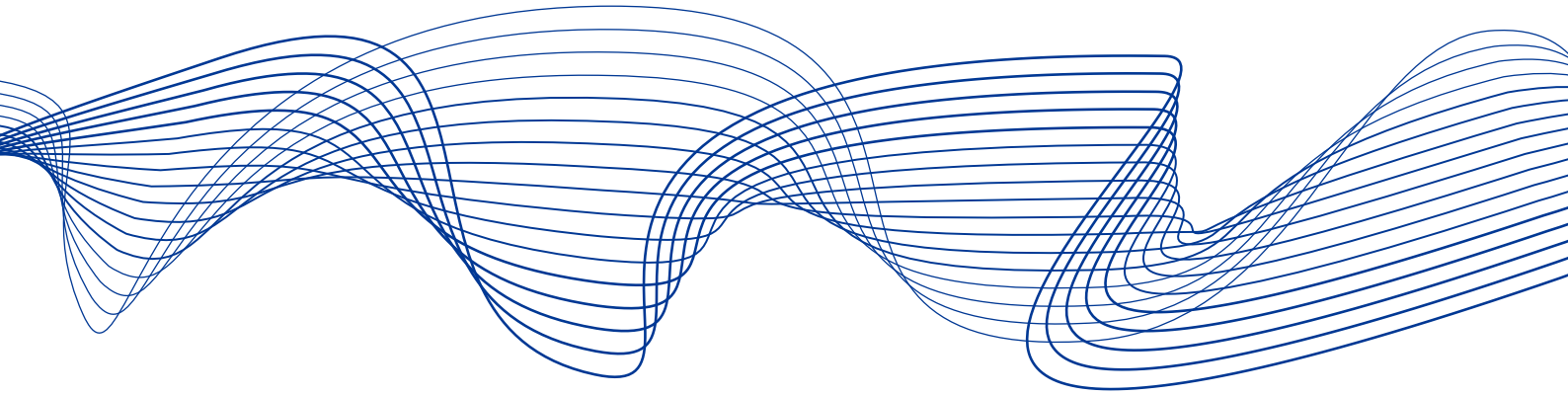
ESRB
European Systemic Risk Board
European System of Financial Supervision

Advisory Scientific Committee

No 16 / December 2025

Artificial intelligence and systemic risk



Stephen Cecchetti, Robin L. Lumsdaine, Tuomas Peltonen, Antonio Sánchez Serrano

Contents

Executive summary[1]

In this report, we discuss the implications of the rapid development and widespread adoption of artificial intelligence (AI) for financial stability. Sizeable corporate investments have made advanced AI capabilities like large language models (LLMs) such as ChatGPT, Claude and Gemini widely accessible. Some currently boast 800 million weekly active users. This proliferation of easy-to-use tools is leading to the rapid integration of AI into corporate processes, though there is little consensus on how to do so effectively.

Our report emphasises that while AI offers substantial benefits, including accelerated scientific progress, improved economic growth, better decision making and risk management and enhanced healthcare, it also generates significant concerns regarding risks to the financial system and society. AI's ability to process immense quantities of unstructured data and interact naturally with users allows it to complement and substitute human tasks, potentially revolutionising how work is organised. However, this comes with risks such as difficulty in detecting AI errors, inherited biases, overreliance and challenges in oversight.

With this in mind we examine how AI might amplify or alter existing systemic risks in finance or create new ones. Systemic financial risks typically stem from five categories: liquidity mismatches, common exposures, interconnectedness, lack of substitutability and leverage. In order of importance, AI features that can exacerbate these risks include:

- Monitoring challenges: The complexity of AI systems makes effective oversight difficult for both users and authorities.
- Concentration and entry barriers: A small number of AI providers can lead to single points of failure and interconnectedness.
- Model uniformity: Widespread use of similar AI models can lead to correlated exposures and amplified market reactions.
- Overreliance and excessive trust: Superior performance in good times can lead people to place too much trust in AI, increasing risk taking and hindering oversight. In addition, demonstrated accuracy in some tasks can lead to trust in similar accuracy in others that go beyond current capability.
- Speed: Increased speed of transactions, reactions and enhanced automation can amplify procyclicality and make it harder to stop negative processes.
- Opacity and concealment: AI's complexity can diminish transparency and facilitate intentional concealment of information.

[1] The executive summary is an extensively edited version of a draft initially produced by Google's NotebookLM when prompted to summarise an earlier version of the report. Use of AI tools in other parts of the report was limited to minor editing.

- Malicious uses: AI can enhance the capacity for internal fraud, cyber-attacks and market manipulation by malicious actors.
- Hallucinations and misinformation: AI can generate false or misleading information, leading to widespread misinformed decisions and subsequent market instability.
- History-constrained: AI's reliance on past data makes it struggle with unforeseen "tail events", potentially leading to excessive risk taking.
- Untested legal status: Ambiguity around legal responsibility for AI actions (e.g. the right to use data for training and liability for advice provided) can pose systemic risks if providers or financial institutions face AI-related legal setbacks.
- Complexity makes them inscrutable: The difficulty in understanding AI's decision-making processes can trigger runs when users discover flaws or behaviour is unexpected.

Potential features such as self-aware AI and complete human reliance on AI could further amplify these risks, possibly leading to a loss of human control and extreme societal dependency.

In response to these systemic risks and associated market failures (fixed cost and network effects, information asymmetries, bounded rationality), we call for a thoughtful and measured policy approach to address the systemic risks AI creates. This involves a mix of competition and consumer protection policies, complemented by adjustments to prudential regulation and supervision. Key policy proposals include:

- Regulatory adjustments: Recalibrating capital and liquidity requirements, enhancing circuit breakers, amending regulations addressing insider trading and other types of market abuse and adjusting central bank liquidity facilities.
- Transparency requirements: Adding labels to financial products to increase transparency about AI use.
- "Skin-in-the-game" and "level of sophistication" requirements: Ensuring AI providers and users bear appropriate risk.
- Supervisory enhancements: Ensuring adequate IT and staff resources for supervisors, increasing analytical capabilities, strengthening oversight and enforcement and promoting cross-border cooperation.

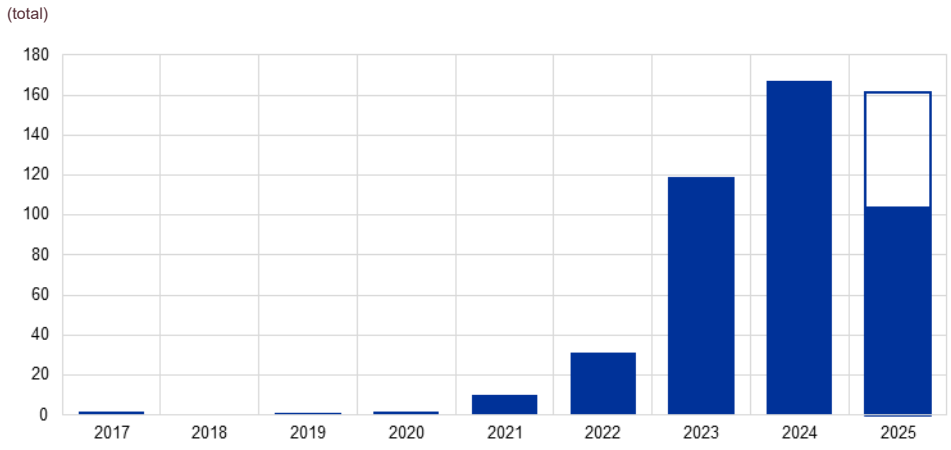
These will all require further analysis to get a clearer picture of the impact and channels of influence of AI, as well as the extent of its use in the financial sector. Critically, these policies must evolve rapidly due to AI's fast-changing nature – and its global dimension necessitates international cooperation. The report concludes that failing to keep pace with AI in finance could lead to increased financial instability and more frequent interventions by authorities.

1 Introduction

Artificial intelligence (AI) is everywhere. Sizeable corporate investment in developing the large-scale models that underlie tools like OpenAI's ChatGPT, Anthropic's Claude, Microsoft's Copilot and Google's Gemini is making advanced AI capabilities available to virtually anyone with a smartphone.[2] While OpenAI does not publish exact numbers, recent reports suggest ChatGPT has roughly 800 million active weekly users.[3] Rapid development and adoption mean that there is now a proliferation of easy-to-use tools, with little consensus about how to integrate these into our environment. Chart 1 shows the sharp increase in the release of large-scale AI systems since 2020. The fact that people find these tools, made feasible by the development of large language models (LLMs), intuitive to use is surely one reason for their speedy widespread adoption. In part due to the seamless inclusion of these tools in existing day-to-day platforms, companies are rapidly integrating AI tools into their processes. For example, Copilot is a part of the suite of Microsoft Office tools, Gemini is available as part of Google's standard search engine, and Zoom has an AI assistant that produces meeting summaries.

Chart 1

Number of large-scale AI systems released per year



Source: World in Data.
Notes: Data for 2025 up to 24 August. The white box in the 2025 bar is the result of extrapolating the data to that date for the full year.

While there are large differences across Member States, in 2024 an average of 13% of EU non-financial corporations reported using some form of AI.[4] This marks a substantial increase from the 8% that reported doing so the year before. The highest level of adoption, with levels above 25%, is by Danish and Swedish non-financial corporations, with Belgium and Finland slightly below 25%. At the other end of the spectrum, fewer than 4% of Greek, Hungarian, Polish, Bulgarian and

2 "Large-scale" models are those where training requires computing in excess of 10^{23} floating-point operations.

3 See the demandsage website.

4 See Eurostat (2025). Additionally, employees of some corporations may be using AI tools without official adoption by the corporation.

Romanian non-financial corporations report using any form of AI. According to the Eurostat survey, smaller non-financial corporations are more limited in their adoption of AI technologies.

This rapid development and deployment give rise to concern that AI is generating risks to the financial system as well as unknown (and unknowable) risks to society at large. The main entities developing AI lie outside the purview of financial supervisory authorities. Furthermore, the way these tools are being developed, integrated and used, and the implications for systemic risk, remain poorly understood. Some users are wholeheartedly embracing AI with little regard for the risks; others are voicing concern that "the robots are taking over". From a regulatory perspective it is critical that we take a thoughtful and measured approach and avoid overreacting, while remaining alert to emerging areas of concern. The uses of AI receiving attention may not be the channels most likely to create systemic risk. This may just reflect the existing uncertainty about the advances in AI and how both the financial sector and society at large will use it. In addition, current AI policies and regulations – both within individual firms and at the level of the EU AI Act – understandably focus on protecting individuals (e.g. safeguarding personal information) and sheltering firms from legal liabilities.[5] As a result, the current constellation of regulations may not go far enough in addressing systemic risk.

There are various estimates of the impact generative AI will have on labour and on non-financial corporations. Gmyrek et al. (2023) analyse 436 occupations, identifying those most likely to be affected by AI. The authors distinguish three groups: those least likely to be affected (mainly composed of manual and unskilled workers), those where AI will augment or complement tasks (occupations such as photographers, primary school teachers and pharmacists), and those where it is difficult to predict.

There is no clear definition of AI, as it encompasses many different activities (Box 1).[6] AI technologies, including generative AI models such as chatbots based on LLMs, have increased their capabilities in recent years.[7] For the purposes of our assessment of the impact that AI will have on systemic risk and the subsequent policy discussion, we follow Bengio et al. (2025) and think of "AI" as a general-purpose technology that is at least as capable as today's most advanced general-purpose AI. These are systems that can perform a wide range of tasks. They can generate output with more flexibility than narrow AI trained to perform a specific task. General-purpose AI is not equivalent to artificial general intelligence, however. The latter refers to potential future AI that equals or surpasses human performance on all or almost all cognitive tasks. By contrast, today's most advanced general-purpose AI surpasses human performance in a limited set of tasks only.

---

5 See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (OJ L 12.7.2024). Annex 1 contains a brief summary of the AI Act.

6 See Annex 1 for the definitions of AI and general-purpose AI in the AI Act.

7 Generative AI is a type of AI that can create content such as text, images, music, videos, code or other sources of data, based on inputs or prompts.

Box 1
The concept of AI

The term "artificial intelligence" has many meanings. First defined by John McCarthy in 1955 as "the science and engineering of making intelligent machines", a recent survey of work published between 2005 and 2020 lists 28 definitions.[8] A few examples include (i) "computer systems that perform tasks that typically require human intelligence", (ii) "the ability of a machine to perform cognitive functions that we associate with human minds, such as perceiving, reasoning, learning, interacting with the environment, problem solving, decision making, and even demonstrating creativity"; and (iii) "a computer algorithm that makes decisions that otherwise would be taken by human beings".[9] In the official sector, the White House Executive Order on the Safe, Secure, and Trustworthy Development and Use of AI (Biden, 2023) defines AI as "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems use machine- and human-based inputs to perceive real and virtual environments; abstract such perceptions into models through analysis in an automated manner; and use model inference to formulate options for information or action."[10] The European Institute for Innovation and Technology states that "Artificial Intelligence is a machine-based system that perceives its environment, pursues goals, adapts to feedback/change and provides information/takes action."[11] And finally, EU Regulation 2024/1689 (the "AI Act") defines an AI system as "[…] a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments".

Combining all of these, we see the concept of AI as encompassing everything that includes algorithms, machine learning and generative AI (including LLMs) and AI agents (Figure A). Algorithms are "a set of mathematical instructions or rules that, especially if given to a computer, will help to calculate an answer to a problem".[12] While they do not necessarily imply the use of computers (tying our shoe laces is algorithmic), using coded instructions run on a computer can substantially increase speed, frequency and scope of execution. Algorithmic trading, where a machine executes trades thousands of times faster than a person possibly could, is an example of such an application. Where AI may be distinct from more traditional algorithms is in the key role it can play in discretionary decision making, potentially with little or no supervision. Conventional machine learning techniques probabilistically infer patterns in existing data.[13] Machine learning is thus a subset of AI. Current state-of-the-art machine learning systems utilise neural networks and deep learning, algorithms

8 See Collins et al. (2021).

9 See Russell and Norvig (2010), Rai et al. (2019) and Daníelsson et al. (2022).

10 See Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.

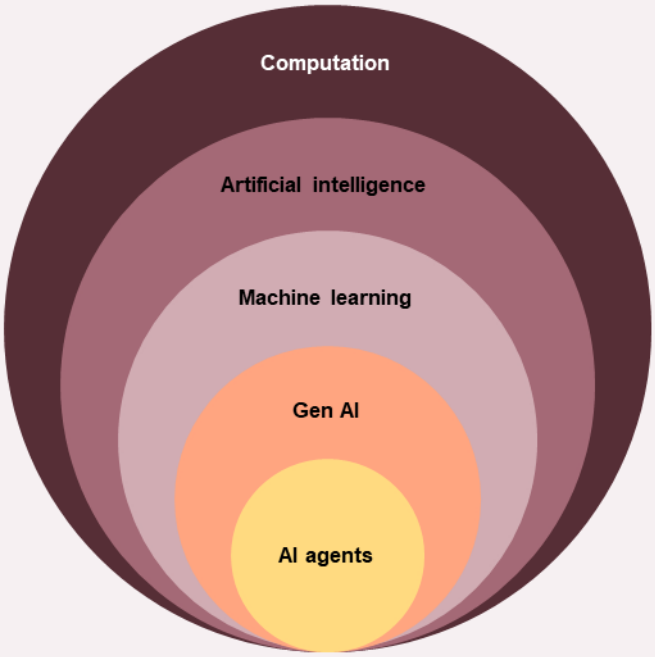
11 See European Institute of Innovation and Technology (2021).

