

3. Artificial intelligence challenges for the financial system

By Claire Brousse, Olivier Fliche, Jules Joyez and Julien Uri

Artificial intelligence technologies offer numerous opportunities for the financial sector, with applications including customer service, process optimization, risk management, and market activities. However, due to its nature, artificial intelligence presents challenges in data management, modelling, and governance. It could exacerbate financial stability risks, particularly procyclicality, market volatility, market concentration, and cyberattacks. Moreover, the potential for misuse, such as information manipulation, could destabilize the financial system. The potential impacts of artificial intelligence remain difficult to quantify as they will depend on the scale and modalities of its adoption. Nevertheless, a governance framework and risk management approach tailored to artificial intelligence-specific challenges are necessary to mitigate risks. The European artificial intelligence Act, adopted in May 2024, is specifically designed to promote trustworthy artificial intelligence.

3.1. The development of artificial intelligence presents opportunities for the financial system and the economy

Artificial intelligence technologies have undergone a remarkable acceleration in recent years.

Artificial intelligence encompasses a wide array of technologies designed to create systems capable of mimicking human cognitive functions. The development of these technologies is underpinned by various scientific disciplines, including statistics, data analytics, computer science, and cognitive science. The European Union's AI Act¹¹⁸ defines artificial intelligence as an automated system designed to operate with varying degrees of autonomy, capable of adapting after deployment, and able to infer from input data how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

The origins of artificial intelligence date back to the 1940s and 1960s¹¹⁹ with the creation of a first mathematical model of a neural network in 1943¹²⁰ and the publication of an article by Alan Turing in 1950 questioning for the first time a potential intelligence of machines, in which he imagines the Turing test¹²¹. The development of artificial intelligence has evolved with technological developments over the following decades and has experienced a marked acceleration in the field of machine learning, with deep learning technologies¹²². The increase in computing power and the acceleration of research on the subject have made it possible to develop artificial intelligence techniques with increasingly sophisticated technologies and an increasing number of applications.

Generative artificial intelligence has emerged as the most widely recognized branch of artificial intelligence among the general public. Leveraging deep learning-based machine learning algorithms, it produces a variety of new content formats, including text, images, video, and audio. In November 2022, OpenAI, an artificial intelligence research firm, unveiled ChatGPT, a prototype chatbot employing generative artificial intelligence based on large language models to generate text. Trained on both publicly available data and user inputs, ChatGPT garnered unprecedented attention, attracting one million users within just five days—an unparalleled rate of adoption in the history of web applications.

¹¹⁸ [AI Act | Shaping Europe's digital future \(europa.eu\)](#)

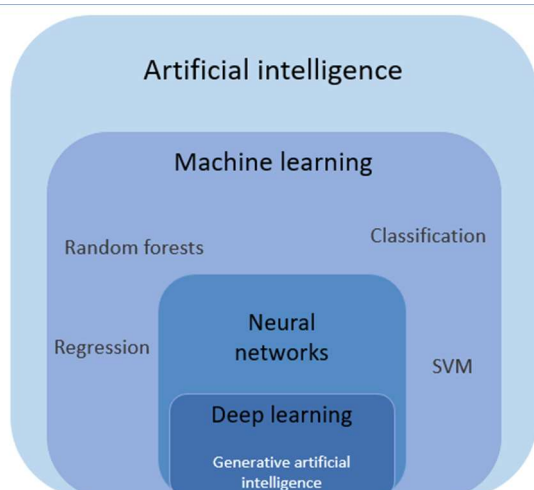
¹¹⁹ [History of \[?\]\[?\]](#)

¹²⁰ McCulloch, Warren S. Pitts, Walter. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 1943, vol. 5, p. 115-133.

¹²¹ The Turing test judges whether a machine can imitate a human conversation such that it is indistinguishable from a human. Turing, Alan M. *Computing machinery and intelligence*.

¹²² Machine learning is a subset of AI that concentrates on developing computer programs designed to learn as they are developed and used. Deep learning is a branch of machine learning that employs neural networks learning techniques.

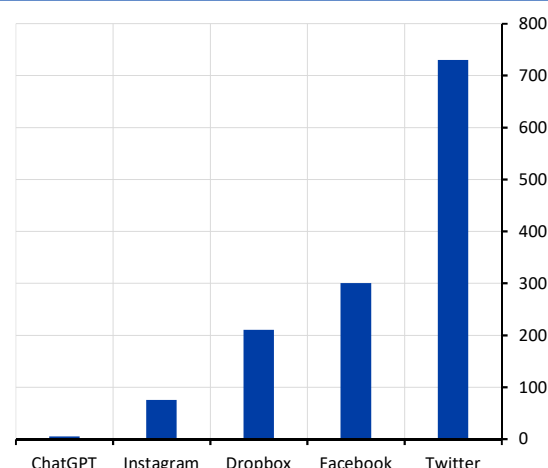
Chart 3.1: Artificial intelligence fields of study



Source: [left]: Banque de France; [right]: SimilarWeb.

Note: ChatGPT was launched online on 30 November 2022.

Chart 3.2: Number of days for a service to reach 1 million users since its launch



Artificial intelligence has already found multiple applications across various industries

In 2022, a survey of over 1,400 international companies revealed that 50% had adopted artificial intelligence technologies (generative or otherwise) in their operations (see Chart 3.3). This survey¹²³ indicates that while artificial intelligence adoption doubled between 2017 and 2018, it has remained relatively stable in recent years, suggesting that recent artificial intelligence development has been primarily driven by increased applications among existing artificial intelligence adopters. In France, according to a recent survey¹²⁴, 35% of companies report using artificial intelligence technologies¹²⁵.

These technologies are primarily used to optimize service operations, develop new products, and manage finances. Artificial intelligence is particularly used in the manufacturing sector for predictive inventory management, predictive maintenance of industrial equipment, quality control inspections, and waste management. Artificial intelligence also benefits the agricultural sector through refined weather forecasting to anticipate events that may impact harvests. In the service sector, artificial intelligence enables improvements in customer service (personalized recommendations), dynamic pricing (tourism and transportation), and cybersecurity (system monitoring, analysis of vulnerabilities and threats in real time). Finally, artificial intelligence is also used in healthcare for decision support and computer-assisted surgery.

