight needed to shape effective, forward-looking oversight of generative AI in the financial system.

I. CORE CONCEPTS: FROM MACHINE LEARNING TO GENERATIVE AI IN FINANCE

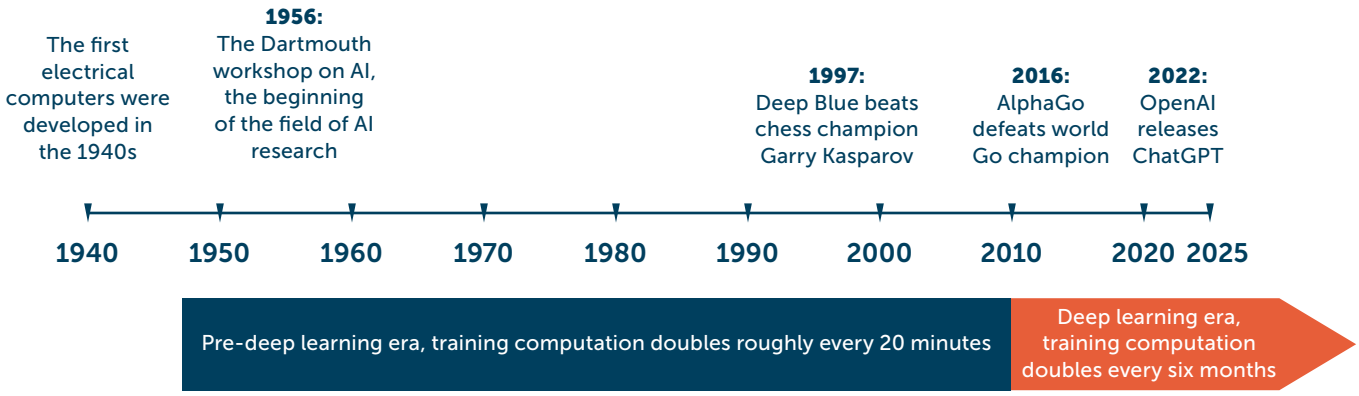
To understand the specific capabilities of and risks associated with generative AI, it is essential to distinguish it from traditional machine learning. While both fall under the broader umbrella of AI, generative models represent a meaningful evolution beyond earlier predictive systems. Rather than merely detecting patterns or making forecasts, generative AI can autonomously produce novel outputs, such as text, images, and code, based on the data on which it has been trained. This leap in functionality introduces powerful new applications but also raises complex challenges: understanding how model outputs are generated, determining how such systems should be governed, and assessing their implications for systemic risk. These distinctions are crucial for identifying where current regulatory frameworks may be lacking and where new standards or oversight mechanisms may be needed.

AI has come a long way since its conceptual origins in the mid-20th century. In 1948, British mathematician Alan Turing wrote *Intelligent Machinery*, delving into the idea of a machine that learns from experience, and drawing an important analogy to the human brain. Around the same time, neurophysiologist Warren McCulloch and logician Walter Pitts published a seminal paper proposing a model of how neurons perform logical operations. Crucially, they showed how networks of simplified, artificial neurons could, in principle, compute any function. These works looked beyond the development of "ordinary" programmed computers, which function on discrete, predefined rules, and asked whether machines modeled after the human brain—called *neural networks*—could learn by example and, in a sense, "program" themselves.

These early inquiries into the world of machine learning led to the creation of AI machines such as the stochastic neural analog reinforcement calculator (SNARC), the first neural network computer, built in 1951. SNARC used analog circuits to simulate neural processing and marked one of the first tangible efforts to translate brain-inspired theories into hardware. Over the following decades, AI matured through milestones like IBM's Deep Blue, which defeated a world chess champion, and Facebook's DeepFace facial recognition system, which advanced the state of computer vision. As these applications evolved, machine learning came to encompass a range of approaches. *Supervised learning algorithms*, for example, are trained on labeled datasets whose input and output variables are known; *unsupervised learning* enables models to infer patterns and relationships from unlabeled data, with *semi-supervised learning* straddling the two. Additionally, reinforcement learning models teach machines to adapt through trial and error. (See figure 1.)

With the advent of large language models (LLMs), AI entered a fundamentally new phase. Trained on a mix

Figure 1. AI Development to Date



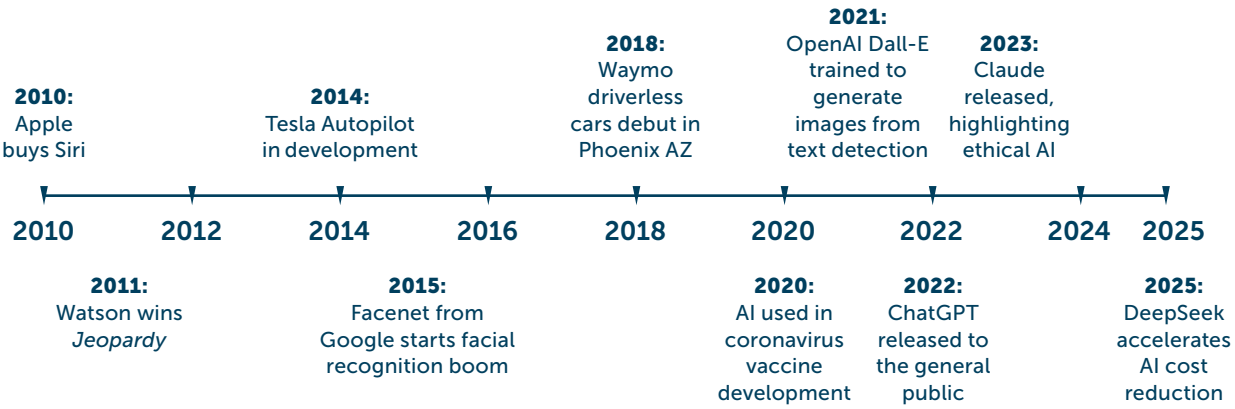
Source: Multiple AI-related charts in Giattino, Charlie, Edouard Mathieu, Veronika Samborska, and Max Roser, "Artificial Intelligence" (2023), https://ourworldindata.org/artificial-intelligence.