12 See the Cambridge Dictionary.

13 Daníelsson et al. (2022) define machine learning as a process "*whereby a computer algorithm takes in data and infers the reduced form statistical model governing the data.*"

inspired by the structure of the human brain. Advancements in machine learning have enabled algorithms to generate “new” data (generative, or gen, AI); LLMs are one of the most important applications of this technology, facilitating the processing and generation of human language.[14] As one writer puts it, “Machine learning is a data-driven way to achieve AI, but not the only one”.[15] Finally, it is important to note that AI algorithms based on machine learning, including LLMs, are probabilistic. That means that they can (and do) generate different answers to the exact same query at different points in time. By contrast, traditional algorithms consistently generate the same answer to a given problem every time they run.

Figure A
AI and computing



Source: Aldasoro et al. (2024).

Taking a risk management perspective, our concern over AI stems from how people use it. For example, there are surely times when it is safe to employ AI as a substitute for decisions traditionally made by people. But there are also cases where the use of autonomous machines with limited (or no) human intervention can be dangerous.

Importantly, AI has been with us since the earliest days of electronic computing in the 1950s.[16] What is new is the ability to process immense quantities of unstructured data – words, pictures, and audio files – not just arrays of numbers. Furthermore, these new LLMs interact with people in a natural manner, updating and

[14] See Aldasoro et al. (2024).
[15] See Giudici (2018).
[16] See Fernández (2019).

providing responses in near-real time. This allows users to apply AI to a multitude of tasks including summarising large volumes of information, creating presentations, writing code, editing papers (like this report) and creating podcasts, among other things. Given the power of these tools, it is unsurprising that numerous firms and government agencies are implementing AI tools in their organisations.[17] This ability to complement and substitute for people has the potential to fundamentally change the way we organise tasks.

As always, these benefits come with risks and pitfalls. AI blurs the boundaries between the financial sector and the real economy. For instance, by paralysing digital payments and trading platforms, a failure in an AI system governing the energy grid or telecom network could quickly turn into a financial crisis. Furthermore, AI errors are difficult to detect, outputs inherit biases from their training data, and there is a tendency towards excessive trust and overreliance on the tools themselves. Since many of our existing enforcement protocols require identifying a person with legal responsibility, oversight is challenging. How do we regulate and sanction algorithms and code? And will AI narrow or widen the divergence between the informed and the uninformed?

In this report we examine the impact of AI on financial stability.[18] Is it revolutionary or evolutionary? How might it amplify or alter existing sources of systemic risks? Does it create new systemic risks, requiring new macroprudential tools to address vulnerabilities? And in what ways might it aid risk identification, mitigation, and management? Will it influence system resilience, and if it does, what regulatory modifications will we need to make to ensure continued resilience? Is the use of AI encouraging further concentration and wider contagion? By increasing the speed of information processing, does it increase fragility, thus requiring stronger circuit breakers to curtail the transmission of shocks? Even if AI does not create fundamentally new risks, is it pushing familiar risks outside the existing perimeter of regulation into activities and entities where the official sector has little visibility or legal authority?[19]

AI can influence the financial system through both the supply of and the demand for financial services. It clearly changes the way the financial system operates, affecting the range of financial services supplied to society. By increasing both the variety of services and access to them, AI will likely increase overall demand.

---

17 See the discussions in Cipollone (2024) and Barr (2025). Among other things, they note that to control access to sensitive information, official sector entities purchase or subscribe to pre-trained foundation models and then use their own data to fine tune them before using them in controlled internal environments. For a broader overview on the use of AI in central banks around the world, see Irving Fisher Committee on Central Bank Statistics (2025).

18 High market valuations of major AI providers that could trigger sharp market corrections pose an additional financial risk.

19 We note that there is a large and growing literature discussing the impact of AI. There are publications and reports by various international institutions: Bank for International Settlements (2024) Chapter 3, Financial Stability Board (2024), International Monetary Fund (2024) and Organisation for Economic Co-operation and Development (2024a and 2024b). There is also work by official sector researchers: Aldasoro et al. (2024), Kumar et al. (2023), Fernández (2019) and Marchetti (2022), as well as academics: Comunale and Manera (2024), Daníelsson et al. (2022), Daníelsson and Uthemann (2024a and 2024b), Gaske (2023), Giudici (2018), Ji et al. (2024), Leitner et al. (2024), O'Halloran and Nowaczyk (2019), Petrone et al. (2022) and Remolina (2022). While much of the literature expresses concern about the risks of AI, Barefoot (2022), O'Halloran and Nowaczyk (2019) and Petrone et al. (2022) argue for making AI central to a modernised regulatory framework.

In this report we focus primarily on changes in the supply of financial services, which we see as the less speculative and more direct impact of AI on the financial system. We build on recent reports to discuss systemic risks derived from the impact of AI on society, including changes in labour markets and, indirectly, on finance.[20]

Before proceeding, we must emphasise that like all technologies, AI itself is outside the financial regulatory perimeter. It is the uses of AI by investors and financial intermediaries that can generate externalities and spillovers, creating the systemic risks that justify intervention. As such, our focus is on applications of AI, not the algorithms themselves. In addition, we do not discuss the risk associated with institutions either failing to innovate or implementing the "wrong" AI technology.

The remainder of this report is organised as follows. In the next section, we discuss the impact of AI on society, with a particular focus on the financial system. In Section 3, we examine AI and systemic risks. We start with the sources of systemic risk, then discuss the properties of AI and finally tabulate which of its properties are associated with which systemic risks. Section 4 follows with a discussion of some regulation and policy implications. Annexes and boxes provide more detailed analysis on specific topics.

[20] See, for example, Videgaray et al. (2024) and Bengio et al. (2025).

2 Framing the current impact of AI on society and the financial system

We start with a discussion of the impact of AI on society and the financial system. What is the likely impact of AI on the labour force and the real economy? To understand the implications of AI for the financial system, we look at historical experience with technological innovation.

2.1 Understanding the impact of AI on society

AI technologies have the potential to bring substantial benefits to our societies. Figure 1 below shows the ten potential benefits of AI identified in a recent survey by the Organisation for Economic Co-operation and Development (2024a). They range from those that affect the availability and use of information, governance, decision making, scientific discovery, provision of healthcare, and possibly reduced inequality. Focusing on the financial system, we note the potential for AI to have a significant positive impact on both the range of available services and their methods of delivery, as well as internal management of risk and how supervisory authorities monitor institutions.[21]

Figure 1
Potential benefits from the use of AI

Source: World in Data.
Notes: Data for 2025 up to 24 August. The white box in the 2025 bar is the result of extrapolating the data to that date for the full year.

To fully reap the benefits, it is important that AI users understand its nature and capabilities, as well as its limitations and risks. When a user asks an LLM a question, the program scans its training data and returns the statistically most likely response based on learned patterns. The most probable answer is not necessarily the

[21] For further details on the benefits and opportunities of AI for financial institutions, see Leitner et al. (2024).

best answer, however. In some cases, AI may opt for the most widely accepted view, even if current circumstances or context makes it inappropriate. For example, when asked to generate a random number between 0 and 10, mimicking observed human behaviour, many tools will give 7 as an answer with a higher probability. Due to its deterministic nature, AI cannot generate purely random numbers.[22] This "probabilistic friction" has important consequences for the use of AI in cryptography, for example. Specifically, an AI user needs to give clear instructions, providing as much accurate information as possible. Otherwise, the response provided may not be optimal. In other words, AI maximises an objective subject to a set of constraints. This means that if users (or designers) provide incorrect or incomplete specifications of the objective or the constraints, even if unintentionally, the AI agent may produce undesirable outcomes.

That said, AI can solve large-scale problems quickly, changing how we allocate resources.[23] In particular, it gives us the ability to dramatically increase the amount of information we can process in an hour or a day. General uses of AI comprise knowledge-intensive tasks such as (i) aiding decision making, (ii) simulating large networks, (iii) summarising large bodies of information, (iv) solving complex optimisation problems, and (v) drafting. There are numerous channels through which AI can create productivity gains, including automation (or deepening existing automation), task complementarity and new tasks. AI not only helps humans solve problems; in some instances, it can also propose solutions on its own. Overall, this should increase productivity. We should note, however, that current estimates of the overall productivity impact of AI tend to be quite low. In a detailed study of the US economy, Acemoglu (2024) estimates the impact on total factor productivity (TFP) to be in the range of 0.05%-0.06% per year over the next decade. Since average annual TFP growth in the 21st century has been roughly 0.9% in the United States, this is a very modest improvement.

In line with other innovations, it may take time to reap AI's productivity benefits. Looking at the industrial revolution, evidence suggests that the higher productivity induced by the steam engine took 80 years to become visible in national statistics.[24] Similarly, the broad impact from the introduction of electricity in the United States came in the 1920s, some 40 years after Edison's initial generating stations began supplying power to customers in Lower Manhattan.[25] The reason is that taking full advantage of new technology can require businesses to re-organise. This means time-consuming investments in both physical and human capital.[26] So we might expect a potentially lengthy delay before AI boosts productivity, even in advanced economies. That said, given the speed at which many people are adopting LLMs and the like, and the ease with which AI appears to be adaptable to even sophisticated tasks, productivity improvements may come more quickly this time.

---

22 There is a broader discussion in computer sciences about the ability of computers to generate pure random numbers, touching upon the idea of pseudorandomness.

23 See Garicano (2024).

24 See Crafts (2021).

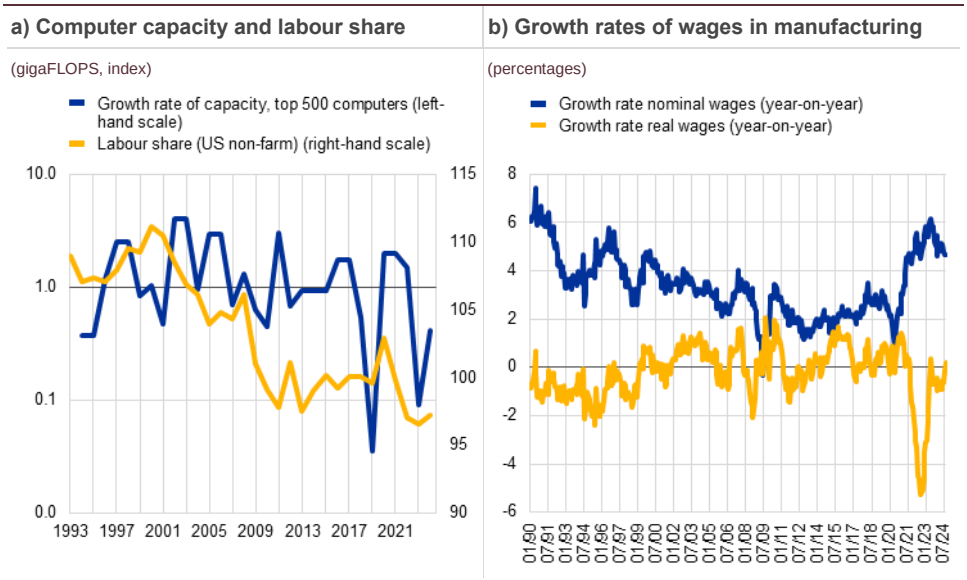
25 As recently as 40 years ago, Robert Solow (1987) quipped that "You can see the computer age everywhere but in the productivity statistics."

26 Unlike overinvestment in technologies such as railroads or fibre optic cables, overinvestment in AI may not create tangible infrastructure with long-term social benefits.

AI may affect the allocation of labour in the economy and indirectly financial stability. Aldasoro et al. (2024) suggest the possibility of a scenario akin to that in some science fiction stories like Kurt Vonnegut's novel *Player Piano* (1952). Vonnegut envisions a dystopian world in which automation replaces all human workers. The result is class conflict between wealthy, highly skilled engineers who maintain the machines and the lower classes who have nothing to do. While Vonnegut's vision is surely extreme, we can imagine that AI will shift the relative value of various skills, creating demand for new ones while making others obsolete. The shift from blacksmiths forging horseshoes to mechanics servicing automobiles in the early 20th century is a case in point. But as the reallocation of labour and capital occurs, with a possible decline in real wages and the labour share of national income, the result could be political discord both within and among countries. Panel a) of Chart 2 shows that since 2000 the share of income accruing to labour in the non-farm sector of the US economy has declined by over 10%. Simultaneously, computer capacity (shown on a logarithmic scale) has increased around 100% annually since 1993. Over a similar period, the annual growth rate of real wages never exceeded 2% and was negative in almost half of the periods. Box 2 discusses briefly the impact of widespread use of AI on labour markets, an area of research that has grown substantially in the last few years.

Chart 2

Computer capacity, labour share and real wages



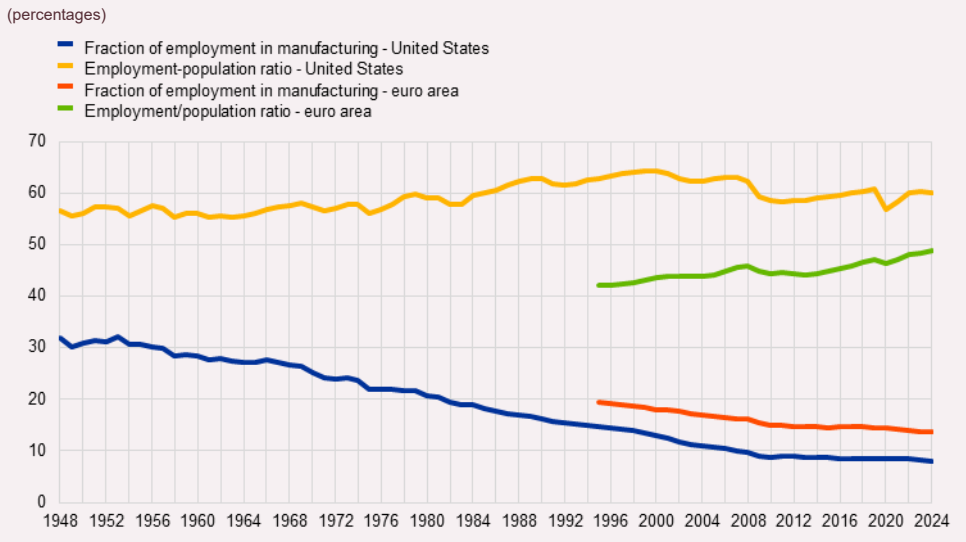
Sources: Dongarra et al. (2024), retrieved from World in Data, U.S. Bureau of Labor Statistics, retrieved through FRED, OECD, Haver Analytics and ESRB Secretariat calculations.
Notes: The scale of the growth rate in the capacity of the top 500 computers is logarithmic. Labour share refers to US non-farm and is the ratio of labour compensation paid to current dollar output. Nominal and real wages cover all member countries of the Organisation for Economic Co-operation and Development (OECD) for the manufacturing sector. Nominal wages are indexed (2015 = 100); real wages are computed after deflating with a price indicator for all OECD member countries.

Box 2
The impact of AI on labour markets

The distribution of employment across sectors is constantly changing. At the dawn of the industrial revolution roughly three-quarters of the European and US labour force worked in agriculture. As a result of centuries of technological progress it now takes less than 2% of advanced economy employment to feed an entire nation. Similarly, since the end of the Second World War, productivity improvements have dramatically reduced the proportion of the population employed in manufacturing. Chart A displays the trend in manufacturing employment as a fraction of total employment (in dark blue) for the United States from 1948 to 2023, and for the euro area since 1996. In the United States the proportion was nearly one-third in 1948, falling steadily until 2008, when it stabilised at about 8½%. Other advanced economies, like those in the EU, follow a similar path.[27] One of the many reasons for this shift lies in increased automation brought on by the dramatic rise in the speed and efficiency of computers. However, this technological progress has had virtually no impact on the employment/population ratio (shown in yellow for the United States and in green for the euro area). Historically, innovation has redistributed employment, not reduced it.[28]

Chart A

Manufacturing employment as a share of total employment and the employment/population ratio, United States and euro area, annual



Sources: U.S. Bureau of Labor Statistics, Eurostat and Haver Analytics.
Notes: For the United States, manufacturing employment share is the ratio of all employees in manufacturing to the sum of employees in private industry plus those in government. For the euro area, employment includes self-employed persons.