Generative artificial intelligence is Accelerating Technological Advancement and Expanding artificial intelligence Applications. Generative artificial intelligence is driving a technological surge that is broadening the spectrum of artificial intelligence applications. By enabling the rapid and simple generation of content, generative artificial intelligence can facilitate the automation of various tasks, such as customer service through chatbot integration. This technology has already captivated businesses (see Chart 3.4). While most companies currently employing generative artificial intelligence have limited deployments, they anticipate a significant expansion of its applications in the near future¹²⁶. Key obstacles to broader generative artificial intelligence adoption include data privacy concerns, trust issues, transparency challenges, and a shortage of skilled professionals¹²⁷.

AI has the potential to exert a profound influence on the economy, though the extent and nature of this impact remain uncertain. By streamlining the production of goods and services, artificial intelligence could temporarily

¹²³ The state of AI in 2022—and a half decade in review | McKinsey

¹²⁴ “Les employeurs face à l’Intelligence Artificielle” – a survey of a sample of 3,000 French companies with ten or more employees about their use of AI, June 2023 – France Travail.

¹²⁵ Structural analysis of France 2023 (note d23-124).

¹²⁶ Generative AI: Differentiating disruptors from the disrupted | MIT Technology Review

¹²⁷ Data Suggests Growth in Enterprise Adoption of AI is Due to Widespread Deployment by Early Adopters (ibm.com)

boost productivity growth. A 2023 study of US customer support agents revealed a 25% average increase in worker productivity five months after implementing generative artificial intelligence, with the most significant benefits accruing to less experienced and lower-skilled employees¹²⁸. Additionally, artificial intelligence could foster sustained productivity growth by automating idea generation. However, measuring these macroeconomic effects is challenging, as their magnitude hinges on the policies enacted¹²⁹. A March 2024 report by Anne Bouverot and Philippe Aghion estimates that artificial intelligence -driven productivity gains in France could augment GDP by €250 billion to €420 billion¹³⁰. An IMF article¹³¹ from December 2023 suggests that growth gains might be limited if artificial intelligence displaces human workers, but they could be substantial if artificial intelligence development complements human capabilities.

Chart 3.3: Use of artificial intelligence by international companies in 2022

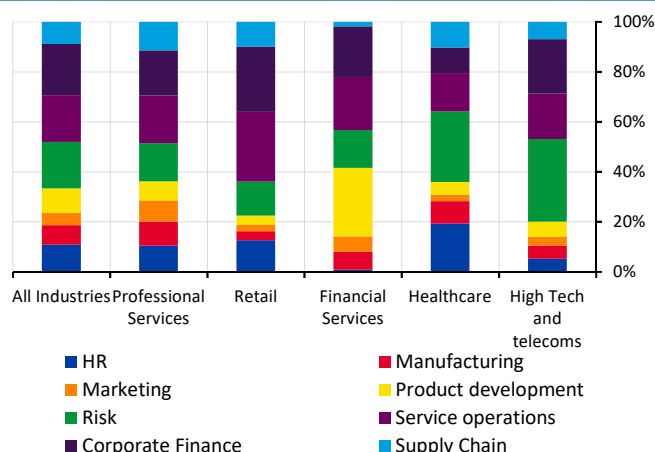
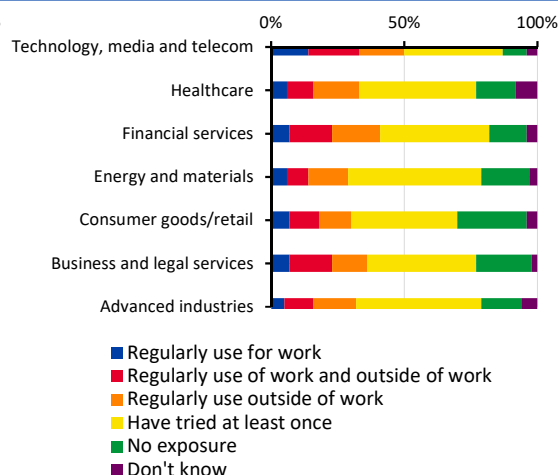


Chart 3.4: Exposure to generative artificial intelligence tools in international companies



Sources: [left]: 2022 McKinsey Survey data; [right]: 2023 McKinsey data.

The financial sector has been a pioneer in adopting artificial intelligence technologies and is poised for even greater utilization in the coming years.

The financial sector has made considerable use of artificial intelligence for many years. The integration of artificial intelligence into financial services dates back to the 1980s with the introduction of artificial intelligence -powered tools for personalized financial and tax advice. Since then, financial institutions have gradually incorporated artificial intelligence into their market activities, such as forecasting, financial analysis, and decision-making.

Banks and insurance companies, have extensively leveraged artificial intelligence. Four primary use cases, each with its associated benefits, can be identified¹³²:

- **Enhanced customer experience and satisfaction:** Through various support systems, decision support tools, and recommendation engines, including chatbots and voicebots, artificial intelligence enables personalized customer experiences. Generative artificial intelligence further enhances customization capabilities, improving the efficiency of marketing and sales processes.
- **Optimized value chain:** artificial intelligence automates previously manual tasks, optimizing internal processes and reducing costs. Its ability to process unstructured data amplifies these benefits. A prime

¹²⁸ Brynjolfsson E., Li D., Raymond L (2023), "Generative AI at Work", *NBER Working Paper Series*.

¹²⁹ Trésor-Éco No. 341 (April 2024), "Les enjeux économiques de l'intelligence artificielle" (economie.gouv.fr)

¹³⁰ Commission de l'intelligence artificielle, *25 recommandations pour l'IA en France*, report coordinated by Anne Bouverot and Philippe Aghion, 13 March 2024.

¹³¹ Acemoglu (D.) and Johnson (S.), *Rebalancing AI*, December 2023.