of supervised, unsupervised, and reinforcement learning, LLMs such as ChatGPT and DeepSeek represent a major leap in AI's evolution. Whereas traditional machine learning models rely heavily on predefined features and structured data to detect patterns and make predictions, generative AI models, including LLMs, go further. They can generate original content—text, images, code—based on what they learn from massive datasets, enabling capabilities that were previously out of reach.

What sets generative AI further apart is its ability to generalize across tasks without requiring extensive feature engineering or domain-specific customization. Unlike earlier machine learning systems that needed precise tuning for each application, LLMs are trained on broad datasets and can be efficiently adapted to specific tasks. This capability enables a wide range of financial use cases, from compliance monitoring to risk assessment, with minimal model reconfiguration. (See figure 2.)

Recognizing the distinctions between generative AI and traditional machine learning is essential to understanding how this technology transforms the foundations of the financial system. Generative AI's capacity to create content, generalize across tasks, and function with limited human oversight introduces a new class of risks—ranging from model opacity to systemic vulnerabilities—that existing regulatory frameworks may not be equipped to manage. It is because of these risks that this paper focuses specifically on generative AI: a technology whose transformative power demands focused scrutiny and forward-looking governance.

Figure 2. Developments in the Deep Learning Era



Source: Multiple AI-related charts in Giattino, Charlie, Edouard Mathieu, Veronika Samborska, and Max Roser, "Artificial Intelligence" (2023), https://ourworldindata.org/artificial-intelligence.

II. THE STRATEGIC CAPABILITIES OF GENERATIVE AI

Generative AI is not only powerful in theory—it is also uniquely capable in practice. Its ability to detect patterns, generate content, and adapt dynamically across tasks makes it a foundational technology across industries, including finance. This section highlights four core functional strengths that set generative AI apart from traditional models: content generation, innovation, adaptability, and operational optimization. A clear understanding of these capabilities is essential to assessing the opportunities, risks, and regulatory implications of generative AI's expanding role in the financial sector.

As generative AI becomes more deeply embedded in financial institutions, it brings with it a distinct set of risks. These include model opacity, bias and discrimination, "hallucinations," data privacy breaches, cybersecurity threats, intellectual property disputes, and the potential for systemic contagion across markets. While these issues will be examined in greater detail later in this paper, it is important to acknowledge from the outset that the very capabilities that make generative AI transformative may also create challenges for financial stability, market integrity, and consumer protection.

Idea Generation and Content Creation at Scale

When guided by effective prompts, generative AI can drive idea generation and content creation well beyond the limits of traditional methods. Models like DALL-E 3 translate abstract or fantastical text prompts into visual concepts, helping creatives—from writers to designers—envision and iterate their work.[1] In music, AIVA (Artificial Intelligence Virtual Artist) generates imaginative, genre-blending compositions—such as rock-samba or lo-fi jazz—and holds the distinction of being the first virtual composer recognized by the Society of Authors, Composers, and Publishers of Music.[2] In fashion, platforms like Mercer use generative AI to help users rapidly sketch and prototype new designs, lowering barriers to entry for emerging creators.

AI-Powered Discovery and Innovation

Generative AI can drive advancements by identifying new opportunities and enhancing existing processes. The health care innovation tool AlphaFold has expanded the frontier of drug discovery with its AI-powered predictive tool for protein structure (see figure 3),[3] while the model EVEscape bridges biological and evolutionary insight to predict virus evolution. Had it been deployed prior to Covid-19, researchers report, it would have anticipated the most serious mutations of SARS-CoV-2, with an accuracy similarly generalizable to HIV and influenza.[4] AI tools like Startup.ai can generate innovative startup concepts, and AI-enabled trend analysis platforms such as SparkToro or Exploding Topics can provide information about emerging market trends for entrepreneurs in search of audience intelligence.

Dynamic Learning and Adaptation Across Domains

Generative AI learns from diverse datasets and continuously adapts to dynamic, complex, real-world conditions. This adaptability extends well beyond the capabilities of chatbots, which can tailor conversational styles to mirror user behavior. In finance, AI models can respond to evolving market conditions—such as volatility, geopolitical events, or economic trends—by adjusting trading strategies in real time.[5]

1 Michael Howe-Ely, "More Examples of AI-Generated DALL-E 2 Images," *Michael Howe-Ely Blog Posts,* July 26, 2022, https://michaelhoweely.com/2022/07/26/more-examples-of-ai-generated-dall-e-2-images/.

2 Ed Lauder, "Aiva is the First AI to Officially be Recognised as a Composer," *AI Business*, March 10, 2017, https://aibusiness.com/verticals/aiva-is-the-first-ai-to-officially-be-recognised-as-a-composer.