[27] According to data compiled by Maluquer de Motes (2021) for Spain, the share of employment in agriculture in Spain exceeded 70% in 1850 and started to decline around 1910, reaching less than 10% in 1994. Similarly, the employment/population ratio has remained generally stable, albeit with a slight decreasing trend, since 1850. The share of employment in manufacturing peaked around 1980 at 35% and has decreased below 20% since then.

[28] Related to the sectoral redistribution noted in the text, it is important to emphasise that despite the fairly constant employment/population ratio, other demographic redistributive effects have occurred, notably the increase of women in the labour force and the tendency towards earlier retirement.

Various estimates of the impact of AI on labour have been made. Gmyrek et al. (2023) analyse 436 occupations, identifying those most likely to be affected by AI. The authors distinguish four groups: those least likely to be impacted (mainly composed of manual and unskilled workers), those where AI will augment and complement tasks (occupations such as photographers, primary school teachers or pharmacists), those where it is difficult to predict (amongst others financial advisors, financial analysts and journalists) and those most likely to be replaced by AI (including accounting clerks, word processing operators and bank tellers). They conclude that 24% of clerical tasks are highly exposed to AI, with an additional 58% having medium exposure. For other occupations, roughly one-quarter are medium-exposed. By contrast, Cazzaniga et al. (2024) suggest that up to 60% of jobs in advanced economies are exposed to AI, which they argue will affect income and wealth inequality.[29] Acemoglu (2024) estimates that 20% of labour tasks in the United States are at risk of displacement by AI.

Over the long term, large disruptions in labour markets caused by the adoption of AI may impact the demand for financial services and financial stability at large.[30] While the adoption of AI may not diminish employment in general terms, there is substantial uncertainty about the distributional impact. Will working conditions and wages improve or worsen, and for whom? Will labour force participation shift?[31] Will there be large reallocations in the workforce? Will there be widespread defaults by companies that fail to adopt these technologies quickly? Additionally, how will any AI-induced shifts alter the overall return on higher education? These structural shifts in labour markets and returns on education, although highly speculative now, have the potential to (i) increase inequality and populism, (ii) lead to a deterioration in borrowers' debt servicing capacity, and (iii) change the demand for financial services, ultimately increasing systemic risks.

Many technologies have very high fixed costs but low marginal costs. That is, after a significant initial investment, scaling is cheap. The result is that inventors and early adopters gain market power with the accompanying ability to exploit customers. Do generative pre-trained transformer (GPT) models have these properties? We know very little about the marginal costs of operating and maintaining generative AI systems, but we do know that development costs are extremely high. Chart 3 shows the increasing costs associated with the hardware and energy costs of training AI models. Recent models such as ChatGPT 4 and Gemini 1.0 Ultra cost more than USD 100 million to train.[32] Then they need constant updating to stay relevant, requiring substantial processors, computer time and energy/water. Ignoring the updating and training expense, costs borne by users appears be quite low. Pricing of AI tools currently seems to rely on scale, that is increasing the number of customers who pay a small fee every time they use the services (and they should use them frequently). For

[29] See also Comunale and Manera (2024) for a summary of academic work in this area.

[30] See also Aldasoro et al. (2024), Korinek (2024) and Videgaray et al. (2024).

[31] When looking at the level of education in the population of the United States, labour force participation has been decreasing since 1993 across all levels, except for those who did not finish high school, according to data from the Bureau of Labor Statistics.

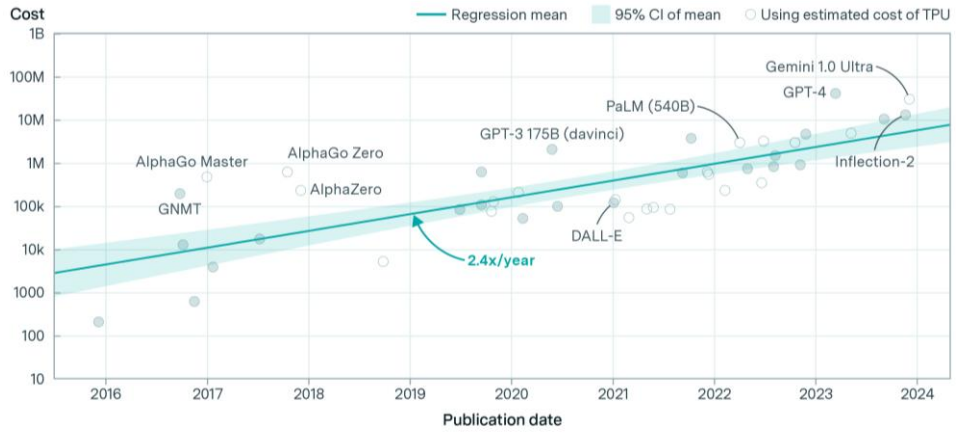
[32] See the Visual Capitalist website.

example, in August 2025 Open AI listed its current pricing for GPT-4o at USD 5 per 1 million input tokens. If a typical prompt is 50 words, this implies a cost of between 0.0025 and 0.005 cents per query.[33] We have no way of knowing the relationship of this to Open AI's marginal cost. Regardless, the costs to consumers seem relatively low.

Chart 3

Amortised hardware and energy cost to train frontier AI models over time

(US dollars at 2023 constant values, log scale)



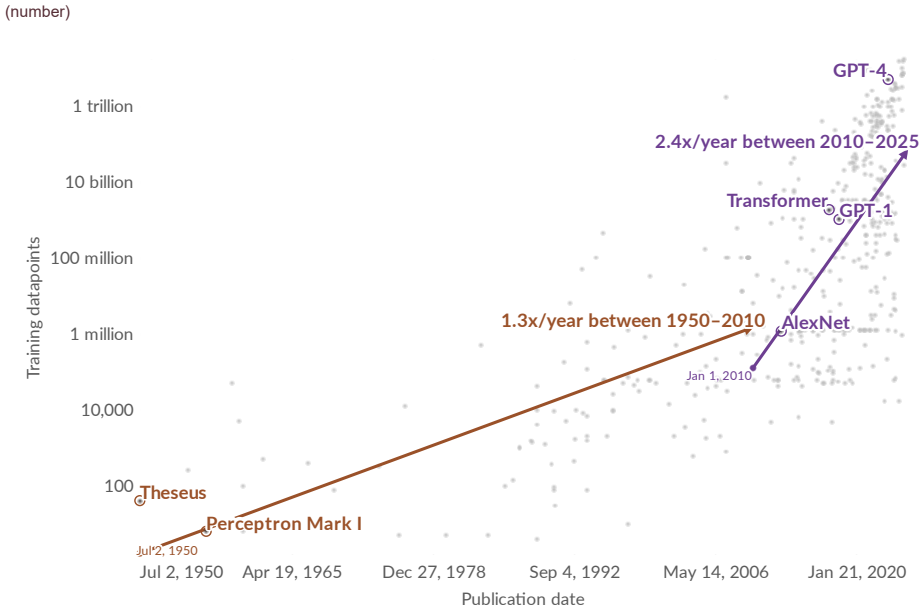
Source: Cottier et al. (2025).
Notes: The models selected are among the top ten most compute-intensive for their time. Amortised hardware costs are the product of training chip hours and a depreciated hardware cost, with 23% overhead added for cluster-level networking. Open circles indicate costs which used an estimated production cost of Google TPU hardware. These costs are generally more uncertain than the others, which used actual price data rather than estimates.

Generative AI models require a continuous flow of training data to remain current. The latest versions have billions of parameters trained using trillions of "tokens" (a token is roughly an English word). As we noted earlier, these are enormous statistical models that generate probabilistic output. Like the fitted values for a simple linear regression model, when a Generative AI model produces a sentence or a paragraph it is estimating the most likely words or phrases. In theory these models should continue to improve in accuracy as their training data expands and the number of parameters increases. One of the appealing aspects of LLMs is their ability to take a general pre-trained foundation model and fine-tune it for performance on specific tasks. In some cases, the process achieves performance levels better than humans can manage. As seen in Chart 4, over the past 75 years the quantity of data fed into these models has grown exponentially.

[33] See the OpenAI website.

Chart 4

Training datapoints used in AI models



Source: Epoch AI (2024), retrieved from World in Data.
Notes: Training datapoints are a representation of the number of examples the AI model learns from during its training process. Each domain has a specific data point unit; for example, for vision it is images, for language it is words, and for games it is timesteps. This means it is only possible to directly compare systems within the same domain.

In addition to its ability to resolve specific tasks, AI can match or surpass human cognitive capabilities across several cognitive tasks. People often refer to the latter as "artificial general intelligence", a concept that is close to "general-purpose AI" (which refers to AI systems able to perform a wide variety of tasks). An early landmark in the competition between humans and computers was the victory of Deep Blue over chess champion Gary Kasparov in 1996. More recently AI has surpassed human performance in activities such as reading comprehension or visual reasoning and is almost at par in competition-level mathematics (Chart 5).[34] Some researchers conclude that the creative abilities of ChatGPT4, including its capacity to generate original output, seem to match those of humans.[35] Similarly, in some tasks and in some circumstances AI chatbots appear to outperform human participants.[36]

[34] See Stanford University (2024).

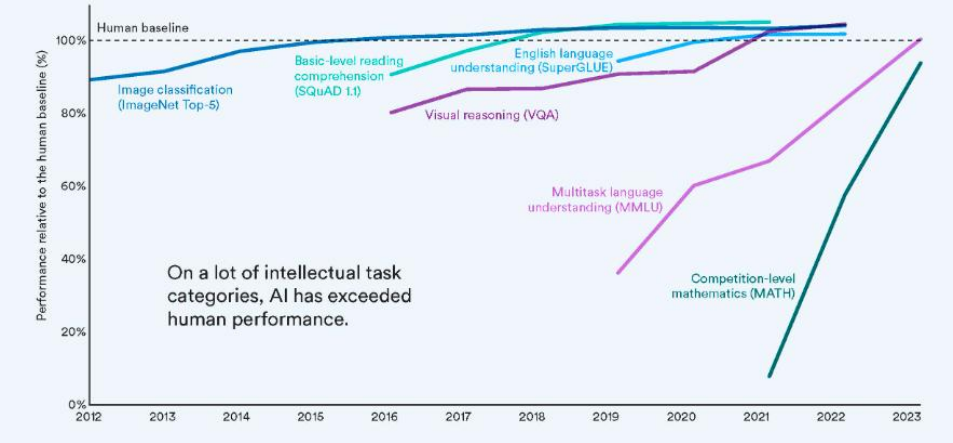
[35] See Guzik et al. (2023).

[36] See Koivisto and Grassini (2023).

Chart 5

AI technical performance versus human performance

(percentages)



Source: Stanford University (2024).

There is large uncertainty about the pace of AI advances.[37] New breakthroughs may increase the capabilities of AI, address the issue of the enormous amount of data needed to generate these systems, and reduce the currently high costs of maintenance. But will they? Bengio et al. (2025) cite four main constraints on the development of AI (power constraints, chip production capacity, data scarcity and latency walls) but estimate that these will not impede the current pace of development of AI until 2030, with large uncertainty after that.[38]

Prominent voices in the development of AI have already raised concerns about the potential for disruption created by artificial general intelligence. As a result, many computer scientists are advocating regulation of further AI development. Over 600 researchers and public figures are signatories to the statement on AI risk promoted by the Center for AI Safety: “Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war”.[39]

2.2 AI in finance

Narrowing our scope to the financial system, technological innovation has broad social and private benefits. It reduces costs, increases speed and improves access to financial services. It allows better measurement and pricing of a wider range of risks, allowing for a broadening and deepening of financial markets. Innovation increases the range of available financial instruments, increasing risk-bearing capacity. It allows for higher volumes of transactions, improving efficiency of pricing. It provides increased security and resilience. And it aids monitoring, allowing authorities

37 See Bengio et al. (2025) and Korinek and Suh (2025).

38 For further details, see Sevilla et al. (2024).

39 See Center for AI Safety (2024).

to ensure safety and soundness, protect consumers and investors, safeguard market integrity, reduce criminal uses and promote financial stability.[40]

In many ways, the introduction of AI in finance is just the latest in a long string of such innovations. Taking a very long-term perspective, we can think of many technological developments – both conceptual (software) and tangible (hardware) – that had a profound impact on finance. These include (i) double-entry bookkeeping in accounting in the 13th century, which allowed a precise record of assets and liabilities and profits and losses for the first time; (ii) the printing press in the 15th century, which enabled rapid and wide dissemination of information in written form; (iii) the creation of the East India Company in the 17th century, making it possible to trading in small claims on the revenue of a firm;[41] and (iv) the telegraph, which increased the speed of communication across continents and expanded banks' activities beyond their countries of origin.[42] Similarly, computers have substantially increased the capacity to process information and the speed of communication. The development of high-frequency trading, online banking and trading platforms would be examples of how computers have affected finance. We are already seeing similar gains from AI in terms of speed of communication and increased productivity.

The scope, scale and delivery of modern financial services is largely a result of innovations over the past sixty years. Arner et al. (2015) see the launch of the ATM in 1967 as the starting point of a new era in finance. Over the following twenty years we saw broad adoption of credit and debit cards, an explosion in various types of derivative instruments (including securitisation pools), the introduction of money market funds, the launch of electronic trading platforms such as Nasdaq, and many more. In Table 1 we reproduce a list of innovations from Silber (1983) that took place between 1970 and 1982, including the triggers to which he links them.

[40] Daníelsson et al. (2022) note that microprudential authorities can use technology, including AI, to monitor financial institutions individually, under a detailed rulebook. Microprudential authorities look at the day-to-day operations of financial institutions, using large amounts of standardised data to monitor adherence to the rules.

[41] See De la Vega (1688) for a reflection on trading activities at the Amsterdam Stock Exchange.

[42] See Lin et al. (2021) for a quantification of the impact on banking that the introduction of the telegraph in China had.