¹³² See for example two ACPR studies on the digital transformation of France's financial sector: *La transformation numérique dans le secteur bancaire français, Analyses et Synthèses* No. 131, January 2022; *La transformation numérique dans le secteur français de l'assurance, Analyses et Synthèses* No. 132, January 2022.

example in banking is the rise of customer self-service, allowing branch advisors to focus on higher-value activities.

- Risk management: artificial intelligence is used to manage financial risks (e.g., credit risk), operational risks, and compliance risks. It is particularly effective in combating fraud, anti-money laundering, and counter-terrorism financing (AML/CTF) through systems that monitor suspicious transactions and automate alert management.
- Market activities: artificial intelligence is instrumental in improving trend identification and integrating vast volumes of data. For instance, market participants employ machine learning for algorithmic trading to optimize execution strategies based on market conditions. Asset managers also utilize these technologies for portfolio construction and asset valuation. However, it is important to note that some financial institutions have made misleading claims to their clients regarding their use of artificial intelligence, leading to regulatory actions by authorities such as the Securities Exchange Commission¹³³.

In practice, many use cases combine some or all of these characteristics. For instance, in the banking sector, machine learning systems can assess customer creditworthiness based on their financial data (generating credit scores similar to the U.S. credit score). This leads to reduced loan processing times (improving customer experience and distribution process efficiency) through highly automated processes (optimizing internal operations) while enhancing default prediction through better modelling (reducing risks)¹³⁴. Similarly, in insurance, artificial intelligence systems are used to personalize pricing based on customer profiles, within regulatory limits, and to reduce claim processing time and costs through automated image analysis and advanced report reading capabilities. A survey conducted by Moody's Analytics in November 2023 revealed that 30% of financial institutions surveyed are currently piloting or actively using artificial intelligence for risk management and compliance activities¹³⁵.

The rise of generative artificial intelligence is poised to unfold a spectrum of applications in the financial sector. While traditional artificial intelligence is already widely used by financial institutions, the deployment of generative artificial intelligence within this industry appears limited and primarily focused on internal process improvements (risk identification, code development, document generation and synthesis, fraud and money laundering prevention). A December 2023 survey¹³⁶ indicates that a vast majority of financial institutions anticipate a surge in generative artificial intelligence applications within three years, with significant implications for their business. Broader development could extend to customer-facing areas.¹³⁷ For example, unlike current -powered chatbots that provide predefined responses, generative artificial intelligence can offer personalized information to users, such as their account balance, recent transactions, or tailored savings product recommendations. Robo-advisors, online platforms that provide financial advice and automated portfolio management, represent an example of "classic" artificial intelligence applications that could be augmented with generative artificial intelligence to enhance customer data collection, and offer personalized product recommendations and advice in natural language¹³⁸. The development of these customer-facing applications necessitates addressing transparency and accountability concerns.

Artificial intelligence-based technologies can facilitate financial system supervision (a concept known as "SupTech"). Artificial intelligence offers opportunities to improve the efficiency of supervisors in carrying out their missions, including risk control solutions¹³⁹. The Banque de France and the Autorité de contrôle prudentiel et de résolution (ACPR) have developed several projects in this area (see box). Similarly, the Monetary Authority of

¹³³ [SEC.gov | SEC Charges Two Investment Advisers with Making False and Misleading Statements About Their Use of Artificial Intelligence](#) Delphia and Global Predictions told customers and potential customers that they were using AI, when in fact they were not. The SEC noted that these firms claimed to be using an AI model, but were not in fact doing so and had therefore misled their customers.

¹³⁴ [Artificial intelligence and machine learning in financial services - FSB 2017](#)

¹³⁵ [AI in compliance \(moody's.com\)](#)

¹³⁶ [2023 IIF-EY Survey Report on AI ML Use in Financial Services - Public Report - Final.pdf](#)

¹³⁷ [Generative Artificial Intelligence in Finance - December 2023 \(oecd-ilibrary.org\)](#)

¹³⁸ Robo-advisors operate according to systems of rules, proposing asset allocations based on the information provided by customers or using complex algorithms that may be employed to monitor portfolios continually and adjust them dynamically. See in particular: [Robo-Advising: Less AI and more XAI? - 2021 \(institutlouisbachelier.org\)](#)

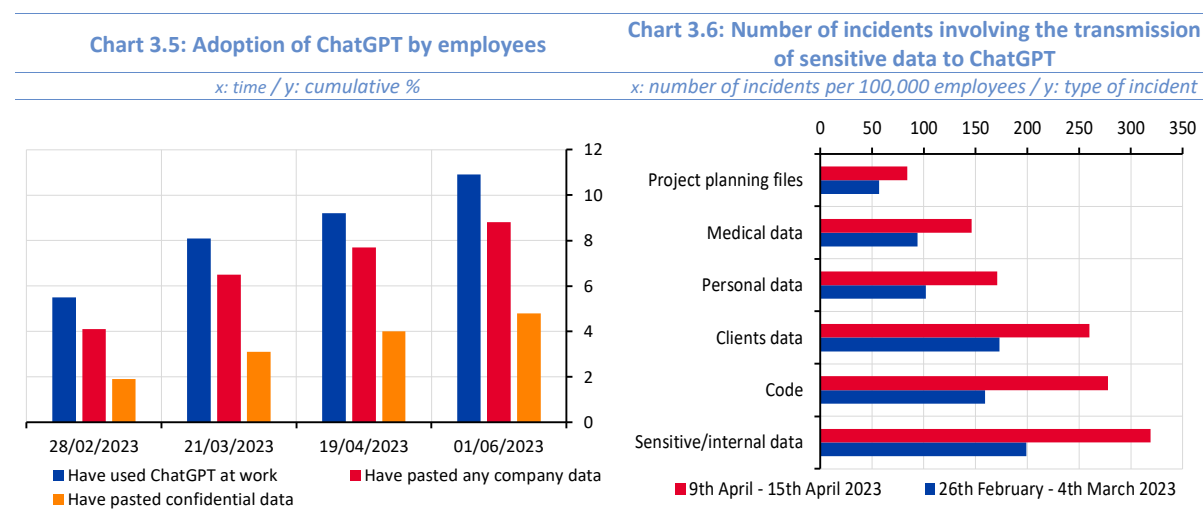
¹³⁹ [The financial supervisor in the age of AI | Banque de France \(banque-france.fr\)](#)

Singapore (MAS) uses artificial intelligence to analyse documents for red flags and to analyse transaction data to detect potential market manipulation¹⁴⁰. The Securities and Exchange Commission (SEC) also employs artificial intelligence to identify fraudulent activities and market manipulation.