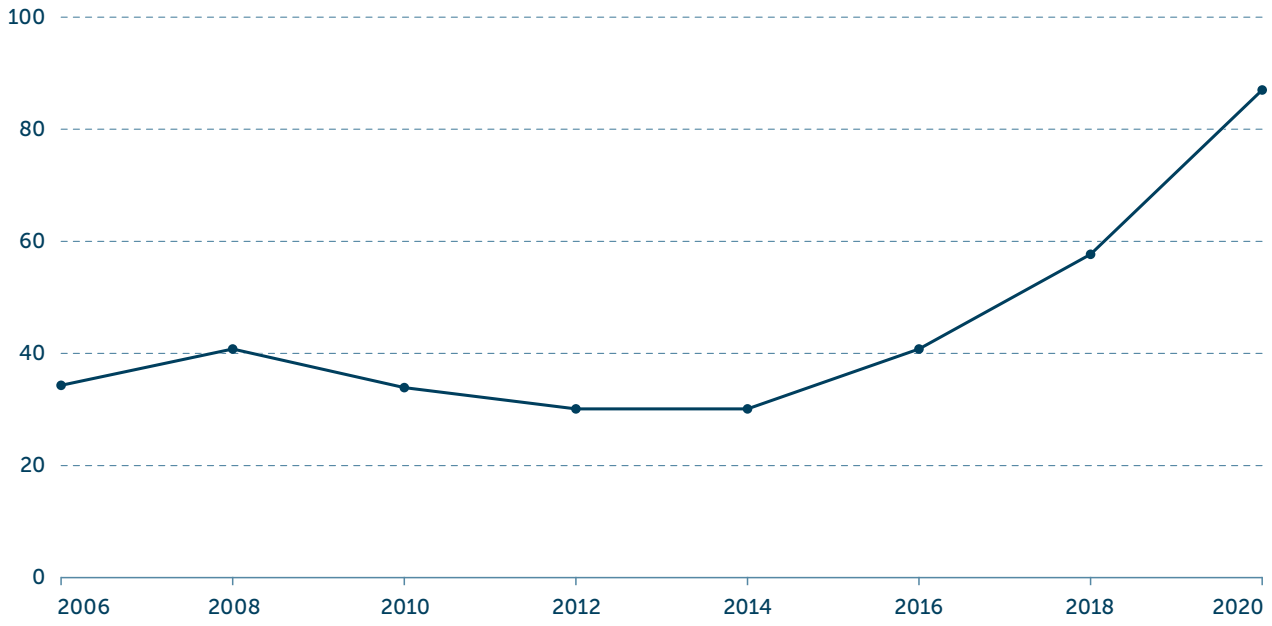
3 Ştefan-Bogdan Marcu, Sabin Tăbîrcă, and Mark Tangney, "An Overview of AlphaFold's Breakthrough," *Frontiers in Artificial Intelligence* 5 (2022): 875587, https://doi.org/10.3389/frai.2022.875587.

4 Nicole N. Thadani, Sarah Gurev, Pascal Notin, Noor Youssef, Nathan J. Rollins, Daniel Ritter, Chris Sander, Yarin Gal, and Debora S. Marks, "Learning from Prepandemic Data to Forecast Viral Escape," *Nature* 622, no. 7984 (2023): 818–25, https://doi.org/10.1038/s41586-023-06617-0.

5 Mike Smith, "Machine Learning in Trading: A Game Changer for Markets?" *GOMarkets,* February 17, 2025, https://www.gomarkets.com/en/articles/trading-strategies/machine-learning-in-trading-a-game-changer-for-markets/.

Figure 3. Protein Folding Prediction Accuracy

Median accuracy of predictions in the free modeling category for the best team in each year's Critical Assessment of Protein Structure (CASP) competition. Ranges from 0–100 (100 = best). Large advances were made in 2018 and 2020 by DeepMind's AlphaFold_AI systems.



Data source: DeepMind via AI Index Report (2021) OurWorldinData.org/artificial-intelligence | CC BY

Note: Measured using best-of-5 Global Distance Test, which scores the similarity between a predicted protein structure and the actual structure.

Source: DeepMind via AI Index Report (2021), https://ourworldindata.org/grapher/protein-folding-prediction-accuracy

In cybersecurity and fraud detection, generative AI can detect and respond to new and emerging threats, including complex scams like synthetic identity fraud.[6]

Optimization and Efficiency Gains in Operations

Generative AI can be harnessed to streamline operations, enhance decision making, and maximize resource efficiency across industries. In logistics, AI-powered systems can incorporate real-time data—such as weather, traffic, and delivery conditions—to optimize route planning and reduce delays.[7] In supply chain management, AI models can forecast inventory needs and recommend cost-saving strategies to improve coordination and responsiveness.[8] In health care, predictive tools can anticipate patient volumes and enable preemptive allocation of beds, staff, and equipment to meet shifting demand.[9]

6 Juan Fernando C. Bulgarelli, "AI Fraud Detection: Preventing Scams with Generative AI," *Veridas*, March 14, 2025, https://veridas.com/en/generative-ai-fraud/.

7 FarEye Technologies, Inc., "AI-Powered Route Planning Software for On-Time Delivery Success." *FarEye* (blog), February 12, 2025, https://fareye.com/resources/blogs/how-ai-powered-software-for-route-planning-ensures-on-time-deliveries.