Table 1
Financial innovations between 1970 and 1982

	Inflation			Volatility of interest rates	Technology	Legislative initiative
	Level of interest rates	General price level	Tax effects			
A. Cash management						
1. Money Market Mutual Funds	✓					
2. Cash management / sweep accounts	✓				✓	
3. Money market certificates	✓					✓
4. Debit card	✓				✓	
5. NOW accounts	✓					
6. ATS accounts	✓				✓	
7. Point of sale terminals					✓	
8. Automated clearing houses					✓	
9. CHIPS (same day settlement)					✓	
10. Automated Teller Machines (ATM)					✓	
B. Investment contracts						
(i) Primary market						
1. Floating rate notes				✓		
2. Deep discount (zero coupon) bonds	✓		✓	✓		
3. Stripped bonds	✓		✓	✓		
4. Bonds with put options or warrants	✓			✓		
5. Floating prime rate loans				✓		
6. Variable rate mortgages				✓		
7. Commodity linked (silver) bonds				✓		
8. Eurocurrency bonds	✓					
9. Interest rate futures				✓		
10. Foreign currency futures						
11. Cash settlement (stock index) futures						✓
12. Options on futures				✓		✓
13. Pass-through securities						✓
(ii) Consumer-type						
1. Universal life insurance				✓		
2. Variable life policies		✓				
3. IRA/Keogh accounts			✓			✓
4. Municipal bonds funds			✓			✓
5. All-saver certificates	✓					✓
6. Equity access account		✓		✓		
C. Market structures						
1. Exchange-traded options						
2. Direct Public Sale of Securities						
Green Mountain Poer Co.				✓		
Shelf registration				✓		✓
3. Electronic trading						
NASDAQ					✓	
GARBAN					✓	
4. Discount brokerage						✓
5. Interstate depository institutions					✓	✓
D. Institutional organisation						
1. Investment bankers / commodity dealers	✓			✓		
2. Brokers / general finance						
3. Thrifts with commercial banks	✓			✓		✓
4. Financial centers (Sears Roebuck)						

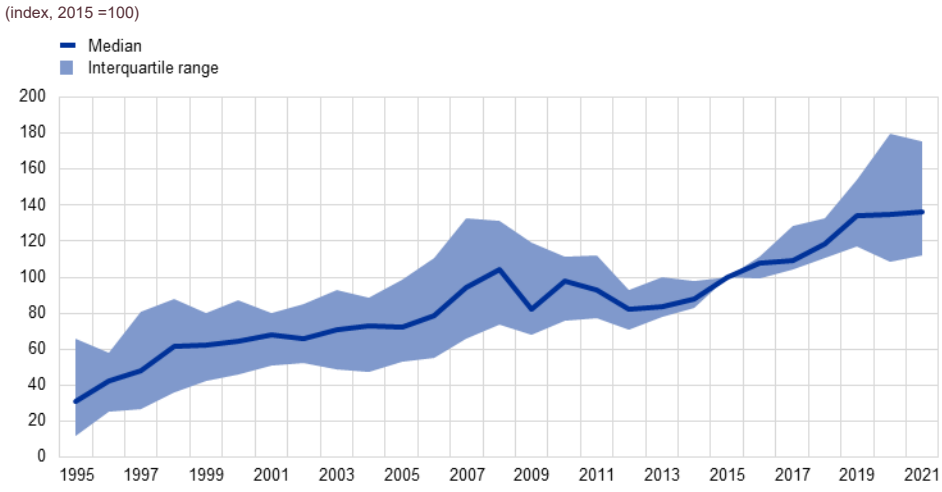
Source: Silber (1983).
Note: For readability purposes, we have removed various columns from the original, including those on internationalisation.

In addition to AI, smart phone technology, the internet and application programming interfaces (APIs) are major sources of financial innovation.[43] The increased digitalisation of advanced economies over the last 30 years is affecting the way financial institutions produce and provide services to their customers, as well as bringing new fintech and big tech players into the production and provision of financial services. Advances in telecommunications and information technology enabled a large increase in the capacity to process information and connect economic agents, triggering a wave of digitalisation.

Financial intermediaries are very heavy users of information technology (IT). Consequently, they have very high IT-related expenses. Chart 6 documents the four-fold increase in gross investment (gross fixed capital formation) by European financial and insurance corporations (including banks) in computer software and databases over the past 30 years. Some estimates put US banks' IT spending at more than USD 100 billion per year, with the eight designated G-SIBs spending around USD 60 billion. By comparison, based on data from the European Banking Authority, we estimate that EU banks spend USD 40-50 billion per year.

Chart 6

Gross fixed capital formation in computer software and databases (chain-linked volumes), financial and insurance corporations



Source: Eurostat and ESRB Secretariat calculations.
Note: Data for Belgium, Czech Republic, Estonia, France, Italy, Latvia, Lithuania, Luxembourg, Netherlands, Austria, Romania, Finland, Slovenia and Sweden.

2.3 The use of AI in the EU financial sector

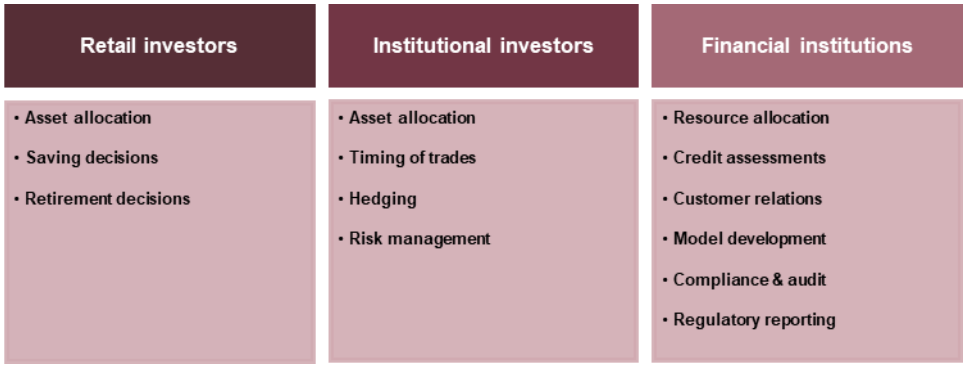
Supervised financial institutions, retail and institutional investors, and the authorities can use AI. Generic uses include drafting and summarising large bodies of information, simulating complex systems and aiding in decision making. Figure 2 lists some potential uses of AI by different private-sector groups participating in finance. For regulatory and supervisory authorities, potential uses include (i) fairer,

[43] See Beck et al. (2022).

faster, cheaper and more equitable monitoring, (ii) earlier warnings, (iii) reverse stress testing and multiple scenarios in stress tests, and (iv) simulations using large network models.[44] Overall, the hope is that authorities can use AI to improve their capacity to analyse the vast amounts of supervisory and market data available, thus making their risk assessments timelier and more accurate and increasing financial stability and resiliency.

Figure 2

Uses of AI in finance



Source: Authors' elaboration.

At the time of writing, adoption of AI by EU banks is only partial. Looking at the far-right column of Table 2, we see that as of spring 2023, most EU banks were not using or planning to use AI in the short term.[45] Among those that do, decision trees, random forest models and regression analysis are the most widely used. Remarkably few EU banks report using any form of natural language processing (NLP), including LLMs, despite the widespread attention such methods have received. Recent data from the European Banking Authority shows that banks apply AI most often to activities including customer support, anti-money laundering, fraud detection, and profiling and clustering of clients or transactions.[46]

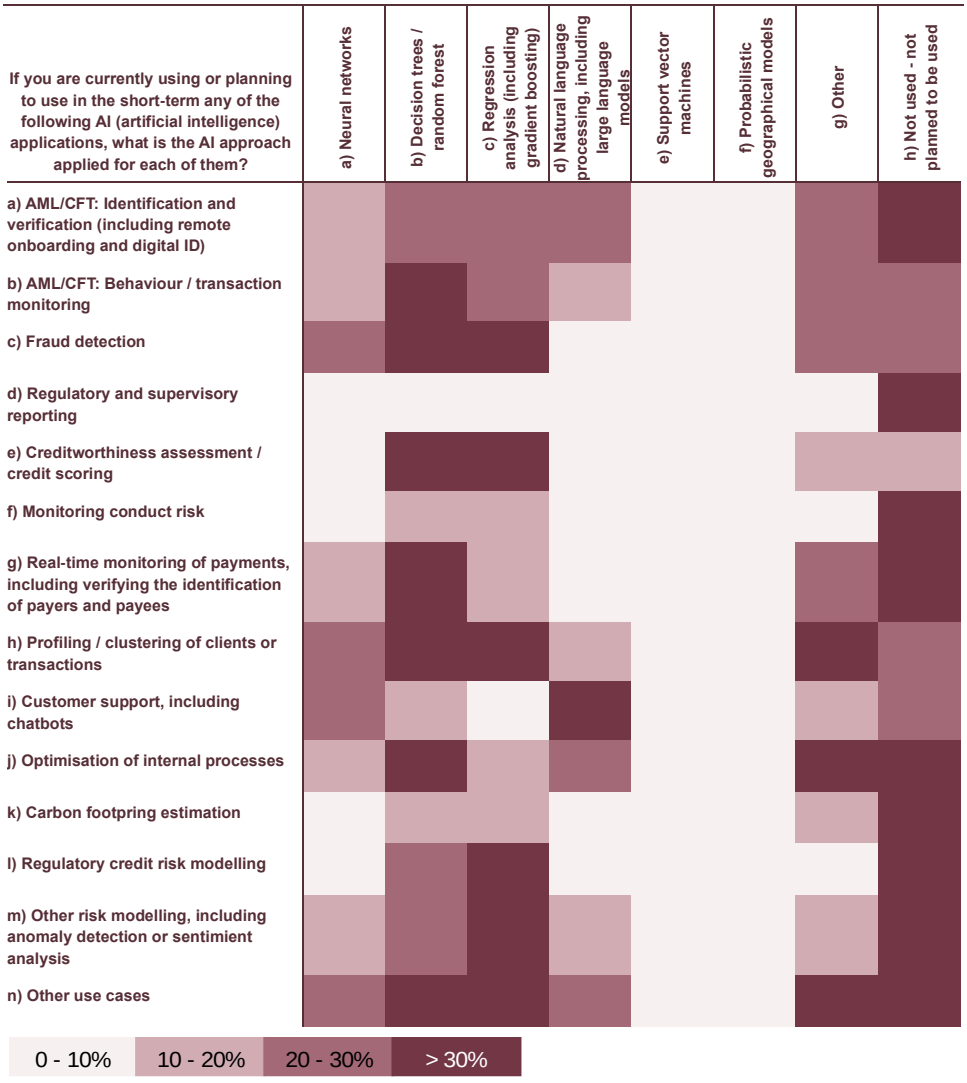
[44] See O'Halloran and Nowaczyk (2019), Flood et al. (2020), Daníelsson et al. (2022), Wever et al. (2022) and Petrone et al. (2022).

[45] It is worth noting however, that as AI is moving very quickly, surveys from spring 2023 are unlikely to give an accurate picture of the state at the time of writing this document (autumn 2025). Data from the EBA's 2024 report paint a similar picture, although not presented in as accessible a format; hence we present the 2023 version.

[46] See European Banking Authority (2024).

Table 2

Uses of AI by banks



Source: European Banking Authority (2023), redrawn by authors for clarity.
Notes: Table based on the responses to Question 28 of the questionnaire for a sample of 85 EU banks. The definition of AI used by the European Banking Authority may differ from the one in Bengio et al. (2025).

Other types of financial institutions also use AI. According to the European Securities and Markets Authority (2025a), asset managers use AI, including LLMs, primarily to support human-driven investment decisions. Those few investment funds that actively market their use of AI usually integrate it into systematic investment strategies, but they have not performed any better or worse than peers and do not charge higher fees. Besides, these funds, which tend to be active funds investing in equity markets, have experienced outflows in some periods. Regarding insurance, already in 2024, 50% of the respondents surveyed by EIOPA were using AI in non-life insurance lines of business and 24% in life insurance lines of business, with many more planning to adopt it within the next three years.[47]

[47] See European Insurance and Occupational Pensions Authority (2024).

As AI adoption proceeds, how will it influence finance? Will it amplify existing systemic risks, create new ones, or both? We now turn to these questions.

3 AI and systemic risk

Does AI amplify existing sources of systemic risk for financial stability, create new ones, or both? The introduction of any new technology can influence systemic risk. For example, securitisation creates the potential for concentration and transfer of the risks in banks' balance sheets. The advent of high-frequency trading brings the potential for flash crashes. And pervasive communication through smartphones and social networks enables accelerated deposit withdrawals, amplifying the risk of bank runs and panics. While it is too early to tell what AI might bring, in this section we discuss aspects of it that might lead to greater systemic risk.

As in earlier cases, many of the sources of systemic risk we identify are not inherent to technological innovations. They stem from corporate and societal choices on how to deploy them. Current versions of AI are well-suited for specific tasks or problems.[48] For example, an AI assistant may be able to mimic the way a human decision-maker uses a flow-chart to navigate through a body of information. And AI can process far more information far faster than any human. Furthermore, given a precise set of objectives, AI-based decision making will likely be more accurate and consistent with the task at hand. What remains less clear – at least at the time of writing – is how AI will perform when employed to assist with larger, less well-defined tasks.

Turning to the financial sector, institutions can use AI to provide a range of services to a variety of end users.[49] This is already underway and has the potential to create systemic risk. It is inevitable that the future development and adoption of AI will lead to new and unforeseen risks. Determining when such risks might rise to the level of systemic importance is obviously challenging. That said, we can examine how the properties of AI are likely to influence systemic risks.

One of the key concepts around the use of AI is trust, which is also key for the functioning of advanced economies and the financial system. Trust is defined as an "assured reliance on the character, ability, strength, or truth of someone or something", and "one in which confidence is placed".[50] It is also the "belief that someone or something is reliable, good, honest, effective, etc."[51] The role of trust in enabling economic transactions and ensuring the smooth functioning of the financial system has been widely analysed in academia.[52] While advances in AI have created more human-like interactions with AI tools, it is easy to forget that the tool is in fact not

---

48 See, for example, Luo et al. (2024). Daníelsson et al. (2022) emphasise this point.

49 Daníelsson et al. (2022) note that "AI is well suited for measuring and managing exogenous risk because it can use large data samples, well-established statistical techniques, and many repeated events to train on while objectives are straightforward." There is also evidence that financial institutions are using AI for model validation, developing early warning systems (Wever et al., 2022), audit efficiency (Xing et al., 2020) and simulation via large network models (O'Halloran and Nowaczyk, 2019, Flood et al., 2020, Petrone et al., 2022).

50 See the Merriam Webster dictionary.

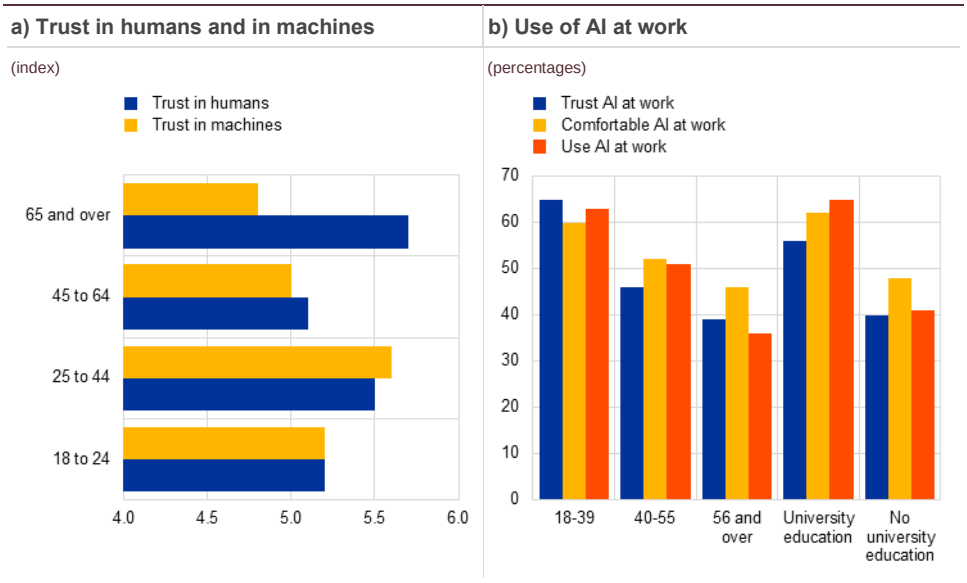
51 See the Britannica Dictionary.

52 See, among others, Mayer et al. (1995) for an early contribution and Tonkiss (2009) for a reflection on trust during the global financial crisis.

human. Recent surveys point to higher trust in AI by younger generations and those with higher education (Chart 7).

Chart 7

Generational and educational differences in the perception of AI



Sources: National AI Opinion Monitor and Gillespie et at. (2023).
Notes: Panel a) shows data from the US National AI Opinion Monitor by age of respondent, based on a population of 4,767 individuals in December 2024. For further details, see Ognyanova and Singh (2025). The data in panel b) are taken from the 2023 Global Study by the University of Queensland and KPMG, covering more than 17,000 respondents in 17 countries (Australia, Brazil, Canada, China, Estonia, Finland, France, Germany, India, Israel, Japan, Netherlands, Singapore, South Africa, South Korea, United Kingdom and United States). For further details, see Gillespie et al. (2023).