3.2 Inherent risks of Artificial Intelligence

While the potential opportunities of artificial intelligence are numerous, its deployment carries inherent risks that could be amplified by generative artificial intelligence. Firstly, artificial intelligence raises concerns related to the quality and confidentiality of the data used for its training. The autonomous and opaque nature of artificial intelligence models also poses the risk of difficulty in understanding and trusting the results provided. Finally, excessive concentration of power among industry players would exacerbate these risks for users and be detrimental to the system as a whole.

Artificial intelligence, through its algorithms and learning processes, gives rise to risks related to data management, particularly data confidentiality. This risk is especially pronounced in generative artificial intelligence systems that learn not only from publicly available data but also from user-provided data (to improve their responses): these tools can thus transmit information learned from one user by generating content for others. According to a June 2023 report by Cyberhaven¹⁴¹, 10.8% of employees have used ChatGPT at work, and 4.7% of them have already provided confidential company information to the chatbot, with this percentage tending to increase over time. In June 2023, confidential data accounted for 11% of all data copied and pasted directly into the tool. These behaviours pose a significant threat to companies, as sensitive information can be widely disseminated. Furthermore, data transmitted to conversational tools can also be stolen during cyberattacks. To mitigate this risk and protect the confidentiality of financial data, several banking groups including JPMorgan, Bank of America, Goldman Sachs, and Deutsche Bank have banned the use of the ChatGPT online service for their employees. However, these risks could be lower in the case of tools reserved for internal company use, with strict data management rules.



Source: Cyberhaven.

More generally, the quality of results obtained from artificial intelligence technologies is directly dependent on the quality of the data. The use of inadequate data during the training of the tool, or during its use, can lead to erroneous results¹⁴². Some technologies are limited by their historical data or by the available data sources. For example, GPT 3.5 was trained on historical internet data up to January 2022, so a user may obtain incorrect or outdated information when querying the tool (e.g., no consideration of the war in Ukraine). At this stage, the information generated by generative artificial intelligence tools primarily comes from content written by humans. However, a strong adoption of these tools could lead to a feedback loop: the content they generate could become

¹⁴⁰ [Written reply to Parliamentary Question on use of artificial intelligence in supervision of financial institutions \(mas.gov.sg\)](#)

¹⁴¹ [Report](#) created based on data provided by 1.6 million users of Cyberhaven software worldwide.

¹⁴² [Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers \(oecd.org\)](#)

one of the main sources of information for generative artificial intelligence models. This recursive phenomenon poses a risk of losing the diversity of original information in favour of a standardization of responses that could be potentially biased, erroneous, or outdated¹⁴³.

AI can perpetuate historical biases present in the data or create new ones due to its functioning. This can create blind spots or even lead to a lack of fairness. A model trained on biased data will perpetuate - or even reinforce - these historical biases and potentially discriminatory practices. This is especially true since a biased model generates biased results which are then used in its learning process, reinforcing the biases (through recursion). For example, the algorithm used by Apple and Goldman Sachs for credit line allocation decisions reportedly offered lower loan amounts to women than to men, all else being equal ¹⁴⁴. Furthermore, biased decisions made by machine learning models can occur even with high-quality data, due to approximations, flaws in the algorithm's construction, or the fact that correlations between variables may be specific to the databases used (reflecting, for example, the characteristics of a historical sample).

Furthermore, artificial intelligence poses risks related to models that can undermine the reliability of results. For example, "hallucinations" are false responses presented as facts by tools using generative artificial intelligence with large language models. These hallucination problems in generative artificial intelligence models are inherent in the nature of large language models, which estimate a probable response based on the words provided by the user and can give abnormal or even false answers. For instance, a company asked ChatGPT to produce an article about the results of a company called Tesma, and ChatGPT created a coherent article but with invented and completely false financial figures. This phenomenon can also occur when the model is trained on false data. Moreover, the same question asked at two different times can generate different answers, highlighting a problem of temporal consistency. A January 2024 article from the Bank for International Settlements (BIS) also illustrates the limitations of generative artificial intelligence models, which invent false answers when faced with novel problems¹⁴⁵.

Technologies using artificial intelligence produce complex results, and their opaque functioning implies difficulties in understanding these results. The results of systems based on artificial intelligence are inherently complex to interpret for a human mind, due to the nature of certain modelling used (neural networks, etc.), which rely essentially on non-linear relationships. Some artificial intelligence firms may sometimes themselves reinforce the opacity of the functioning of tools, in order to protect their intellectual property. As a result, some artificial intelligence tools thus constitute black boxes, whose parameters and functioning are not visible to the entities integrating them into their services and whose sources are not mentioned.

Clear governance is necessary to reduce the risks associated with potential biases, lack of transparency and the reliability of the tool. Clear governance requires defining lines of responsibility for the development and supervision of artificial intelligence-based systems throughout their lifecycle, from development to deployment, along with explicit designation of responsibility for the results produced by the model.

The lack of competition among providers of artificial intelligence-based tools could exacerbate the risks. The costs required for the development of artificial intelligence-based models could entail the risk of dependency on oligopolistic suppliers. At this stage, OpenAI represents 39% of the generative artificial intelligence market compared to 30% for Microsoft, 8% for Amazon and 7% for Google ¹⁴⁶. The concentration of these technologies in a limited number of suppliers (relying on their advantages in terms of data or computing power) could amplify the risk of recursive use of data provided by artificial intelligences for model learning. Models would be trained with data from a limited number of artificial intelligence-based tools and would lose diversity.