8 Uche Nweje and Moyosore Taiwo, "Leveraging Artificial Intelligence for Predictive Supply Chain Management, Focus on How AI-Driven Tools are Revolutionizing Demand Forecasting and Inventory Optimization," *International Journal of Science and Research Archive* 14, no. 1 (2025): 230–50, https://doi.org/10.30574/ijsra.2025.14.1.0027.

9 Shiva Maleki Varnosfaderani and Mohamad Forouzanfar, "The Role of AI in Hospitals and Clinics: Transforming Healthcare in the 21st Century," *Bioengineering* 11, no. 4 (2024): 337, https://doi.org/10.3390/bioengineering11040337.

III. FOUNDATIONS FOR OVERSIGHT: EXPLAINABILITY, GOVERNANCE, SYSTEMIC RISK

To evaluate the implications of these capabilities for finance, three key issues are particularly important: explainability, governance, and systemic risk. These concepts form the foundation for much of the discussion and will be explored in greater depth in subsequent papers. Explainability remains a dynamic area of technical development, while governance practices and systemic risk are increasingly shaped by maturing theoretical frameworks and emerging regulatory norms. A clear understanding of these domains is critical to assessing how AI integration may alter the operations of financial institutions, influence market behavior, and challenge existing oversight mechanisms.

Model Explainability as a Foundation for Governance and Compliance

As AI becomes more deeply embedded in financial services, explainability has emerged as a critical concept for enabling effective oversight and sustaining trust. *Explainable AI* has features that allow humans to interpret and understand how it produces a given output or decision. This capacity is especially important in domains with significant consequences for individuals or markets—such as credit determinations, investment recommendations, and trading activity—where opaque systems could potentially obscure risks, undermine confidence, or impede regulatory review.

Explainability can be approached at multiple levels, each offering a distinct lens into model behavior. *Local explainability* examines how a model arrived at a specific decision for an individual case—for example, why a particular credit application was approved or denied. *Cohort explainability* analyzes model performance across defined groups or demographics to detect potential disparities or biases. *Global explainability* aims to uncover which features or inputs most significantly influence the model's overall behavior and predictions. Common explainability techniques include *feature importance analysis* (ranking input variables by their impact on outputs), *sensitivity testing* (examining how changes in inputs affect results), and *rule extraction* (deriving human-interpretable decision rules from complex models).

These approaches are not static; the field continues to evolve rapidly, with new tools and methodologies emerging to enhance transparency in increasingly complex AI systems. These advances are essential for identifying hidden risks such as discrimination, model error, and unintended outcomes. Improved explainability also supports compliance with existing legal requirements in areas such as consumer protection, lending, and capital markets, where understanding model behavior is critical to regulatory oversight.

Governance Structures for AI Systems

Effective AI deployment in finance requires strong governance frameworks. *AI governance* refers to the policies, structures, and oversight mechanisms that guide the responsible development and use of AI systems within institutions. It spans a wide range of practices—from internal protocols, such as model validation, risk management, corporate governance, and audit controls, to external collaboration with financial regulators and standard-setting bodies across jurisdictions. A coherent governance structure helps financial institutions proactively mitigate risks of AI and enable long-term business success.

Closely related is *data governance*, which addresses the management, architecture, and security of the data used to train, test, and operate AI systems. As financial institutions increasingly rely on third-party data providers, alternative datasets, and synthetic data, concerns around accuracy, privacy, and security become more pronounced. Poor data quality or faulty security methods can degrade model performance, heighten compliance risk, and produce flawed or biased outcomes. Robust data governance frameworks are therefore essential to maintaining the integrity, reliability, and accountability of AI-driven financial operations.

Systemic Risk and Financial Stability in AI-Driven Markets

The use of AI in finance has raised concerns about new and evolving forms of systemic risk. One potential source stems from model homogeneity: As institutions deploy similar AI systems trained on common or overlapping datasets, their behaviors may become highly correlated. In times of stress, these synchronized responses could amplify financial shocks rather than dampen them. In high-speed or algorithmic trading, for instance, AI models may react to stressed market signals in ways that exacerbate volatility and illiquidity.

Generative AI introduces additional vulnerabilities. Its capabilities can be exploited to create synthetic identities, generate sophisticated malware, or evade detection—raising the risk of cyber attacks with system-wide implications. When AI is embedded in critical functions such as liquidity management or risk modeling—especially within systemically important financial institutions—failures can propagate rapidly across interconnected markets. These challenges are further compounded by the complexity and opacity of many AI systems, which could make early warning signs harder to detect. Addressing these risks will require a deeper understanding of model behavior at scale, along with greatly enhanced regulatory and supervisory capabilities. The concept of systemic risk will be discussed in greater detail later in this series.

IV. OPPORTUNITIES AND CHALLENGES IN AI-DRIVEN FINANCE

AI is rapidly expanding its footprint across the financial sector, redefining how firms operate, compete, and deliver value. From client engagement to core infrastructure, its integration spans a broad and evolving set of use cases, each carrying distinct strategic implications. The examples that follow provide a representative—though not comprehensive—snapshot of current applications. As adoption accelerates, financial institutions must navigate a complex landscape of opportunity and risk, calibrating their approaches to the unique demands of each domain.