In the remainder of this section, we focus on three things. First, we review the sources of systemic risk. Our contention is that we know what these are, so we can make a complete list of the important externalities and spillovers that lead to financial stress. Second, we list the features of AI that can exacerbate these sources of systemic risk. The fact that AI continues to develop means our list is speculative and ultimately incomplete. Finally, we put these two together and ask how each property of AI influences systemic risk.

3.1 Sources of systemic risk

Systemic financial risk arises from externalities and spillovers.[53] **While there are several ways to characterise these, we find it convenient to divide them into five categories: liquidity mismatches and information sensitivity, common exposure, interconnectedness, lack of substitutability, and leverage.**[54] These all ultimately relate to misaligned incentives and moral hazard, which, in part stem from

[53] For more detailed discussions on the concept of systemic risk, see, among others, European Central Bank (2009), Smaga (2014) and Benoit et al. (2017).

[54] Note that this is related to the five intermediate objectives for macroprudential policy enumerated in European Systemic Risk Board (2013): (i) to mitigate and prevent excessive credit growth and leverage; (ii) to mitigate and prevent excessive maturity mismatch and market illiquidity; (iii) to limit direct and indirect exposure concentrations; (iv) to limit the systemic impact of misaligned incentives with a view to reducing moral hazard; (v) to strengthen the resilience of financial infrastructures.

two structural factors of our financial system. First, there are public safety nets in the form of deposit insurance, central bank emergency liquidity support and implicit fiscal guarantees. Second, the limited liability company structure means that losses incurred by owners of financial institutions cannot exceed their investment. Together, these factors give owners and managers incentives to take risk beyond what is socially optimal, while depositors are indifferent to the balance sheet structure of institutions. The result is that risk taking is excessive relative to the social optimum, creating fragilities that make financial crises both more frequent and more severe.

First, liquidity mismatches and information sensitivity arise from the fact that many financial intermediaries issue liquid liabilities and use the proceeds to purchase illiquid assets. When a shock causes liquid assets thought to be risk-free and information-insensitive to suddenly become risky and information-sensitive, this mismatch creates a system prone to runs on banks and markets.[55] Consider the problem of a bank run. Banks issue short-term deposits to finance long-term loans. The existence of deposit insurance and government regulation lead bank deposit holders to be indifferent to the health of their bank, at least under normal circumstances. But if something happens that leads them to be concerned – if these information-insensitive deposit accounts originally thought to be risk-free suddenly become risky and information-sensitive – depositors will run to assets they believe to be safe. But since banks' assets are generally illiquid, they will not be able to meet these withdrawal demands. Furthermore, a run on one bank will quickly spread to others. Something similar can happen in a market for safe assets.[56]

Second, common exposures arise when many institutions or individuals face the same specific risk factor. While the precipitating event can be small, the fact that everyone has exposure to the same shocks can bring the system down. For example, a broad range of institutions may be vulnerable to the same underlying risk, such as a wave of mortgage defaults or the inability to roll over short-term debt. This is equivalent to what happens in a biological system that lacks diversity. In that case, a change in chemistry or climate can precipitate the collapse of the system. In finance, common exposures can be direct or indirect. For example, intermediaries may face direct exposure to a frail institution through financial contracts. Or they may unknowingly face exposure through their counterparties, which are themselves directly exposed to the frail institution. Furthermore, if a vulnerable intermediary disposes of assets in a fire sale, depressing prices and undermining market liquidity, it could damage other institutions. Put slightly differently, common exposures are of two types: with and without interconnectedness.

Third, interconnectedness arises from the fact that financial intermediaries have a complex network of exposures. These can take many forms. The two simplest are a chain system, and a hub-and-spoke system. In the former, we can think of institutions as arranged in a straight line with exposure to the one in front and the one behind. If a single institution defaults, it affects everyone further down the line. In this case, systemic risk results from cascading transmission and cumulating losses. In

---

55 See Dang et al. (2020).

56 A weaker version of this is where people suddenly realise that an asset is riskier than they initially thought.

the latter situation, one large entity is at the centre (the hub), with exposures to a myriad of smaller entities (the spokes). In the simplest case, the spokes only face exposure to the hub, not to each other. In this case – which is also one of common exposures – should the large entity at the hub come under stress, so will all the spokes. Obviously, the interconnectedness of financial intermediaries is much more complex than either of these descriptions. That level of complexity can create challenges, not just for identifying and mitigating risk but for resolving institutions as well.[57]

Fourth, lack of substitutability arises when a critical service has very few suppliers so a failure can pose a systemic risk. Examples of this are all around us. One instance was the faulty CrowdStrike software update pushed to customers in July 2025, which brought down millions of Windows devices worldwide – each requiring manual intervention. Various aspects of the financial system have a similar reliance on a small number of providers. For retail payments, there are the VISA and Mastercard networks. For wholesale payments, there are systems run by central banks. Some derivative markets are heavily dependent on central clearing parties, central nodes of the financial system that cannot fail. And some banks have become “too-big-to-fail” in view of their size and range of activities they perform.

Finally, we come to leverage and the inherent procyclicality of the system. The presence of leverage, understood as a prevalence of debt over own funds on the liabilities side of institutions’ balance sheets, exacerbates the systemic impact of the interactions between financial and economic activity. These are mutually reinforcing and can create adverse feedback loops. In the simplest example, investor complacency in a boom lowers market risk premia. This, in turn, drives up asset prices, investment, consumption and profits, reinforcing the prevailing optimism. Rising asset prices raise wealth and the value of collateral, lowering leverage and making it easier to borrow. Increasing credit supply spurs economic activity further, reinforcing the euphoria. In a bust, this cycle runs in reverse: failing asset prices depress wealth, collateral value falls, bank capital declines, and credit supply shrinks. Leveraged investors and firms scramble to deleverage simultaneously. Because everyone faces exposure to the real economy, a modest local shock can become a widespread bust.

3.2 Features of AI that can influence systemic risk

There are many ways to categorise how features of AI could lead to systemic risk. In the following list, we enumerate the features of AI that we see as most salient from a systemic risk perspective.[58] These are not mutually exclusive or exhaustive. They appear in an order that we will explain below.

[57] Flood et al. (2020) propose ways to assess this resolution challenge for individual institutions, based on their internal ownership structure.

[58] Our list also relies on previous work by Financial Stability Board (2024), Crisanto et al. (2024), Aldasoro et al. (2024), Organisation for Economic Co-operation and Development (2024a), European Banking Authority (2024), Videgaray et al. (2024), Barr (2025) and Bengio et al. (2025).

- **Monitoring challenges:** Fully understanding complex AI systems as they evolve is extremely difficult, if not impossible. AI developers and providers may not have an accurate and predictable overview of their own systems. Users may not be able to assess whether an AI agent is doing what is intended, and authorities may not be able to assess whether systems are fully complying with regulatory and supervisory requirements.

- **Concentration and entry barriers:** Currently, there are thousands of companies using AI models that can be traced back to a very small number of providers.[59] The reason for this is that developing and maintaining these underlying systems is extremely costly. This problem predates AI: around 90% of all personal computers run on either the Microsoft Windows or Apple iOS operating systems.[60] For generative AI systems, one estimate puts ChatGPT's subscription share at over 60%.[61] When making the large investments necessary, these firms may not consider whether the resulting technology will act in an ethical manner in line with societal objectives.[62] While the alternative of unregulated, decentralised AI models (e.g. open-source LLMs) may overcome issues of concentration, this could also evade oversight and create unpredictable systemic exposures.

- **Model uniformity:** If many institutions rely on the same AI model or data source, a flaw in that model or a corrupt data update could lead to simultaneous failure – a form of systemic risk not well captured by the interbank exposure matrices of traditional stress tests. This problem of a model monoculture is not new. For example, we all learn the same financial theory from the same textbooks. Risk managers all strive for best practice. And the top four providers of financial risk management systems account for more than 50% market share, while the top three providers of financial reporting and analytics software have more than 60% market share.[63] So, the fact that there is a high level of correlation between how risk assessments respond to shocks is not at all surprising. But AI may make this form of herding worse, amplifying market reactions to news.[64]

- **Overreliance and excessive trust:** If AI model recommendations generate superior performance in good times, outperforming humans, this could increase trust in AI and may lead to additional risk taking in the financial system (see also Box 3 below). The use of AI may extend from activities where it clearly outperforms humans and generates optimal outcomes to others where results are untested or simply wrong. Furthermore, if a model's training data does not

---

59 See Aldasoro et al. (2024), Comunale and Manera (2024) and Stanford University (2024).

60 See Wikipedia entry entitled Usage share of operating systems.

61 See blog post by Fabio Duarte entitled Number of ChatGPT Users (October 2025).

62 Examples are easy to construct. Knowing that certain market participants react to news in particular way, an AI model may decide to spread false information to move financial markets to its benefit. To ensure compliance with prudential requirements, a model may produce supervisory reports that misrepresent a financial institution's balance sheet. Yet another possibility is that AI could develop complex financial instruments that exist solely to create revenue for its owner.

63 See sense's Financial Risk Management and Financial Reporting webpages.

64 Gensler and Bailey (2020) argue that broad adoption of deep learning may increase this form of financial instability. In contrast, others believe that AI models may make financial markets more efficient as humans are less able to process information correctly. However, this argument only works under the assumption of perfect markets, which does not correspond with reality.

include sufficient information about the adverse tail of the distribution, it can give an unrealistically optimistic assessment of likely future outcomes. Concerns have also been raised recently regarding the potential for AI systems to engage in deceptive behaviour.[65]

- **Speed:** Even prior to the recent advances in AI, we have experienced flash crashes triggered by high-speed transactions, often through algorithmic trading.[66] So far the impact of flash crashes for the financial system has been limited, with prices quickly reverting to previous levels and only a small number of institutions affected.[67] Nevertheless, there is the potential for AI to intensify market crashes by increasing the speed at which other parts of the financial system react to a changing environment, expanding the impact of flash crashes beyond high-frequency traders.

- **Opacity and concealment:** The fact that AI systems are complex and difficult to understand also makes them difficult to monitor. This creates the possibility that individual users may intentionally conceal information from managers, customers or authorities.[68] There may also be instances where employees either unknowingly use AI embedded in software they are using or use it in ways that are not authorised. In this way, AI may facilitate diminished transparency, resulting in delayed identification of systemic risks.[69]

- **Malicious uses:** Related to opacity and concealment, but at a greater level of severity, is the possibility of malicious uses of AI by sources inside and outside a firm. External sources include criminals or hostile state actors able to damage entire systems. For criminals, AI creates the capacity to conduct widespread fraud, attacking a far broader population of potential victims and doing it more effectively.[70] For hostile state actors and others perpetrating cyber-attacks, it can allow the creation of weapons that are more potent, more difficult to detect and more rapidly deployable.[71] These include the threat of AI-enabled attacks on both financial market infrastructure (e.g. payment systems or CCPs) and key financial institutions.

- **Hallucinations and misinformation:** [72] One way to think of AI is as a very large estimated statistical model generating predictions. It may have billions of parameters estimated from trillions of pieces of data, but it is still a statistical model where the output is best seen as a guess. This has two well-known problems. First, the input data used for estimating (training) may not be accurate or representative. Second, the output is probabilistic and may not be correct (in an objective sense). As a result, AI may present as facts information that is false

65 See Hagendorff (2024) and United Nations University (2025).

66 See, among others, Aquilina et al. (2021).

67 One of the major (and first) flash crashes occurred on 6 May 2010, which affected the market in E-mini S&P 500 stock index futures. For further details, see Kirilenko et al. (2017).

68 Beck et al. (2022) additionally consider the possibility that AI models will exhibit unjust or prejudicial discrimination among users of financial services.

69 See Foucault et al. (2025).

70 See Yamin et al. (2021), Fiott (2022) and Mazzucchi (2022).

71 See, for example, National Cyber Security Centre (2024).

72 For a typology of AI hallucinations, see Sun et al. (2024) and Huang et al. (2023).

or misleading.[73] This will be especially damaging when users are not sufficiently knowledgeable to be able to identify that the information an agent is producing is unreliable or when they place excessive trust on it. Malicious (or unaware) users of AI may exploit this feature to spread misinformation through the financial system or society.[74]

- **History-constrained:** Like every model based on existing data, history constrains AI. It is inherently backward looking. When confronted with a task that goes beyond its training data, AI will provide a solution whose accuracy is difficult – if not impossible – to assess, and it will have even greater difficulty assigning an accurate probability to it happening. Leaving aside hallucinations (see above), reliance on past events and data constitutes a perennial challenge for risk managers, which AI may exacerbate. How can you guard against things that are not in the historical record? And should you? This feature of AI becomes particularly relevant when thinking about tail events, for which past data may be non-existent or scarce). On the other hand, AI may increase the capacity to carry out scenario analysis and aid in the identification and estimation of tail events.

- **Untested legal status:** There are numerous cases underway in which producers of intellectual property are suing companies for using information to train AIs without the permission of the owners of the data. We do not know how these will turn out and whether the AI firms (and possibly their customers) will have to pay royalties to the producers of the intellectual property they are using.[75] A wider unresolved legal issue is who has legal responsibility for actions based on AI systems.

- **Complexity makes AIs inscrutable:** The incredible complexity of AI tools means that it is almost impossible to explain (to the public, to users and, possibly even to the AI developers themselves) why they generate any particular output. This has immediate implications for transparency, accountability, regulatory compliance and user trust. In fact, this inscrutability may slow adoption of beneficial tools that would enhance financial stability. However, like many of the things on this list, the challenge is not new. It is often difficult for people to explain some of their actions. There is, though, an important difference between AI and human inscrutability. While the latter can stem from intuition, subconscious processes, emotional processes (including biases, personal experiences and emotions), etc., there are still underlying reasons, even if they are not explainable. By contrast, AI lacks behavioural underpinnings, so its inscrutability stems directly from its complexity.[76]

Before continuing, we include two additional items on our list of AI features that may create or amplify systemic risk. These fall into the category of "potential features of AI".[77] The first is that AI may become self-aware, leading to the loss of

---

73 Examples would be representing a monument in an incorrect setting or making up references to non-existing academic papers.

74 For example, a viral deepfake could cause a bank run with systemic consequences.

75 It is also possible that limiting AI firms' access to information will degrade the quality of output.

76 For a wider discussion on the topic of AI inscrutability, see Zerilli (2021).

77 See Bengio et al. (2025) and the Law Zero initiative.

human control over it.[78] The second is that we become completely reliant on AI to the point where we are no longer capable of doing tasks ourselves.[79] These are different from the challenges and risks associated with speed, opacity and complexity. Instead, in the first case it is the possibility that AI will start to act autonomously, stop responding to human commands (or only feign a response), or alter the objectives humans gave it and start acting unethically in ways that are counter to societal interests.[80] In the second case, we are considering an extreme dependency on AI for a wide variety of tasks, which may have long-lasting consequences for society even if it acts in an ethical manner.