Generative artificial intelligence could increase the risks of cyberattacks. Although most chatbots are built with internal safety features, such as refusing to respond to a direct request to create a cyberattack, they are generally

¹⁴³ The curse of recursion: training on generated data makes models forget. Shumailov, Ilia, Shumaylov, Zakhar, Zhao, Yiren, et al. 2023.

¹⁴⁴ Bias, fairness, and other ethical dimensions in artificial intelligence – Bank Underground

¹⁴⁵ Testing the cognitive limits of large language models (bis.org)

¹⁴⁶ The leading generative AI companies (iot-analytics.com)

not programmed to refuse the creation of computer code or text that could indirectly serve to launch cyberattacks.

3.3 Artificial intelligence poses significant risks to financial stability

Artificial intelligence presents a range of risks that could potentially exacerbate existing vulnerabilities in the financial system. The widespread adoption of artificial intelligence models by market participants could heighten the risk of herding behaviour, leading to amplified market volatility and procyclicality. Moreover, the deployment of artificial intelligence could lead to increased market concentration and the emergence of new systemic risks. Finally, the financial sector's heavy reliance on technology makes it particularly susceptible to the cyber threats posed by artificial intelligence.

Artificial intelligence (AI) has the potential to exacerbate market procyclicality and volatility. The increasing use of artificial intelligence in trading strategies and investment decisions could create a risk of homogeneous recommendations and herd behaviour, leading to heightened market procyclicality¹⁴⁷. When multiple market participants employ similar artificial intelligence tools, it can result in a build-up of one-sided trades, amplifying market movements during adverse conditions. For instance, if an algorithm recommends that several large investors sell an asset, the subsequent price decline could trigger sell recommendations for other investors, fuelling the downward price movement and impacting associated hedging positions. This can occur very rapidly, leading to a "flash crash." Such stress events can generate significant liquidity demands for weaker market participants, potentially forcing them to sell assets and spreading stress to other asset classes. Given the high degree of interconnectedness among financial institutions, these movements could have systemic consequences. Furthermore, if models are not fully understood, users may struggle to predict their behaviour under various market conditions and may find it difficult to adjust their strategies during periods of underperformance. Additionally, artificial intelligence can increase the risk of market manipulation. For example, the dissemination of deepfakes, synthetic media created by generative artificial intelligence, could trigger massive market movements and severe volatility.

More broadly, inaccurate risk assessments can lead to significant financial losses for institutions. Underestimation of risks could result in excessive lending or substantial pricing errors, causing significant losses. Similar to traditional statistical models, training artificial intelligence models on inherently incomplete data can amplify losses during crises. Since models are typically trained on historical data, they may not accurately capture the behaviour of an asset during extreme events. Moreover, using a model with incomplete data can lead to incorrect loan pricing by banks or systematic losses on insurance products. Such errors could potentially erode customer confidence, triggering bank runs and, if widespread, could have systemic implications.

The widespread adoption of artificial intelligence (AI) could lead to increased vendor concentration, creating dependency risks and potentially concentrating the financial system. Similar to the risks associated with a large number of market participants using the same software, the widespread use of a limited number of artificial intelligence tools could pose operational risks in the event of a technology failure or outage, hindering business operations. Furthermore, the dominance of a single artificial intelligence provider could amplify interconnections within the financial system and facilitate the spread of shocks¹⁴⁸. The growing prevalence of AI-based risk management systems, such as BlackRock's Aladdin and MSCI's RiskMetrics, could reinforce this trend by increasing the concentration of specialized players¹⁴⁹. Moreover, the development of artificial intelligence tools requires significant upfront investments, placing larger financial institutions in a better position to develop these technologies compared to smaller players. These technological disparities could further contribute to increased concentration within the financial system.

¹⁴⁷ [Intelligence artificielle: enjeux pour le secteur financier - Olivier Fliche, Su Yang - Fintech-Innovation Unit, ACPR](#)

¹⁴⁸ [SEC.gov | "AI, Finance, Movies, and the Law" Prepared Remarks before the Yale Law School](#)

¹⁴⁹ [Artificial intelligence and systemic risk \(lse.ac.uk\)](#)

Generative artificial intelligence could exacerbate the cyber risks to which the financial system is highly exposed. Generative artificial intelligence enables cyber attackers to create more sophisticated and harder-to-detect attacks. This technology can be used to automate the creation of highly personalized phishing emails in multiple languages with flawless grammar. Similarly, generative artificial intelligence systems can mimic a person's voice (or even image) to impersonate them. Additionally, generative artificial intelligence facilitates attacks involving malicious code. Conversational artificial intelligence agents can generate or modify code, enabling more sophisticated attacks. Artificial intelligence also enables the industrialization of cyberattacks by automating their design and execution, as artificial intelligence can adapt its strategy based on detected vulnerabilities. Furthermore, generative artificial intelligence has enabled a new category of cyberattacks known as indirect prompt injection, which involves manipulating prompts to bypass the tool's built-in security and allowing the generative artificial intelligence to retrieve sensitive information itself. Unlike previous methods, the artificial intelligence tool doesn't simply aid in creating malicious tools; it acts as a malicious actor itself.

Generative artificial intelligence solutions are being developed to bolster cybersecurity efforts, aiding in threat detection and dynamically strengthening security measures. Preparedness for cyberattacks is an essential component of cyber resilience, and by analysing vulnerabilities in real-time, detecting cyberattacks, and simulating attack scenarios, artificial intelligence can significantly improve cyber resilience. Furthermore, artificial intelligence can automate the response to attacks, implementing tailored protocols based on the specific threat.

The inherent risks of AI, such as potential biases, errors, and the opacity of models, can have a detrimental impact on the reputation of financial institutions. Errors or biases stemming from artificial intelligence models could become systemic within an organization or market, eroding customer trust. Beyond the risks to financial institutions, potential biases and difficulties in understanding model outputs can jeopardize customer and investor protection. A study by the ACPR highlights that conversational explanations provided by robo-advisors tend to inflate customer confidence, making them more likely to follow incorrect recommendations¹⁵⁰. Consequently, customers may be misinformed about investment risks and make suboptimal decisions.