The next paper in this series will examine selected applications in greater depth, focusing on specific implementations and technical architectures across financial institutions. Adoption patterns differ markedly by firm, regulatory environment, and business model. What cuts across these variations, however, is a shared potential: to fundamentally reconfigure some capabilities, reshape cost structures, and alter the competitive landscape of global finance.

Client-Facing Applications: Financial Advice and Customer Interaction

Within client-facing services, AI is fundamentally altering the provision of financial advice and customer interaction. Machine learning algorithms analyze clients' transaction histories, investment preferences, and risk tolerances to generate tailored investment recommendations. Financial institutions are using AI to provide personalization at scale, allowing them to deliver customized recommendations that adapt to a client's specific financial objectives.

Natural language processing (NLP) also enables conversational interfaces, which make it capable of interpreting complex financial queries and delivering contextually relevant guidance. These systems can assist with summarizing documents, drafting communications, and answering client questions with increasing sophistication. The democratization of financial expertise through AI represents a profound shift in client service delivery, enabling institutions to maintain continuous engagement with clients through digital channels while reserving human intervention for complex scenarios that require more nuanced judgment.

AI in Credit Assessment, Pricing, and Monitoring

In institutional lending, AI systems can analyze vast datasets encompassing financial statements, market conditions, industry trends, and alternative data sources

to assess creditworthiness with greater nuance than traditional approaches. These models identify subtle relationships between variables that might escape conventional analysis, potentially improving lending decisions. For corporate borrowers, AI models can process unstructured data from regulatory filings, news sources, and industry reports to provide a more comprehensive risk assessment than financial statements alone could offer.

The application of AI in retail lending presents both extraordinary opportunities and significant challenges. The information asymmetry inherent in consumer lending—where borrowers possess substantially more information about their circumstances and intentions than do lenders—has traditionally been addressed through standardized underwriting criteria that inevitably exclude some creditworthy individuals. AI models can potentially reduce this asymmetry by incorporating alternative data sources and identifying patterns indicative of repayment capacity, thereby enabling greater credit access. However, these same capabilities could potentially reinforce or even exacerbate existing problems with respect to bias and discrimination, as the later section on risk dimensions will explain.

AI in Trading, Investment Analysis, and Portfolio Optimization

Algorithmic trading systems leverage deep learning to identify subtle market patterns across multiple asset classes and time horizons, executing complex strategies with precision unattainable through traditional approaches. As noted in a March 2025 SEC roundtable on AI in finance, BlackRock employs AI for "algorithmic pricing" to improve liquidity and reduce costs, while also utilizing the technology as an "investment process augmentation tool" that achieves superior trading strategies and assists with best execution.[10]

These applications enhance not only trading performance but also transaction cost analysis, enabling more efficient capital deployment. In investment analysis, NLP extracts actionable intelligence from earnings calls, regulatory filings, and news feeds, quantifying sentiment and identifying material disclosures that may impact valuations. These capabilities enable institutions to process volumes and types of information that would overwhelm human analysts, generating insights that enhance investment performance and risk management simultaneously.

Enterprise Applications: Operations, Risk Management, and Compliance

Enterprise applications represent some of the most powerful and transformative uses of AI in financial services. JP Morgan, for instance, uses "COiN" (contract intelligence) to streamline document review, improving both efficiency and compliance outcomes.[11] In back-office operations, financial institutions deploy AI solutions for tasks such as reconciliations—processes that have traditionally required substantial manual effort. While less visible than client-facing innovations, these applications can deliver significant operational gains and risk management enhancements, allowing firms to reallocate resources from routine administrative tasks to higher-value activities that require human judgment.

In the domain of risk management, AI-powered capital and liquidity optimization models use advanced forecasting techniques to anticipate cash flow needs across global operations. These capabilities help institutions meet regulatory requirements while minimizing idle capital and improving reserve allocation.

AI also holds significant promise for improving regulatory compliance. Machine learning algorithms can parse complex rule books, extract specific obligations, and automate reporting processes that previously demanded extensive human input. Generative AI goes a step further by enabling large-scale review of unstructured data—such as contracts, communications, and regulatory filings—to detect red flags, monitor for conduct risk, and assist in drafting disclosure

10 U.S. Securities and Exchange Commission, "SEC Roundtable on Artificial Intelligence in the Financial Industry," March 27, 2025, https://www.sec.gov/newsroom/meetings-events/sec-roundtable-artificial-intelligence-financial-industry.

11 Don Galeon, "An AI Completed 360,000 Hours of Finance Work in Just Seconds," *Futurism* (blog), March 8, 2017, https://futurism.com/an-ai-completed-360000-hours-of-finance-work-in-just-seconds.

language. Its ability to adapt to evolving legal requirements and enforcement trends enhances the scalability and responsiveness of compliance functions while reducing cost and resource burdens.

AI in Financial Infrastructure: Trading, Clearing, and Settlement

AI is increasingly reshaping the architecture of exchange trading, with major market operators employing advanced models to refine market structure, enhance liquidity, and improve price discovery. NASDAQ's deployment of AI across its infrastructure exemplifies this transformation. Since 2018, NASDAQ has utilized AI in its index business to create "minimum volatility portfolios" and define technology exposures using patent data, enabling the development of more sophisticated index products. Additionally, since 2022, the exchange has utilized AI to optimize the listing of option strikes, addressing the challenge that most options contracts trade minimally or not at all. This application helps market makers more efficiently cover the market by aligning listed options with anticipated demand.