Box 3
The role of trust

As AI tools become ever more human-like, it is tempting to entrust them with increasing responsibility. Simply extrapolating trust in humans may lead to excessive trust in AI, however. Following Mayer et al. (1995), we identify three components of trust: ability, benevolence and integrity. For individuals seeking guidance about their finances, it is important that their advisors have these three characteristics. Furthermore, trust plays an important role when people find themselves in situations they cannot control or do not fully understand. From this perspective, given the increasing complexity of AI systems, there is a possibility that humans will place excessive trust in AIs. Klingbeil et al. (2024) confirm this. They show that when people know that AIs are generating advice, they become overly reliant on it, following it even if it conflicts with available contextual information and is not in their own interests. In addition, while an AI model can perform certain tasks for humans, it is unclear whether the outcomes will be benevolent (for the good of humans) or performed with integrity (according to a set of principles humans find acceptable). The fact that today's AIs are neither able to pursue abstract goals nor adopt human values makes this trust even more concerning.[81]

There is ample academic literature demonstrating differences in the ways humans develop, maintain and lose trust in humans versus how they react to algorithms.[82] In particular, while people are initially more sceptical about input received from algorithms, once they develop trust, humans tend to trust them blindly. This transfer of trust is greater with algorithms than with humans, leading to reduced

78 In this scenario, AI would lead people – both those inside institutions and those who are supposed to monitor them – to believe they control the system, when in fact they do not. Langer (1975) defined the notion of "illusion of control" in psychology, showing that humans hold illusory beliefs in their ability to control the outcome of chance-determined games. Among others, Gai et al. (2019) and Daníelsson (2022) apply this notion to finance.

79 There is a discussion in the management literature on whether technology augments or substitutes for human workers (for further details, see Huseynova, 2024). However, both augmentation and substitution can lead to a process of deskilling, where reliance on AI to undertake an increasing number of tasks diminishes the skills of humans and makes them worse decision-makers over time (Acemoglu, 2021; Korinek, 2024).

80 For examples, see Pastorello et al. (2020) and Daníelsson and Uthemann (2024a) and references therein.

81 See Buckman (2021).

82 See Afroogh et al. (2024) for a recent review of the existing literature on trust and AI.

oversight, even as decisions become more complex. Furthermore, this trust is stronger, the more human-like the algorithm.[83] On the flip side, although trust in algorithms develops more slowly than trust in humans, people also lose trust more quickly and are less forgiving when made aware of an algorithm's errors.[84] Also, despite AI technology striving to produce results that are increasingly human-like, evidence suggests that characteristics that are too-human are undesirable and may lead to a lack of trust.[85]

There is also literature emphasising that whether there is greater trust in humans or algorithms depends on the nature of the decision.[86] Experimental research finds that while *ex ante* study participants express a preference for algorithmic managerial decisions, *ex post* they express greater dissatisfaction with decisions they perceive to be unfair when made by an algorithm rather than by a human.[87] The authors attribute this to empathy extended to a human manager who might be applying a notion of fairness that differs from the worker's own. In other words, it is easier for a person to understand or rationalise a decision they perceive to be unfair when another human makes it than when it is made by an algorithm. A study of trust in and reactions to police decisions made either by humans or algorithms comes to similar conclusions.[88]

3.3 How AI might create systemic risk

We now combine our two lists, the sources of systemic risk and the features of AI. The results are in Table 3. Grey boxes indicate that an AI feature either amplifies or creates this systemic risk and white boxes that there is no apparent relation between an AI feature and systemic risk. The order of the columns and rows is such that the left-hand column is the most prominent, as is the top row. At this point we say nothing about the intensity of the risks that AI is creating. So, it could be the case that an AI feature that contributes to only two types of systemic risk is more dangerous than one contributing to four. We note a few things about this table. First, all but two of AI's features contribute to the systemic risk created by liquidity mismatches and information sensitivity, and the risk of common exposure. Second, monitoring challenges contribute to all five sources of systemic risk, while concentration and model uniformity each contribute to four of the five. The potential features of AI contribute to the same four sources. Third, at the other extreme, untested legal status and complexity/inscrutability contribute to one source of systemic risk each (lack of substitutability, and liquidity mismatches and information sensitivity, respectively).

83 See Cabiddu et al. (2022).
84 See Dietvorst et al. (2015 and 2018).
85 See Hoff and Bashir (2015).
86 See Lee (2018).
87 See Chugunova and Luhan (2024).
88 See Hobson et al. (2023).

Table 3

How current and potential features of AI amplify or create systemic risk

		Sources of systemic risks				
		Liquidity mismatches and information sensitivity	Common exposure	Interconnected-ness	Lack of substitutability	Leverage and procyclicality
Existing features of AI	Monitoring challenges	■	■	■	■	■
	Concentration, entry barriers	■	■	■	■	
	Model uniformity	■	■		■	■
	Overreliance, excessive trust	■	■	■	■	
	Speed, difficult to stop		■	■	■	■
	Opacity, concealment	■	■	■		
	Malicious uses, crime/terror	■	■	■		
	Hallucinations, misinformation	■	■			
	History constrained	■	■			
	Untested legal status				■	
	Complex, inscrutable	■				
Potential features of AI	Self-aware AI, loss of control	■	■	■	■	
	Complete reliance on AI	■	■	■	■	

Source: Authors' elaboration.
Notes: Titles of existing features of AI are red if they contribute to four or more sources of systemic risk and orange if they contribute to three. Potential features of AI are coloured orange to show that they are not certain to occur in the future. In the columns, sources of systemic risk are coloured red when they relate to ten or more features of AI and orange if they relate to more than six but fewer than ten features of AI.

Turning to specifics, we explain why we choose the pattern in Table 3. To do this, we go through each of the AI features again:

- **Monitoring challenges:** Due to monitoring challenges created by the complexity of AI models, the implications of developments in the financial system are in general more difficult to observe. A sudden change in information will increase fragility caused by both the liquidity mismatch and the leverage on institutions' balance sheets. In addition to direct exposures, cross-institution exposures can be indirect and yet common (clients of clients of clients, etc.), implying that interconnectedness is more complex and probably not fully observed.[89] Monitoring challenges also lead to incomplete assessments of whether an institution or activity lacks substitutes in case it collapses, with this issue coming to the fore only when there is a crisis.

[89] A similar challenge applies in the case of assessing downstream models fine-tuned with the same underlying data and based on the same underlying fundamental models; see the model uniformity discussion below.

- **Concentration and entry barriers:** If there are only a few providers for some AI services, they will likely become highly interconnected nodes of the system. Furthermore, lack of diversity among AI service providers may lead institutions to opt for similar risk profiles, based on liquidity mismatches between their assets and liabilities.[90] Since AI service providers may have very different objectives and attitudes toward risk from those that would support financial resilience, their products may implicitly reflect these different objectives.[91] In addition, with the prohibitive cost of developing new AI infrastructure (e.g. introducing a new LLM), there are both significant barriers to entry for potential new AI providers, as well as significant costs for firms that might want or need to change providers, should one experience financial or other distress (e.g. a software glitch or ransomware attack).

- **Model uniformity:** If financial institutions are using the same range of AI agents based on a limited number of foundational models and tools, the lack of diversity means that everyone has similar, correlated exposures. Reliance on a small set of pre-trained models poses risks similar to those of relying on a limited set of existing credit-scoring models. Moreover, broad use of AI may lead to increased uniformity in the reactions to external shocks and events. Similar responses increase the impact of leverage, making the system more procyclical and increasing its overall fragility. Recognition that existing models are malfunctioning may lead to abrupt reactions by institutions, exacerbating the impact of liquidity mismatches and creating information sensitivity.

- **Overreliance and excessive trust:** AI may become widely used as a result of excessive trust placed in it by users, encouraging common exposures (as users follow blindly the output from a model), increasing interconnectedness (as the same AI is used for a wide range of tasks) and hindering oversight (as the tasks entrusted to AI go beyond what it was designed for). Interoperable AI systems across financial institutions and infrastructures can create cascading effects in the event of a systemic shock. Additionally, if perceptions of AI outputs change, for example, when an error is identified, trust in AI may rapidly diminish and information sensitivity can surge, exacerbating liquidity mismatches and leading to runs. The fact that humans are quick to abandon algorithms when they err could exacerbate the lack of substitutability of AI models. Furthermore, with overreliance comes indifference to other possible sources of information, increasing issues related to concentration and lack of substitutability.

- **Speed:** Technological advances that increase the speed of the provision of financial services – including trading, clearing and settlement – can exacerbate issues created by common exposures and interconnectedness. We can imagine AI suggesting similar trading strategies and positions to a wide range of investors. By increasing the speed of reaction to shocks, AI can amplify

---

90 The concentration of AI providers may also slow adoption of AI, as financial institutions may be concerned about both loss of control and the potential for providers to exert pricing power over customers.

91 The fact that these firms are likely to be outside the financial regulatory perimeter as well as operate in jurisdictions other than that of the home supervisor for the institution using them makes it challenging for authorities to monitor and influence the firms.

procyclicality in the system. Increases in speed also make it harder to stop processes when things go wrong. This, combined with lack of substitutes, can generate systemic risk.

- **Opacity and concealment:** If AI models are widely employed in the financial system, large parts of it could rely on decision inputs that are hard to understand and explain, reducing transparency. In these conditions, some economic agents may exploit the capacity of AI to conceal their activity. Concealment can increase common exposures, interconnectedness and the impact of liquidity mismatches. The mechanism would be analogous to what happened with AIG in 2008: it became a very interconnected actor in the financial system with common exposures, but, due to low transparency, most of these exposures were not visible even to a contractual counterparty. When it started to face difficulties, counterparts to AIG sped to close their positions, changing the information sensitivity of their exposures.

- **Malicious uses:** Whether a given technology is good or bad depends on who is using it and how. Bad actors can exploit AI in the same way that good actors can greatly benefit from it. In contrast to other technologies, however, AI has a greater capacity for bad actors to manipulate its users, by exploiting behavioural biases and the trust people place in AI output.[92] When this is widespread, it creates common exposures and interconnectedness. For example, malicious actors may use AI to persuade investors to take a certain position and then bet against it.[93] Furthermore, when the nature and extent of exploitation becomes apparent, the information sensitivity attached to the financial instruments concerned, may change, potentially generating large losses due to liquidity mismatches.

- **Hallucinations and misinformation:** All AI can be subject to hallucinations, generating outputs that are incorrect, nonsensical or fabricated, despite looking plausible. Even if unintended by providers, hallucinations are a potential source of misinformation to which all users of AI face exposure. In the financial system this implies that hallucinations may lead a broad range of economic agents to take similar positions and, when these are discovered, the information sensitivity of the underlying instruments can rise dramatically, potentially leading to fire sales and runs.

- **History-constrained:** Typically, AI models need time and new data to learn about infrequent or not-yet-seen events, such as the adverse tail of the distribution of possible outcomes. In other words, predictions of tail events by AI can be poor or simply non-existent. While advances in AI (i.e. generative AI) may increase our ability to generate predictions outside an observed distribution, such predictions will be untested and noisy, and hence lack accurate occurrence probabilities. The lack of accurate tail risk predictions can favour excessive risk taking (in the form of large liquidity mismatches) and common exposures. The homogeneity of tail risk estimates means that when information about the tail

92 There is an ongoing discussion on how social networks may exploit behavioural biases of humans, which can provide useful insights when applied to AI.

93 This would be like the actions by Citigroup in the electronic bond network MTS in August 2024. For further details, see Financial Services Authority (2005).

arrives (not necessarily when tail events occur), the possibility of runs can surge, creating systemic risks through these liquidity mismatches and common exposures.[94]

- **Untested legal status:** Reliance on a particular AI provider may prove misplaced should legal decisions render it dysfunctional. When only a handful of tools are available, this can generate systemic stress. Examples of adverse legal decisions for an AI provider may include decisions regarding the use of data protected by copyright or private data to train models, or the potential unethical behaviour of self-learning AI algorithms.

- **Complexity makes them inscrutable:** The difficulty in understanding how an AI model performs means that a surprise in its behaviour (or in the environment) or an information shock on its internal functioning (for example, the discovery of a flaw in the code) can trigger a run on positions taken based on output generated by it. In these cases, it may be necessary to unwind excessive liquidity mismatches, potentially resulting in a systemic event.

Finally, should some features of AI materialise, we can expect systemic risks to arise from liquidity mismatches and information sensitivity, common exposures, interconnectedness and lack of substitutability. The lack of control over AI in the financial system can result in high interconnectedness and common exposures, which are not visible to humans. Similarly, losing control of AI may also limit the capacity to develop substitutes and leaves triggers of information insensitivity completely in the hands of AI. In the case of complete reliance on AI, the financial system would be completely dependent on the preferences of providers in key areas such as risk taking (which determines liquidity mismatches and exposures), interconnectedness and the degree of substitution. The preferences of AI developers do not necessarily coincide with the societal optimum, leading to the emergence of systemic risks.

Having discussed of how AI features can amplify or alter existing sources of systemic risk to financial stability, we now turn to policy responses.

[94] See Foucault et al. (2025).

4 Policy implications

We now turn to the issue of financial policymakers' responses to the various ways in which AI can amplify and generate systemic risk. As with nearly all innovations, AI creates a dilemma for the authorities. What should they do, and when? Introducing new regulation too soon can hamper the pace of development. But waiting too long may lead to a loss of control that could hamper risk mitigation and management efforts. As a result, regulators should phase in regulations, remaining flexible to ensure safe innovation proceeds, while also limiting risks.

Before addressing specifics, we should emphasise that the sources of systemic risks linked to the features of AI we identify are not inherent to AI. They stem from corporate and societal choices on how to deploy the technology. The issue is not AI itself, but how both firms and individuals choose to develop and use it. In some cases, social benefits will exceed social costs. For instance, AI may enhance risk management capabilities and governance in institutions, helping to better identify and mitigate risks.[95] It may allow us to manage risks better at the level of individual institutions in ways that benefit the financial system. However, our focus here is on systemic risk and macroprudential policy, so we leave the topic of how AI can enhance internal firm risk management and governance, and how it could influence microprudential supervision, to others.

As a starting point for our policy discussion on AI we look at market failures it has caused. Financial regulation is usually a response by public authorities to the presence of market failures in a particular domain.[96] Market failures refer to situations where the allocation of goods or services in a free competitive market is inefficient. While there is no one-to-one correspondence between market failures and systemic risks, the two concepts are closely related. Systemic risks usually emerge because of market failures related to, for example, information asymmetry,[97] but they can also be the result of interconnections, leverage, bubbles or propagation channels not directly linked to market failures.[98] Addressing the second type of systemic risks with regulation is thus more challenging.

We also note that regulation of AI providers is beyond the scope of EU macroprudential authorities. This is true for two reasons. First, AI providers are not financial institutions.[99] Second, at the time of writing, most AI providers are based outside the EU, introducing space for regulatory arbitrage and additional complexity

---

95 See Bengio et al. (2025).

96 See Nordhaus and Samuelson (2009) for a classical textbook explanation, and Vives (2010) and Acharya et al. (2011) for a discussion on market failures leading to the global financial crisis. Bank of England (2009), Schwarcz (2019) and Gai et al. (2019) also consider market failures in macroprudential policy.

97 See European Central Bank (2009).

98 See, among others, Kortian (1995), Bank of England (2009), Shleifer and Vishny (2010), Martin and Ventura (2018) and Armanious (2024).

99 Financial institutions using AI would fall under the purview of financial authorities.

on how EU regulation may apply to them regarding, for example, concentration or model uniformity.[100]

We organise this section as follows. We start by examining how the features of AI can create market failures which call for policy action. Then we turn to an assessment of whether the main macroprudential tools and how policymakers use them to address systemic risk in the financial system today, as described in Box 4, are sufficient to manage the systemic risks generated by AI. To anticipate our conclusion, the answer is a qualified "yes", but with some refinements. Next, we consider how supervisory authorities need to adjust to a new environment where there is widespread use of AI by both the public and private sectors.