The ability to understand model outputs is crucial for effective risk management by financial institutions and regulators. Comprehending how an algorithm works is necessary not only to verify calculations¹⁵¹ but also to examine the variables considered and their respective weights in determining the outcome¹⁵². Moreover, understanding the functioning of an algorithm is essential for regulators to ensure compliance with regulatory requirements. In the case of internal credit risk models used by banks, algorithms must be easily interpretable and understandable by both bank management, to enable informed strategic decisions regarding portfolio allocation, and regulators, to facilitate the application of prudential regulations¹⁵³.

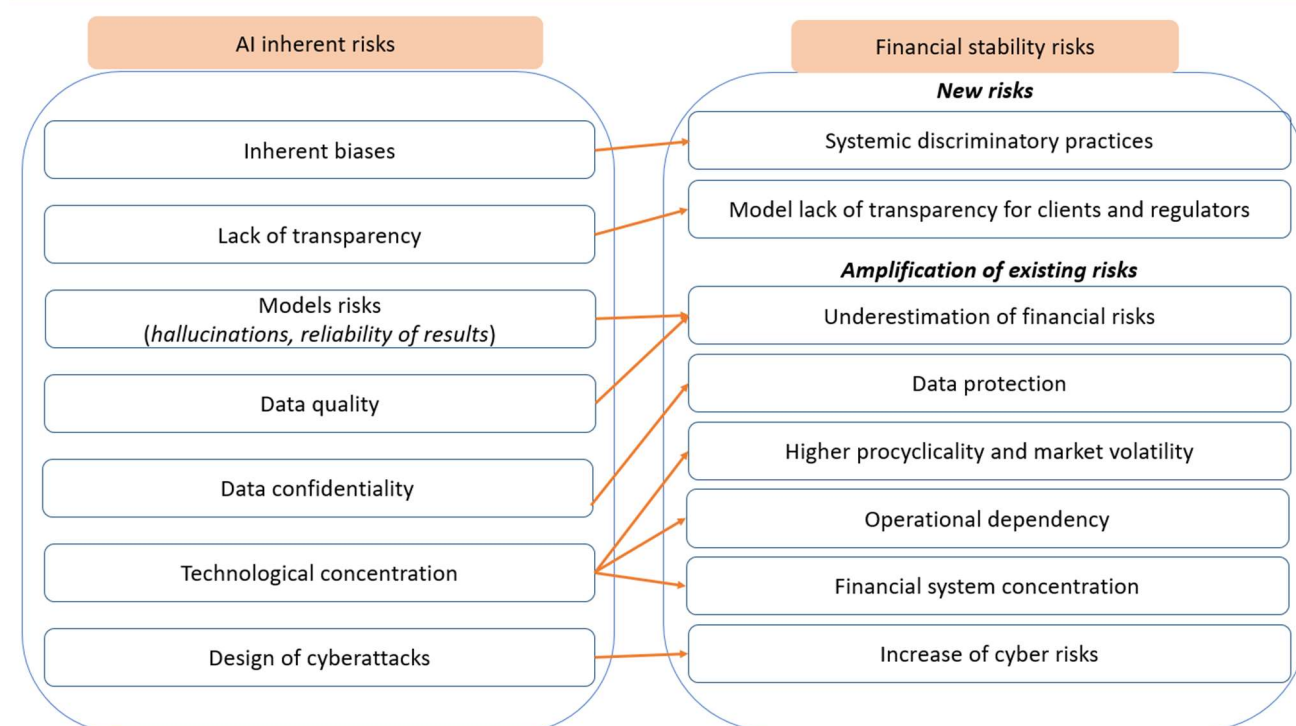
¹⁵⁰ [La motivation du conseil par les robo-advisors: vers un éclairage apporté aux clients? | ACPR \(banque-france.fr\)](#)

¹⁵¹ [Machine Learning and IRB Capital Requirements: Advantages, Risks and Recommendations - Institut Louis Bachelier](#)

¹⁵² [Artificial intelligence in central banking \(bis.org\)](#)

¹⁵³ [Artificial intelligence and machine learning in financial services - Financial Stability Board \(fsb.org\)](#)

Diagram 3.1: Typology of artificial intelligence risks for financial stability



Source: Banque de France.

3.4 Structural risks of information manipulation could impact the financial system

Generative artificial intelligence tools make it easier to create fake information for fraud or geopolitical disruption purposes, with potentially destabilising effects for the financial system.

Generative artificial intelligence tools facilitate the creation of deepfakes, which can be used for fraudulent activities and geopolitical destabilization, with potentially disruptive effects on the financial system. Generative artificial intelligence technologies enable the creation of false videos and recordings by superimposing images and sound onto existing content. This process, known as "deepfaking," can be used for manipulation and fraud. In early February 2024, the governor of the Romanian central bank's image and voice were deepfaked to promote fraudulent investments. In September 2023, the chairman of the Securities and Exchange Commission (SEC), Gary Gensler, warned about the risk of significant one-directional market movements caused by deepfakes if false information led to a loss of confidence in an asset.

These artificial intelligence -generated deepfakes can erode trust in financial institutions, potentially triggering bank runs. The US regional banking crisis in March 2023 highlighted how quickly panic can spread among depositors through social media¹⁵⁴. When artificial intelligence can make false information about banks seem credible, its dissemination can lead to panic and mass deposits withdrawals.

More broadly, disinformation campaigns can destabilize geopolitical situations, leading to significant political, social, and economic consequences. The 2024 World Economic Forum Global Risks Report ranked disinformation as the most severe short-term (two years) threat¹⁵⁵. To mitigate the risk of deepfake-generated disinformation, companies in the sector are increasingly implementing measures to detect or limit their spread. For example, OpenAI has announced the addition of watermarks to images generated by its systems, while Midjourney has banned the generation of images using the faces of political figures. To improve deepfake detection and

¹⁵⁴ Social Media as a Bank Run Catalyst by J. Anthony Cookson, Corbin Fox, Javier Gil-Bazo, Juan Felipe Imbet, Christoph Schiller: SSRN

¹⁵⁵ The world is changing and so are the challenges it faces | World Economic Forum (weforum.org)

prevention, public awareness is essential, as is collaboration among various stakeholders, including specialized companies, social media platforms, telecommunications operators, and media outlets.