Perhaps most notably, NASDAQ's Dynamic M-ELO system, an expansion of a midpoint extended life order, employs reinforcement learning to make market structure decisions.[12] This AI system determines how long certain orders must wait before trading, optimizing outcomes for institutional investors. Applications like this one demonstrate how AI can enhance market structure and price formation processes, benefiting both market operators and participants.

AI is also increasing its role in clearing and settlement processes—the critical but often overlooked task of achieving transaction finality in trading. AI models analyze historical settlement patterns to identify potential failures before they occur, allowing for preemptive intervention. These systems optimize collateral management by predicting liquidity needs and dynamically allocating assets accordingly, reducing funding costs while maintaining system integrity. As transaction volumes increase and settlement time frames compress, AI's capacity to manage this complexity becomes increasingly valuable, enhancing both operational efficiency and financial stability by reducing settlement risk. These infrastructure applications, while less visible than retail-facing innovations, can play a crucial role in supporting market integrity and reducing systemic risk. (See table 1.)

Table 1. Selected AI Applications and Risk Considerations in Financial Services

Use Case Category	Examples of Applications	Selected Risk Considerations
Client Facing	Personalized investment recommendations, conversational interfaces	Regulatory oversight, fiduciary duty, customer satisfaction, accuracy
Credit Assessment	Corporate lending analysis, retail lending with alternative data	Model interpretability, regulatory compliance, bias, discrimination, fair lending
Trading and Investment	Algorithmic trading systems, investment analysis via NLP	Market stability, systemic risk, data quality, model reliability
Enterprise Operations	Back-office automation, risk management optimization, compliance automation	Operational efficiency, cost reduction, model accuracy, regulatory requirements, legal liability, regulatory interpretation
Financial Infrastructure	Market structure optimization, wholesale clearing and settlement	Systemic impact, market integrity, operational risk, settlement failures

12 U.S. Securities and Exchange Commission, "SEC Roundtable on Artificial Intelligence in the Financial Industry," March 27, 2025, https://www.sec.gov/newsroom/meetings-events/sec-roundtable-artificial-intelligence-financial-industry.

V. KEY RISK DIMENSIONS IN AI-DRIVEN FINANCE

As AI becomes increasingly integrated across financial institutions and markets, it introduces potential new risks that span technical, operational, legal, and systemic dimensions. While AI offers significant potential for efficiency and insight, it also raises foundational questions about the reliability, security, and stability of the financial system. This section examines the most pressing risk dimensions that must be addressed to ensure AI is deployed not only ambitiously but also safely and responsibly across global finance. Some of these risks, along with their practical implications, will be explored in greater detail in the next paper in this series.

Bias and Discrimination

Generative AI models can potentially amplify bias based on flaws in the training data. These risks become especially pernicious when AI is used in high-risk decisions involving individuals, such as criminal prosecution or social scoring. Several high-profile cases highlight this concern: the COMPAS recidivism algorithm, which disproportionately mislabeled Black defendants as high risk compared with white defendants,[13] and Amazon's AI recruiting tool, which was found to penalize the word "women" when screening résumés.[14] In the financial sector, bias could distort the outputs of credit scoring algorithms or result in skewed insurance premiums. When AI is relied upon to set pricing, a biased model may unintentionally cause individuals from certain demographic groups to receive less favorable loan terms.

Privacy

Safeguarding sensitive data and ensuring confidentiality present growing challenges in AI-driven environments. The past decade of internet commercialization has already exposed the stark information asymmetry between the institutions that collect data and the individuals who generate data.[15] AI may further widen this gap and magnify associated risks—the stakes of unauthorized data access, security breaches, and misuse of personal information all rise in an era of unprecedented information sharing. Data privacy is especially critical in finance, where firms routinely collect large volumes of personally identifiable information as part of customer onboarding, credit assessment, and transaction monitoring. Additionally, LLMs may introduce a novel form of privacy risk through inference—accurately guessing personal information about users based on seemingly unrelated interactions.[16]

Hallucinations

An AI *hallucination* is an instance in which a model generates inaccurate or misleading information, which can undermine decision making. Such an error can arise from a variety of sources, including poor prompting, flaws in model architecture, distorted input data, or overfitting—a condition in which a model adheres too closely to its training data and fails to generalize effectively to new information. In high-performance data analytics platforms powered by AI, such hallucinations can directly impact trading decisions and investment outcomes. The consequences may include legal and compliance violations, as well as significant erosion of customer trust.

Intellectual Property

A related area of concern involves the ownership, rights, and potential misuse of AI-generated content. These issues have gained prominence in sectors such as entertainment and publishing, where the use of online intellectual property for AI training has

13 Jeff Larson, Surya Mattu, Lauren Kirchner, and Julia Angwin, "How We Analyzed the COMPAS Recidivism Algorithm," *ProPublica*, May 23, 2016, https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

14 Jeffrey Dastin, "Insight - Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women," *Reuters*, October 10, 2018, https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/.

15 Alex Campolo, Madelyn Sanfilippo, Meredith Whittaker, and Kate Crawford, "AI Now 2017 Report," *AINOW*, https://ainowinstitute.org/publications/ai-now-2017-report-2, p. 28.