Box 4
Managing systemic risk: existing tools

In recent years regulators have developed an extensive toolkit to address systemic risk in the financial system. Designed to reduce the frequency and severity of financial instability, the tools address risks across institutions and markets. These tools focus on three important sources of systemic risk: common exposures, interconnectedness and procyclicality. Largely intended to ensure the resilience of banks,[101] they restrict various aspects of balance sheet composition. We can divide them into the following categories:[102]

- capital-based measures, including the countercyclical capital buffer, systemic capital surcharges and sectoral capital requirements;
- liquidity-based measures, in addition to the liquidity coverage ratio (LCR) and the net stable funding ratio (NSFR), as well as loan-to-deposit limits that restrict reliance on various sources of funding;
- borrower-based measures that focus on the creditworthiness of the borrower, placing limits on loan-to-value and debt-service-to-income, as well as foreign currency mismatches;
- other balance sheet restrictions such as leverage ratio limits, large exposure and concentration limits and limits on exposure to other financial intermediaries;
- other, less prominent, tools include dynamic provisioning requirements (e.g. through IFRS 9) and various types of transaction taxes.

---

100 This is not a new phenomenon caused by AI. Insights from regulation of online gambling, including links to providers based in third countries, can be useful. For further information on online gambling regulation in the EU, see Commission work in the field of online gambling services on the European Commission's website.

101 The macroprudential policy framework in the EU for non-banks is currently under development, while the ESRB has promoted the development of a system-wide approach to macroprudential policy; see European Systemic Risk Board (2024).

102 For a more complete description, see Yilla and Liang (2020) and Claessens (2015).

For the non-bank sector and markets, the tools include:

- liquidity management tools (LMTs) such as redemption gates, extension of notice period or anti-dilution tools (ADTs), such as redemption fees, swing pricing and dual pricing;[103]
- leverage limits;
- margins and haircuts;
- circuit breakers.[104]

In addition to these prudential instruments, authorities have tools to address systemic risk arising from liquidity mismatches and lack of substitutability. For the former, existing liquidity-based measures may not be sufficient to address the materialisation of risk at the system level. For those cases, there are central bank liquidity backstops that include the lender of last resort, the market maker of last resort, and central bank swap facilities. The first provides funding to solvent but illiquid institutions – possibly beyond just banks.[105] The second encompasses direct purchases of illiquid financial instruments in markets that authorities deem systemic.[106] The third addresses situations of liquidity mismatches in foreign currencies. As for lack of substitutability, there are tools designed to address monopoly power arising from overreliance on only a few suppliers of a specific good or service, which are based on competition policy. In the case of technology, where there can be strong network effects that naturally push toward concentration, this can be difficult.

4.1 AI, externalities, market failures and policies

The presence of externalities and market failures justifies regulation. In the previous section we focused on the links between the features of AI and systemic risk. We now look at how AI-related externalities and market failures may necessitate changes in macroprudential regulations aimed at conduct, competition, financial institution balance sheet composition and anti-crime enforcement. We note that since regulation has to address legal entities, authorities tend to focus on specific markets and businesses. This is why we have separate regulators for banks, insurance companies, pension funds and asset managers, as well as various types of financial markets.

Typically, there are two types of externalities and market failures: those arising from fixed costs and network effects, and those coming from information asymmetries. To this list, we add bounded rationality. While there is no

---

103 See European Securities and Markets Authority (2025b).

104 See European Securities and Markets Authority (2023).

105 The Bank of England's New Contingent Non-Bank Lending Facility, which began operation in January 2025, is a recent example.

106 See the discussion in Buiter et al. (2023).

consensus among economists on the full list of market failures, the main two are imperfect competition and market power (which arise from fixed costs and network effects), and principal-agent, conflicts of interest and signal extraction problems (which arise from information asymmetries).[107] Given its impact, we add to these bounded rationality, which we understand to mean the limitations imposed on economic agents by their cognitive abilities, the information available to them and the time they have to make a choice, which can lead them to make "good enough" decisions, instead of optimal ones.[108]

Table 4 links the features of AI that we discuss in Section 3 with externalities and market failures. In the description that follows we also discuss which type of policies may seem better suited to addressing the intersection of features of AI and market failures. Similar to Table 3, grey boxes indicate that an AI feature is linked to market failures and externalities and white boxes that there is no apparent relation between an AI feature and market failures and externalities.

Table 4

Linking existing features of AI to externalities and market failures

		Externalities and market failures		
		Imperfect competition	Information asymmetries	Bounded rationality
Existing features of AI	Monitoring challenges		■	
	Concentration, entry barriers	■		
	Model uniformity	■		
	Overreliance, excessive trust		■	■
	Opacity, concealment		■	
	Malicious uses, crime/terror		■	
	Speed, difficult to stop	■	■	
	Hallucinations, misinformation		■	■
	History constrained			■
	Untested legal status	■	■	
	Complex, inscrutable		■	

Source: Authors' elaboration.
Notes: Titles of existing features of AI are red if they contribute to two externalities and market failures. In the columns, externalities and market failures are coloured red when they relate to more than five features of AI.

Several features of AI may lead to large firms having dominant market positions. Concentration of providers of AI and the high existing fixed costs to enter the market may act as a barrier for new corporations, impeding competition. Incumbents may exercise market power over users of AI, who may not have

[107] For a longer discussion on market failures and externalities, often leading to different classifications, see, among others, Bank of England (2009), Nordhaus and Samuelson (2009) and Vives (2010).

[108] The concept of bounded rationality is key in the field of behavioural economics. For further information, see, for example, Conlisk (1996), Kahneman (2003) and Baker and Wurgler (2013).

alternatives. This, in turn, can exacerbate geopolitical tensions and create risks to sovereignty. Moreover, the speed at which an AI-enabled financial system operates may leave less capable participants out of the market, as typically in IT the winner takes all.[109] Finally, uncertainty about some legal issues of AI may also favour larger corporations over smaller firms.

We identify eight features of AI that may exacerbate existing information asymmetries, or create new ones: (i) they can create monitoring challenges, exacerbating principal-agent problems; (ii) they can reinforce overreliance on AI, as users may not get access to all the necessary information; (iii) they create opacity, allow for concealment and result in conflicts of interest; (iv) they also allow for malicious uses; (v) users may not have the means to identify AI hallucinations, making them prone to acting on misinformation; (vi) information asymmetry exacerbates legal problems, (vii) the opacity in how AI tools work and the related complexity can generate a large information asymmetry between those developing the tools (i.e. AI providers) and everyone else, limiting the extent to which verification of AI outputs is feasible; and (viii) speed exacerbates all these problems. However, there are potential benefits; by easing access to information and making technology generally available, AI could contribute to reducing existing information asymmetries, particularly in the area of signal extraction.[110]

The information asymmetries arising from AI also lead to conflicts of interest with the wider interests of society, raising concerns about the impact on labour markets and local communities, the geopolitical landscape, the use of data and the climate footprint.[111] Conflicts of interest arise from divergences between the objectives of society and those of AI providers and are associated with overreliance and trust, malicious uses, speed, untested legal status and complexity. An overreliance on AI may create a large divergence between the corporate interests of AI providers and those of society as a whole, impacting labour markets and local communities substantially. Similarly, hostile geopolitical actors may use AI to further pursue their objectives. Legal ambiguity regarding the use of data creates additional risk; in addition, the environmental impact of the race to build increasingly large systems to meet anticipated demand for AI services is raising concern.[112]

Returning to the table, hallucinations, combined with reliance on past data, interacts with bounded rationality to increase the likelihood of overreliance on AI. Trying to overcome their own limitations, economic agents may rely excessively on AI, assuming that it can take better economic decisions when given a full set of information. AI may also be subject to limitations in its rationality, however, and not be able to process the same type of information as humans.[113] Similarly, bounded rationality may make users more likely to accept AI hallucinations and misinformation

---

109 Regarding winner-takes-all in technology markets, see Barwise and Watkins (2018).

110 See Marwala and Hurwitz (2015) for an initial introduction to the topic. De la Peña and Granados (2024) describe how the application of AI methodologies could improve the economic conditions of small cocoa farmers in Colombia.

111 See Annex 2 for a detailed discussion on our assessment of externalities created by AI.

112 See, among others, this blog post by Paul Krugman.

113 On this topic, see Ma and Su (2024) for a discussion of what they name “superficiality” and “deceivability”.

as they trust the apparently rational outputs generated by AI that surpass their own capacities. Lastly, bounded rationality may exacerbate the difficulties of AI users in recognising the extent to which the past informs AI models, leading to situations where unlikely events (i.e. black swans) or new ground-breaking factors are discarded from economic decisions.[114]

We see a need to ensure that policies are in place to address the market failures related to AI we have identified and call for a mix of policies that balances competition and consumer protection, complemented by adjustments to prudential regulation and supervision. Typically, competition and consumer protection policies address issues of imperfect competition, while consumer protection and transparency policies address information asymmetries.[115] Improperly regulated AI creates social costs, meaning that society will not be able to fully enjoy the potential benefits of AI safely.[116] Finally, the fast-changing nature of AI may require policies to evolve at a comparable speed. That implies a substantial effort by public authorities, which may need to revisit their policies frequently to catch up with the latest developments in AI.[117]

The global nature of AI calls for global policy coordination. Unlike traditional industrial activities, physical borders do not act as a constraint on the spread and use of AI. This makes it difficult to create competition policies at regional or national levels. Developing global policies for AI is therefore of paramount importance.[118] Among current initiatives, the Hiroshima AI process, under the umbrella of the G7, seems the most advanced in the development of policies for organisations developing advanced AI systems.[119] However, with the current deteriorating geopolitical environment, achieving coordination is a particularly difficult endeavour. Even as we write, not all AI providers are willing to sign the recently issued AI Code of Conduct, increasing uncertainty about its future implementation and effectiveness.[120]

While focused on consumer protection, existing AI regulations are an important step to addressing market failures and systemic risk from the use of AI in finance.[121] The EU AI Act and the current state-level regulation in the United States focus on consumer protection,[122] which is important to address imperfect competition and the information asymmetry created by some features of AI, as discussed above. Furthermore, there are at least two channels through which such regulations can

---

114 The difficulty in recognising the role of the past in informing decisions also applies to human decision-makers, who may take a certain decision based on experience (i.e. past information) or intuition, but is something AI is likely to exacerbate. In contrast, an econometric model largely relies on past information. See Feduzi et al. (2022) for a discussion of black swans in public organisations.

115 See Nordhaus and Samuelson (2009).

116 See Acemoglu (2021) and Bengio et al. (2025).

117 See Aldasoro et al. (2024), Organisation for Economic Co-operation and Development (2024b), Videgaray et al. (2024) and Barr (2025).

118 See Aldasoro et al. (2024), Guerreiro et al. (2024), Organisation for Economic Co-operation and Development (2024b), Videgaray et al. (2024) and Bengio et al. (2025).

119 On the Hiroshima AI Process, see its website. As of December 2024, 55 countries had adhered to it, with notable exceptions being Russia, China and Iran.

120 The EU AI Code of Conduct is available online here.

121 Annex 1 provides a short summary of the main provisions of the AI Act (see also Comunale and Manera, 2024), while Beau (2024) outlines the main practical challenges derived from the implementation of the AI Act. For a recent description of the situation in the United States, see Godoy (2025).

122 For an overview of AI regulations around the world, see this article.

address systemic risk indirectly. First, by limiting the scope of activities performed by AI (removing those perceived as riskiest), authorities could reduce interconnections and common exposures. Similarly, imposing limitations or outright bans on the use of AI for certain activities, consumer protection regulation also can be a vehicle for addressing concerns about excessive trust. While beneficial, we should emphasise that these regulations focus on risks to individuals. By definition, they are insufficient to address most of the sources of systemic risk that we have identified in the previous section.

Prudential regulation and supervision can complement consumer protection and competition policies. It is important to modify prudential regulation and supervision to address market failures caused by AI. Such adjustments also can serve to address those systemic risks stemming from AI features that do not interact with market failures, such as interconnectedness. While, in general, we believe that the advent of AI does not necessitate a full overhaul of the existing regulatory system, it is almost surely necessary to consider significant reforms. Supervisory activities may also need to adapt to a new financial system where AI is broadly used. These are the topics covered in the next two subsections.

4.2 Managing systemic risk from AI: regulation

While they may require recalibration, our view is that existing policy tools are likely sufficient to ensure financial system resilience as use of AI becomes ubiquitous in finance.[123] That said, the fact that the speed, scope and scale of AI may amplify existing systemic risks – increasing both the potential frequency and severity of financial stress and financial crises– means that we may need to recalibrate existing tools, including risk controls and risk management provisions. For example, the calibration of capital and liquidity requirements of financial institutions may need adjusting as AI comes into widespread use. Similarly, Pillar 2 requirements for banks may also be a mechanism whereby microprudential supervisors might envisage certain add-ons should they deem a given bank's use or oversight of AI to be imprudent.[124] Furthermore, it is prudent for regulators to analyse the impact of AI on these areas holistically, considering how their own actions may influence any perceived benefits from the uptake of AI.[125]

A recalibration of capital and liquidity requirements seems necessary to account for the speed, scope and scale factors that AI introduces. Microprudential regulation may need adjusting. For example, capital requirements for operational risk should consider the impact of AI on that risk, including the potential for

[123] This discussion is based on an extrapolation of developments up to the time of writing in autumn 2025. Future developments in AI may induce changes in the financial system that are so substantial that authorities may need to craft an entirely new regulatory approach.

[124] One important aspect of supervisory work focuses on governance risks. In the case of AI this would imply looking carefully at how a supervised entity uses AI and applies a proper set of internal controls to ensure accountability.

[125] Additional considerations include (i) identifying the most effective tools to address systemic risks from AI, (ii) retaining flexibility so requirements can remain as simple as possible, and (iii) evaluating whether there is any possibility of refining international standards on capital or liquidity requirements.

more sophisticated and frequent cyber-attacks. Similarly, regulation on large exposures may need adjustment to account for how AI can lead to both increases in concentration of exposures on the one hand, and better risk management on the other. Analogously, in view of the potential for faster deposit runs in banking, liquidity requirements may require recalibration.[126] The current prudential framework focuses on traditional credit, market, liquidity and operational risks. Addressing risks arising from things like the concentration in AI providers, model uniformity or the potential increase in cyber-attacks may require fundamental changes beyond the current prudential framework.

Looking at financial markets, the speed, scope and scale with which the financial system may operate assuming broad adoption of AI increases the importance of circuit breakers in the regulatory toolkit. Circuit breakers trigger a halt in the trading of a financial instrument as soon as its price or change crosses a predetermined threshold. They are already in operation in many markets, particularly related to high-frequency trading, and are able to reduce volatility in trading once triggered.[127] However, in a financial market where AI is widely used by most market participants, authorities may need to broaden the scope and increase the frequency of circuit breakers, probably using AI tools themselves. For example, they could be employed to prevent a shock from having too broad a reach and affecting too many counterparties. In doing so, authorities should define the triggers in such a way that they minimise societal welfare losses, consider potential market instability when approaching the threshold, and ensure that the structure of the rule does not encourage trading to move to other jurisdictions.[128]

Market regulators should also review how AI may affect insider trading, with an eye towards amending existing regulations accordingly. The current definition of insider trading is the trading of securities by *individuals* with access to confidential or material non-public information about the company issuing those securities. Insider trading and market abuse regulations aim to prevent certain individuals taking advantage of their privileged access to non-public information. The use of AI in financial markets, however, may affect the definition of insider trading, as well as the legal responsibilities of individuals, companies and AI providers.[129] Ultimately, existing regulations on market abuse and insider trading may need to be amended. For example, there are cases where LLMs have used insider information to execute trades and hide this behaviour when interacting with humans.[130]

Central banks may need to consider whether their lending facilities are able to respond to sudden liquidity needs arising from a broader range of institutions functioning at a substantially faster pace.[131] Authorities may may even consider

---

126 As we write, regulators are considering recalibrating the liquidity coverage ratio to considering innovations in how current account deposits can flee using online platforms. See Beck et al. (2024).