Chart 3.7: Google search volumes for the terms “artificial intelligence” and “deepfakes”

x: time / y: index

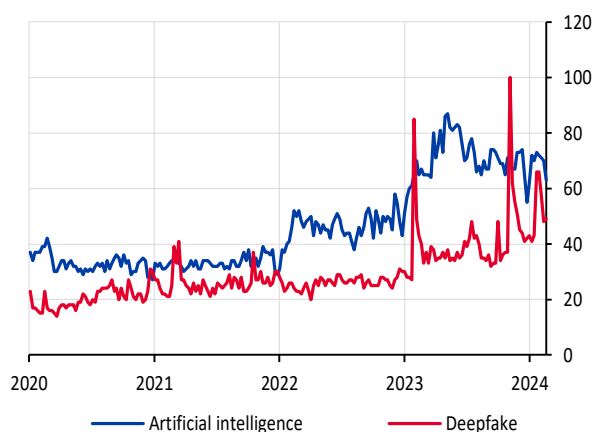
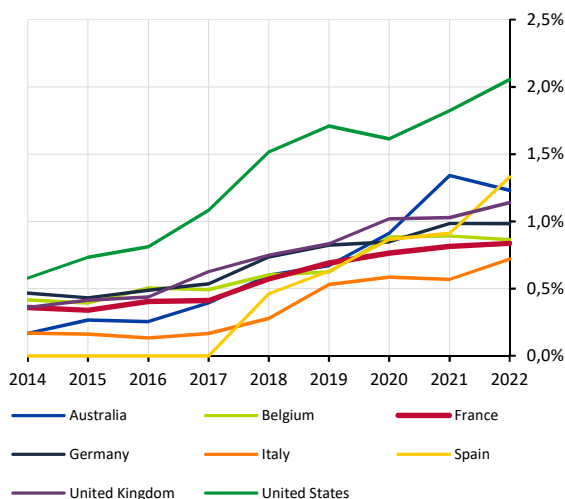


Chart 3.8: Share of artificial intelligence jobs in all job offers

x: time / y: %



Source [left]: Google Trends; [right]: OurWorldInData.

Note: The Google trends shown here are measured as an index which takes the value 100 for the point of highest search interest for the term since 2020.

Automation driven by artificial intelligence could lead to structural economic changes, although the exact scale and nature of these changes remain uncertain¹⁵⁶. Not all countries or economic sectors are equally exposed to artificial intelligence-driven labour market shifts. Developed economies, with their advanced IT infrastructure and a high proportion of automatable tasks (such as those found in manufacturing, legal, accounting, finance, and IT sectors), are likely to be most affected by the rise of artificial intelligence¹⁵⁷. However, in the short term, few jobs will be entirely automated by artificial intelligence. Instead, many roles will gradually integrate artificial intelligence components. Artificial intelligence is therefore likely to cause significant sectoral restructurings, which could destabilize the financial sector if they are not adequately anticipated or occur more rapidly than expected. The widespread adoption of artificial intelligence could lead to substantial job shifts between sectors, potentially resulting in high rates of business failures and impacting the solvency of affected companies or households.

Artificial Intelligence is likely to drive significant sectoral restructurings, which could destabilize the financial sector if they are poorly anticipated or occur more rapidly than expected. The deployment of artificial intelligence could lead to substantial job shifts across various economic sectors, consequently resulting in a high number of bankruptcies within certain industries or companies themselves, with potential implications for the solvency of affected businesses or households. Furthermore, in the markets, the growing adoption and enthusiasm surrounding artificial intelligence have propelled valuations of artificial intelligence-related companies due to investor optimism regarding the future benefits of these new technologies (see Box 1.2, Chapter 1)¹⁵⁸.

The future development of artificial Intelligence will need to consider the constraints imposed by environmental limitations. Training an artificial intelligence model requires significant computational power, which in turn demands a large amount of energy and results in greenhouse gas emissions. According to OpenAI research¹⁵⁹, since 2012, the computational power needed to train artificial intelligence models has doubled every three months. An online search using generative artificial intelligence consumes approximately 4 to 5 times the energy of a traditional online search using a search engine¹⁶⁰. Additionally, generative artificial intelligence systems require substantial amounts of water for processor cooling. The development of generative artificial intelligence models by Google and Microsoft has led to annual increases in water consumption of 20% and 34%, respectively,

¹⁵⁶ Innovation by central banks: the sooner the better | Banque de France (banque-france.fr)

¹⁵⁷ Gen-AI: Artificial Intelligence and the Future of Work (imf.org)

¹⁵⁸ Financial Stability Review, November 2023 (europa.eu)

¹⁵⁹ AI and Compute (georgetown.edu)

according to the companies' environmental reports. The environmental consequences of artificial intelligence will therefore pose potential limitations to its deployment.

Box 3.1: The Banque de France's use of artificial intelligence

By Philippe Grad

The rise of artificial intelligence also presents opportunities for the Banque de France to carry out its missions. To make the most of these new technologies while managing their risks, the Banque de France has adopted an experimental approach. This involves testing the potential contribution of a given technological solution on a specific use case, in order to fully understand its advantages and disadvantages and to minimize uncertainties. If the experiment proves successful, the project is put into production.

In this context, about ten applications now integrate artificial intelligence functionalities, including:

- **N-ACSEL**: a partitioning (clustering) model based on economic data that allows for "profiling" territories. This view, provided to local officials, enables them to better understand their strengths and weaknesses and to benchmark themselves against other territories;
- **ScoreIA**: a credit rating model, incorporated into the company rating application to be used as a decision-support tool by analysts;
- **PAI**: a tool for measuring perceptions and inflation expectations, established by analysing a large panel of press articles;
- **BASTID**: a bank fraud detection system;
- **LUCIA**: a tool for combating money laundering (AML/CFT), enabling the analysis of large volumes of banking transactions and used in the context of on-site inspections of institutions;
- **PLUME**: a system for transcribing voice recordings into text, used in particular in the control of commercial practices in the financial sector.