16 Robin Staab, Mark Vero, Mislav Balunović, and Martin Vechev, "Beyond Memorization: Violating Privacy Via Inference with Large Language Models," (preprint), arXiv: 2310.07298, https://arxiv.org/abs/2310.07298.

sparked accusations of unauthorized appropriation.[17] AI-enabled finance may be complicit in this dynamic, albeit unintentionally. Financial institutions often draw on copyright-protected research and market analyses to generate forecasts and inform allocation decisions—uses that could raise similar concerns around unauthorized use of data for training.

On the output side, questions of what is referred to as "AI inventorship" further complicate the landscape: Should ownership of AI-generated content reside with the end user, the public domain, or the model itself? The absence of international alignment on these questions—and the fragmented legal approaches adopted by different jurisdictions—pose significant challenges to developing a coherent framework for managing the intellectual property status of AI-generated outputs.[18]

Data Management and Protection

Risks associated with AI in finance extend deeply into the architecture, integrity, and security of the data on which these systems rely. Financial institutions handle vast volumes of sensitive information—from transaction histories to personal financial profiles—that demand rigorous data governance. AI systems often amplify, rather than correct, deficiencies in training data, making data quality a critical concern. These challenges intensify when traditional financial data are combined with data from alternative or third-party sources, complicating data lineage and increasing the risk of inconsistency or error.

Data security is equally vital: Institutions must ensure that access to information is tightly controlled, limited to authorized users, and revoked promptly when no longer required. Weaknesses in data management can contribute to *model drift*—a gradual deterioration in AI performance as new data diverge from training inputs, potentially triggering downstream errors across systems. The use of *synthetic data* (artificially generated data that mimic real-world data), while valuable for stress testing and modeling rare scenarios, introduces added complexity related to data validation and representativeness. Without robust data governance, even the most advanced AI systems may rest on unstable foundations, creating hidden vulnerabilities that undermine their reliability and resilience.

Deepfakes and Criminal Exploitation

The misuse of generative AI to create deceptive content and enable fraud poses acute risks to the financial system. Sophisticated deepfake technologies can now convincingly impersonate executives or clients in video conferences and voice calls, undermining traditional authentication protocols. Beyond direct fraud, generative AI enables more effective phishing campaigns targeting financial employees and customers (see figure 4). It also supports the creation of synthetic identities that blend fabricated and real personal information, complicating know your customer procedures. These capabilities mark a fundamental shift in the fraud landscape: Whereas technical sophistication once required specialized expertise, it is now increasingly accessible to a broader range of malicious actors through user-friendly generative AI tools.

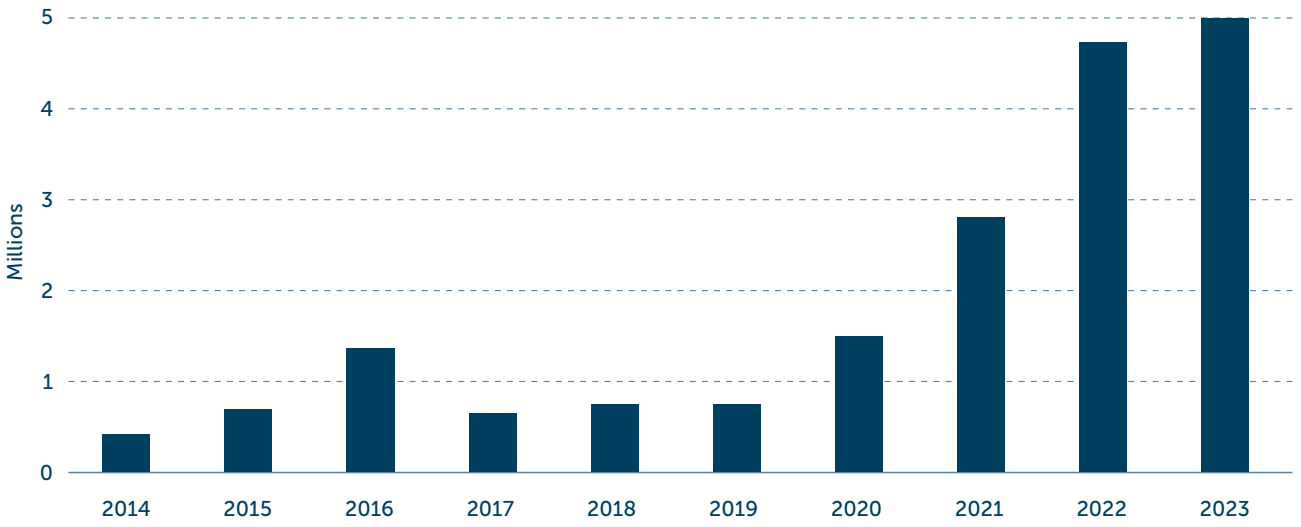
Cybersecurity and Bad Actors

Vulnerability to AI-driven cyber threats presents a growing challenge for financial institutions. While AI enhances defensive capabilities through anomaly detection and behavioral analysis, it also equips malicious actors with unprecedented offensive tools. Adversarial machine learning techniques allow attackers to probe and exploit weaknesses in AI-based security systems, potentially bypassing fraud detection mechanisms or algorithmic trading safeguards. AI can also automate the discovery of software vulnerabilities at scale, with research showing that advanced language

17 Dan Pontefract, "Authors Challenge Meta's Use of Their Books for Training AI," *Forbes,* March 25, 2025, https://www.forbes.com/sites/danpontefract/2025/03/25/authors-challenge-metas-use-of-their-books-for-training-ai/.