Other, more exploratory work using artificial intelligence and other original sources such as satellite data, for example, is also being conducted on the use of pollution data to monitor industrial activity¹⁶⁰, including with external partners¹⁶¹, making it possible to cover many countries with standardised methods.

The Banque de France is indeed in a favourable position to deploy artificial intelligence -based solutions: a diverse and abundant dataset, a platform to centralize this data and organize its governance, infrastructures for innovation, and a broad technical expertise, particularly in data scientist and data engineer teams attached to various departments, working in coordination with the Banque de France's innovation LAB.

In parallel with these achievements, the methodological aspect of designing, developing, and using artificial intelligence systems has also been invested in. This work has led to the establishment of a framework that applies to both development teams and users of artificial intelligence systems. This framework, called "trustworthy artificial intelligence", includes:

- a code of conduct defining the principles and behaviour to be adopted by artificial intelligence developers and users;
- a risk analysis method specific to the technology, its implementation, and its use;
- a toolkit to address risks.

The method is based on seven challenges: i) diversity, non-discrimination, and fairness; ii) security, robustness, and reliability; iii) transparency and explainability; iv) privacy and personal data protection; v) working

¹⁶⁰ [Can satellite data on air pollution predict industrial production? - Banque de France - November 2021](#)

¹⁶¹ [Satellites Turn "Concrete": Tracking Cement with Satellite Data and Neural Networks - Banque de France - June 2023](#)

environment, social and environmental responsibility; vi) human dimension and human oversight; vii) accountability and responsibility.

These challenges, reviewed as part of the artificial intelligence projects, make it possible to identify a hierarchy of risks. Faced with these risks, a toolkit of good practices is applied throughout the lifecycle and is integrated into our internal control organization.

These good practices are organized into three blocks corresponding to the development stages: i) data sourcing; ii) development and validation of the solution; iii) model operation.

3.5. Regulatory pathways for artificial intelligence

The rapid evolution of artificial intelligence and its increasing adoption necessitate a regulatory framework to mitigate the various risks mentioned above.

Within the European Union, The European Parliament and the Council of the European Union adopted, on 13 March and 21 May 2024 respectively, a regulation, known as the Artificial Intelligence Act (AI Act), aimed at fostering innovation and promoting trustworthy artificial intelligence, in compliance with fundamental rights, security and ethical principles,¹⁶² while mitigating the risks potentially presented by the most powerful artificial intelligence models.

This directly applicable regulation distinguishes artificial intelligence systems based on the level of risk they present:

- Artificial intelligence practices posing an unacceptable level of risk to the safety and fundamental rights of citizens, such as subliminal manipulation, social scoring, or predictive policing, will be prohibited by the AI Act (Chapter II).
- High-risk artificial intelligence systems will be regulated. In the financial sector, the two considered use cases are, for banks, the assessment of the solvency of natural persons and, for insurance companies, the assessment of pricing for natural persons in life and health insurance. Before being placed on the market or deployed, these systems will have to be registered and a declaration of conformity will have to be made (covering the following items: risk management processes; high data quality; documentation ensuring traceability and auditability; appropriate degree of transparency and interpretability; measures allowing human oversight; robustness, accuracy, cybersecurity). In France, the ACPR will be the competent authority (market supervisory authority) responsible for overseeing the use of these systems by banks and insurance companies.

Other artificial intelligence systems, sometimes referred to as limited-risk systems and formally grouped under the heading “certain artificial intelligence systems” (Chapter IV, Article 50), will have to comply with transparency requirements, notably by specifying that their content was produced by an artificial intelligence. These requirements will apply particularly to systems that interact directly with human users or that generate content (images, videos, text).

European lawmakers also addressed the specific case of general-purpose artificial intelligence models. These models, defined as those capable of performing a wide range of distinct tasks (Article 3, paragraph 63), and which can be integrated into other systems, are subject to specific requirements (Chapter V). All general-purpose models must therefore comply with documentation and information sharing obligations with deployers. For models presenting a systemic risk, providers must comply with more stringent evaluation and transparency requirements, similar to those imposed on high-risk systems (see above).

¹⁶² The General Data Protection Regulation (GDPR) already states that any person “shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her” (Article 22).

The AI Act will generally apply two years after its entry into force. However, the provisions concerning unacceptable risks and those concerning general-purpose models will enter into application six and twelve months after the entry into force of the regulation, respectively. The articulation between the horizontal rules set out by the AI Act and the vertical regulation specific to the financial sector is not explicitly provided for in the regulation (in particular to give flexibility to supervisors in how to integrate these new missions). This articulation issue will be crucial for financial supervisors, and could be specified in particular in guidelines from the European Banking Authority (EBA) and the European Insurance and Occupational Pensions Authority (EIOPA).

International regulators have begun work on the impact of artificial intelligence on financial stability. The Financial Stability Board (FSB) and the Basel Committee have included artificial intelligence among the priorities of their 2024 work program. On the one hand, the FSB plans to formulate recommendations on taking into account the potential implications of artificial intelligence for financial stability, following an initial report published in November 2017. In a report published in May 2024¹⁶³, the Basel Committee presents the implications of the digital transformation of finance, and in particular the use of artificial intelligence for banks and their supervisors. While the deployment of artificial intelligence could lead to greater automation of certain tasks, this report emphasizes the need to maintain human judgment in the governance and risk management of financial institutions.

Effective enforcement of regulations necessitates robust oversight mechanisms. Financial regulators must therefore prepare to audit artificial intelligence systems deployed within the financial sector, prioritizing those posing the greatest risks to key regulatory objectives (financial stability, consumer protection, anti-money laundering/countering the financing of terrorism). This presents a multifaceted challenge, requiring a deep understanding of the technology. Beyond human resource implications, regulators will need to develop a tailored methodology for auditing artificial intelligence systems. A pilot approach, potentially in collaboration with the private sector, can accelerate the learning curve. Mitigating artificial intelligence risks is a collective responsibility, involving all financial system participants.

¹⁶³ [Digitalisation of finance \(bis.org\)](https://bis.org/digitalisation-of-finance)