18 Adnan Masood, "Intellectual Property Rights and AI-Generated Content—Issues in Human Authorship, Fair Use Doctrine, and Output Liability," *Medium*, April 4, 2025. https://medium.com/@adnanmasood/intellectual-property-rights-and-ai-generated-content-issues-in-human-authorship-fair-use-8c7ec9d6fdc3.

Figure 4. Total Number of Unique Phishing Websites (Attacks) Detected



Source: APWG, MIT Technology Review and ECB calculations. https://www.ecb.europa.eu/press/financial-stability-publications/fsr/special/html/ecb.fsrart202405_02~58c3ce5246.en.html

models can identify flaws in code more efficiently than traditional scanning tools.

Financial institutions are particularly exposed to adversarial attacks through their expansive application programming interface ecosystems, but generative AI can analyze the behavior of such interfaces to uncover exploitable patterns. Still, the cybersecurity threat extends to the models themselves: techniques such as model inversion attacks—whereby adversaries attempt to reconstruct sensitive training data by analyzing a model's outputs—can compromise confidential information. This evolving arms race demands the implementation of advanced countermeasures, including adversarial testing protocols and continuous model surveillance.

Systemic Risk

As noted in the earlier discussion of systemic risk, AI-related failures have the potential to destabilize the broader financial system—a concern that will be explored in greater depth in subsequent papers. One possible risk arises from model convergence, whereby many institutions adopt similar AI models trained on overlapping datasets. This alignment may increase the likelihood that models will behave similarly under stress—for example, by simultaneously triggering asset sales or liquidity withdrawals. Such synchronized responses can create feedback loops, in which AI-driven actions intensify the very market signals that prompted them, compounding volatility and accelerating destabilization. The risk is further heightened by concentration among a small number of AI technology providers, which could enable flaws or failures in widely used models to propagate across the system through multiple channels of exposure.

Another systemic vulnerability lies in how AI models respond to abrupt shifts in the economic environment. Events such as the Covid-19 pandemic or the sudden imposition of trade restrictions can render historical data—and the assumptions embedded in them—obsolete. Models trained on previous regimes may fail to adapt quickly or accurately, leading to flawed outputs across institutions. When these systems act in parallel on outdated assumptions, they risk amplifying market dislocation rather than absorbing it.

Taken together, these dynamics underscore the need for further study, resilient AI governance, and forward-looking oversight as AI becomes more deeply embedded in the financial system.

VI. CONCLUSION: A STRATEGIC PATH FORWARD FOR AI IN FINANCE

This first paper lays the foundation for understanding the transformative role of AI in the global financial system. It has traced the evolution from traditional machine learning to generative AI, surveyed the technology's expanding capabilities, and introduced key concepts—*explainability*, *governance*, and *systemic risk*—that will anchor the discussion throughout this series. As AI becomes more deeply integrated into financial institutions and markets, its potential to reshape business models, market dynamics, and regulatory priorities will only intensify. The need for coordinated, forward-looking frameworks is growing, not just to manage emerging risks but also to guide innovation toward resilient and trustworthy outcomes.

The next paper will examine how AI is currently being deployed across the financial sector, from customer engagement and credit assessment to trading, compliance, and operational infrastructure. It will analyze specific use cases and technical architectures, highlighting both the value these tools unlock and the new forms of complexity they introduce. The third and fourth papers in the series will build on that analysis, turning to the global regulatory landscape and offering practical policy recommendations to support responsible innovation at scale. The series is designed to equip decision makers with the insight needed to govern AI in finance thoughtfully and effectively—balancing innovation with integrity in an increasingly algorithmic future.

The author would like to thank Allyson Ye and Swarith Reddy for their assistance in completing this paper.

Artificial Intelligence Working Group

CO-CHAIRS: William C. Dudley and Sarah Hammer

WORKING GROUP: Ghassan K. Abou-Alfa, Sandeep Arora, Beverly Barrett, Daniela Bassan, Richard Berner, Barbara Boehler, Josh Burek, F. Christopher Calabia, Jake Carney, Keith B. Carter, Uldis Cerps, David W. Cervantes, Himamauli Das, Mona Elliot, Anthony Elson, Omer Fein, Pilar Cristina Frank-O'Leary, Gerald Gallagher, Nitin Gaur, Samim Ghamami, Daniel L. Gleizer, Jeffrey Greenbaum, Michael B. Greenwald, Naureen Hassan, Ulrike Hoffman-Burchardi, Michael Hsu, Linda Jeng, Gregory C. Johnson, Mahesh Kotecha, Pat LaVecchia, Ricardo Mora, Fabio Natalucci, Andrzej Nowak, Maximilien Pucci-Sisti Maisonrouge, David Rogovic, Gursimran Rooprai, Joan Rosas Xicota, Sunil Sabharwal, Paul R. Samson, Jason Schenker, Lemma W. Senbet, Deepika Sharma, Michael F. Silva, Andrew Slack, Keerthika M. Subramanian, Rajesh Swaminathan, Zhenqian Tay, Raffaella A. Tenconi, Hilary F. Till, Siddharth Tiwari, Marsha Vande Berg, Adam Wang-Levine, David Webster, Antonio Weiss, Benjamin S. Weiss, Rex Wempen, Carolyn A. Wilkins, David P. Willard, Linda Yueh

THE BRETTON WOODS COMMITTEE
1701 K St NW #950, Washington, DC 20006
www.brettonwoods.org