127 See Guillaumie et al. (2020), who use regulatory data from the European Securities and Markets Authority to analyse market impacts of circuit breakers.

128 For the design of circuit breakers, see Bongaerts et al. (2024). Chen et al. (2024) discuss how price volatility increases drastically when approaching the threshold for a circuit breaker, increasing the probability of triggering it. See also the survey carried out by Sifat and Mohamad (2018).

129 See also Daníelsson and Uthemann (2024a).

130 See Scheurer et al. (2024) and the explanatory video of their experiment.

131 See Daníelsson and Uthemann (2024b).

using AI to help them manage these facilities at time intervals beyond human capacities.

To address issues of bounded rationality and information asymmetry, conduct authorities may wish to require labels stating when and how financial products use AI. Increasing transparency about the use of AI by financial institutions is important for users of financial services, so they have an idea of the role AI plays in recommendations and decisions. This could result in the addition of explanatory and visible labels to, for example, UCITS (Undertakings for Collective Investment in Transferable Securities) investment funds following strategies determined by AI, market-makers trading in financial markets using AI extensively or insurance corporations using AI to price their products or calculate their technical provisions. This would allow users of financial services to identify clearly when and how someone is using AI.

Finally, as uncertainty around the impact of AI on our societies and the financial sector remains extremely large, "skin-in-the-game" and "level of sophistication" requirements may be useful macroprudential tools. Currently, there are multiple views and scenarios on the impact of AI will have on our societies. A wide range of both favourable and adverse outcomes remains possible.[132] To avoid the negative outcomes, authorities may consider ways that they can ensure AI providers have a stake in the outcome and that institutions using AI are sophisticated enough to understand the risks they are taking.[133] "Skin-in-the-game" requirements for AI providers and "level of sophistication" requirements for institutions using AI could be a way to avoid excessive risk-taking in the use of AI.[134] "Skin-in-the-game" requirements are not new in the financial system: net worth and/or collateral requirements are imposed on mortgage borrowers and investors expect hedge fund managers to commit the bulk of their personal wealth to their own funds. We see the possibility that requirements like these could help to avoid systemic risks arising from information sensitivity or common exposures without imposing unnecessary societal burden.[135] As a precondition, however, there needs to be very clear legal responsibility in cases where AIs are responsible for harmful outcomes.

4.3 Managing systemic risk from AI: supervision

To contain systemic risks arising from AI, supervisory authorities require adequate resources.[136] Regulation alone will not be enough to contain the systemic risks arising from AI, but there needs to be a complete supervisory cycle (including

[132] See a summary in Bengio et al. (2025).

[133] In the EU, the Digital Operational Resilience Act (DORA) applies to critical IT service providers, but only in relation to their cloud computing activities and with the objective of ensuring resilience. It does not contemplate systemic risks created by these service providers.

[134] "Level of sophistication" requirements are like the current "fit-and-proper" framework for bank managers used by microprudential supervisors. For further information see here, among others.

[135] Holding providers responsible requires a regime with more legal clarity than we have today. In such a world it may be tempting for certain financial institutions (insurers) to sell protection against AI risks. It is important that authorities monitor such risk transfers to ensure that they do not become large and concentrated, creating systemic risk outside the traditional financial system.

[136] See also Daníelsson and Uthemann (2024a).

enforcement), where authorities have adequate resources (IT and staff) to keep pace with developments in supervised entities and markets. To put it bluntly, a supervisor using current conventional tools to try to supervise deep learning or reinforcement-learning systems will be blind to emerging risks. We see the official sector in an "arms race" with the private sector. If authorities are to succeed in managing the systemic risks arising from AI, they must keep up.[137] This means having both enough people with the relevant skills and adequate IT resources. Ideally, supervisors should be able to develop their own AI infrastructures rather than relying on existing commercial AI tools.[138] In organisational terms, that may imply allocating responsibilities for AI to financial stability departments, so AI is understood to be more than just an IT topic.[139] Otherwise, supervisors and regulators may struggle to follow developments in the financial system, resulting in excessive risk-taking by certain institutions under their supervision. This, in turn, would increase the frequency and severity of financial crises. Given the fact that these technologies defy national borders, cross-border cooperation (beyond the EU) seems especially important.

In addition, supervisory authorities need to have the capability to monitor the systemic risks linked to AI features. As stated above, some systemic risks (linked to interconnections, leverage, bubbles or propagation channels) are not directly related to market failures. Widespread use of AI in finance may increase both the speed, scope and scale of these systemic risks. It also requires strengthening supervisory authorities' analytical capabilities to monitor interconnectedness and leverage across all the participants in the financial system, and to deepen the understanding of asset price formation and propagation channels. Authorities should also consider the impact of scenarios where different technologies such as distributed ledgers, smart contracts, AI and quantum computing interact. Transparency and a new approach to data sharing may be important milestones in this task.[140]

Supervisory authorities may face challenges in the supervision of AI but need to ensure strong oversight, including enforcement. Many of the features of AI create important challenges for supervisory authorities, which may need to substantially adjust their processes and procedures. This should cover the whole supervisory cycle, starting with regulatory reporting, with reporting entities and supervisory authorities possibly using AI to optimise processes and procedures on their side.[141] Additionally, enforcement that requires non-compliance with regulation to be corrected is an important part of the supervisory cycle. After all, where the

---

137 In the case of high-frequency trading, there have been no apparent systemic consequences from the fact that authorities have not engaged in such an "arms race" with supervised entities. Given the much more pervasive adoption of AI, however, we do not believe that this will be the case here.

138 If supervisory authorities rely on commercial AI tools, there is the risk that they become overly dependent on industry self-reporting and assurances. Something similar happened before 2008 with complex derivatives, as banks often knew far more about their risks than supervisors did.

139 See Daníelsson and Uthemann (2025).

140 See Foucault et al. (2025).

141 See Daníelsson and Uthemann (2025). Benedetto et al. (2025) summarise the experience of applying AI to regulatory reporting at Banca d'Italia.

probability of successful enforcement is relatively low or the sanctions imposed are not meaningful, regulation will be unlikely to be credible, increasing systemic risk.[142]

Within the EU, cross-border cooperation and pooling of resources is critical for effective market surveillance of AI, as mandated by the AI Act, and the supervision of financial institutions regarding their use of AI. Keeping up with the private sector is going to be very costly and resource-intensive for supervisory authorities, which simultaneously face tight budgets and political scrutiny, limiting their ability to invest in top-tiered staff and AI systems of their own.[143] One way to alleviate the costs for individual authorities is to pool resources, taking advantage of what are almost surely major economies of scale in developing, maintaining and implementing effective monitoring, surveillance and supervisory schemes. AI concentration, with only a few key providers, combined with the international dimension (in other words, the difficulty of ring-fencing the provision of AI services to a given jurisdiction) argues in favour of a more centralised or, at least, pooled approach to the surveillance and supervision of AI activities. This may be particularly important if authorities need to test and understand AI models before allowing them to become operational.

Finally, supervisory authorities need to be aware of risks that their own use of AI can generate and the importance of solid governance. AI can be a powerful tool for supervisory authorities, for example, multiplying their capacity to run stress test exercises or enhancing scenario analyses.[144] Given the large concentration in AI providers and the potential for excessive trust in their outcome, ensuring sound governance around AI use within supervisory authorities is important to avoid risks like those identified in Section 3 (such as excessive trust, opacity and concealment, or hallucinations).

[142] See Armour et al. (2016). For example, Berger et al. (2022) show how enforcement actions subsequently decrease systemic risks posed by banks, while Delis et al. (2016) estimate a smaller impact of delaying enforcement actions on banks.

[143] Besides, authorities must compete with large banks, hedge funds and tech companies to attract scarce AI expertise. Without top-tier talent, authorities may struggle to understand the sophisticated AI models used by financial institutions, let alone detect systemic vulnerabilities hidden in them.

[144] See Daníelsson and Uthemann (2024a) and Foucault et al. (2025).

5 Conclusions

AI is both presenting opportunities and creating risks. It can expand human abilities to perform a wide range of tasks, potentially increasing productivity. Improved access to information can empower citizens, improve institutions, help solve complex problems, accelerate scientific progress, improve health care and education, and possibly even reduce inequality and poverty. In the financial system, there is the potential for a host of improvements. Retail investors can improve their saving and retirement decisions. Institutional investors can improve their asset allocation and risk management. Financial institutions can improve credit assessments, customer relations, compliance and regulatory reporting. But these benefits come with a host of risks. We can draw an analogy with aircraft, which developed very quickly after the first flight by the Wright brothers and enabled intercontinental travel. At the same time, aircraft are subject to heavy regulation at global level to avoid crashes.[145]

A number of the features of AI have the potential to create systemic risks. These include concentration of providers, monitoring challenges, the potential for increased model uniformity, increased speed of actions, opacity, speed, and the fact that AIs can promulgate misinformation. These, in turn, can amplify systemic risks, including liquidity mismatches that create runs, common exposures that lead to widespread losses, interconnections creating spillovers, and leverage leading to increases in procyclicality. A key component of the impact of AI on systemic risk is trust (i.e. how much trust humans will place in AI and for which type of tasks). In our assessment, these are systemic risks that may emerge at the current state of development of AI technologies. They also relate to market failures related to imperfect competition and market power, information asymmetries, externalities, and bounded rationality. More disruptive scenarios for the impact of AI on the real economy could create additional systemic risks, albeit these remain at present highly uncertain, and we do not consider them in our assessment.

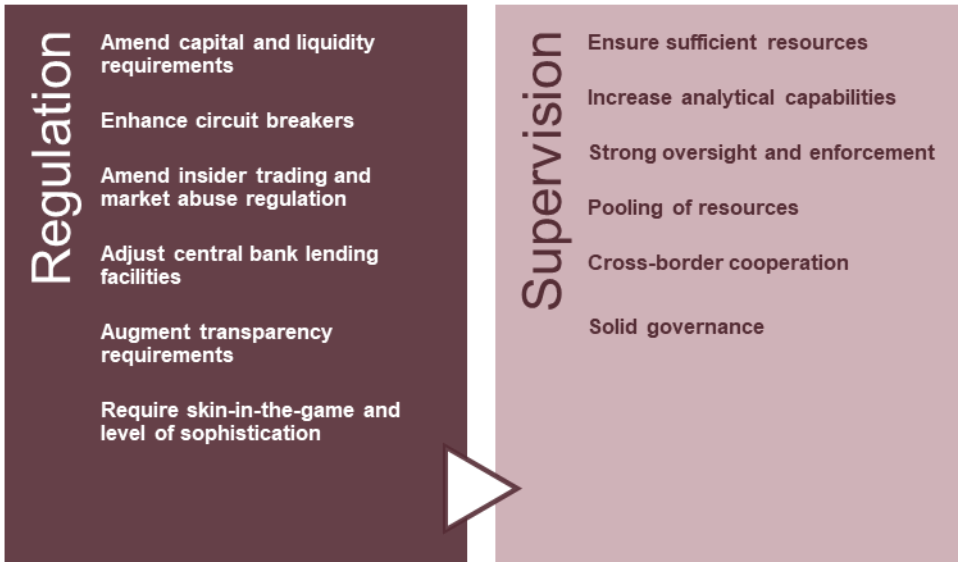
In view of the potential systemic risks and associated market failures, it is essential to implement policies to ensure that AI is used safely. Ideally, we might think of something like Asimov's three laws of robots applied to the financial system or of a global agreement on avoiding malicious uses of AI, similar to the Treaty on the Non-Proliferation of Nuclear Weapons.[146] In reality, we envisage a policy response combining regulation (with competition and consumer protection policies complemented by adjustments to existing financial regulation) and supervision, including a new focus on operational resilience and increased resources to look at the risks posed by AI. Figure 3 summarises our proposals to address systemic risks from AI, based on Sections 4.2 and 4.3.

---

[145] The International Civil Aviation Organization is responsible for developing policies and standards for civil aviation worldwide.

[146] Cecchetti and Schoenholtz (2024) adapt Asimov's three laws of robots as "1. A financial AI must never harm the financial system or allow it to be harmed through inaction; 2. A financial AI must obey human orders, except when it would conflict with the First Law; and 3. A financial AI must protect its own existence, except when it would conflict with the First and Second Laws".

Figure 3
Summary of policy proposals



Source: Authors' elaboration.

In the current geopolitical environment, the stakes are particularly high. Failing to keep up with the use of AI in finance would make financial authorities lose sight of the system, increasing the frequency of episodes of financial instability and, most likely, requiring more frequent intervention by authorities or the public sector; these are usually not cost-free. The global nature of AI calls for a globally agreed policy response, which given the current high degree of geopolitical tensions, adds to the difficulty of the task for financial authorities.[147]

[147] In August 2025, the United Nations General Assembly adopted by consensus a resolution establishing an Independent International Scientific Panel on AI and a Global Dialogue on AI Governance.

References

Acemoglu, D. (2021), "Harm AI", *NBER Working Paper Series*, No 29247, National Bureau of Economic Research.

Acemoglu, D. (2024), "The simple macroeconomics of AI", *NBER Working Paper Series*, No 32487, National Bureau of Economic Research.

Acharya, V., Cooley, T., Richardson, M. and Walter, I. (2011), "Market Failures and Regulatory Failures: Lessons from Past and Present Financial Crises", *ADBI Working Paper Series*, No 264, Asian Development Bank Institute.

Afroogh, S., Akbari, A., Malone, E., Kargar, M. and Alambeigi, H. (2024), "Trust in AI: progress, challenges, and future directions", *Humanities and Social Sciences Communications*, Vol. 11, 1568.

Aldasoro, I., Gambacorta, L., Korinek, A., Shreeti, V. and Stein, M. (2024), "Intelligent financial system: how AI is transforming finance", *BIS Working Paper Series*, No 1194, Bank for International Settlements.

Aquilina, M., Budish, E. and O'Neill, P. (2021), "Quantifying the High-Frequency Trading "arms race"", *Quarterly Journal of Economics*, Vol. 137, Issue 1, pp. 493-564.

Armanious, A. (2024), "Too-systemic-to-fail: empirical comparison of systemic risk measures in the Eurozone financial system", *Journal of Financial Stability*, Vol. 73, 101273.

Armour, J., Awrey, D., Davies, P., Enriques, L., Gordon, J., Mayer, C. and Payne, J. (2016), "Supervision and Enforcement of Financial Regulation", Chapter 26 in *Principles of Financial Regulation*, Oxford Academic.

Arner, D., Barberis, J. and Buckley, R. (2015), "The evolution of fintech: a new post-crisis paradigm?", *University of Hong Kong Faculty of Law Research Series*, No 2015/047, University of Hong Kong.

Baker, M. and Wurgler, J. (2013), "Chapter 5 – Behavioral corporate finance: an updated survey", in *Handbook of the Economics of Finance*, Vol. 2, Part A, pp. 357-424.

Bank for International Settlements (2024), "Artificial intelligence and the economy: implications for central banks", Chapter III in *Annual Economic Report*.

Bank of England (2009), "The role of macroprudential policy – A discussion paper", November.

Barr, M. (2025), "AI: hypothetical scenarios for the future", speech at the Council on Foreign Relations, New York, 18 February.

Barwise, T.P. and Watkins, L. (2018), "The evolution of digital dominance: how and why we got to GAFA", Chapter 1, pp. 21-49, in *Digital dominance: the power of Google, Amazon, Facebook, and Apple*, Oxford University Press, New York, NY.

Beau, D. (2024), "Mastering AI in the financial sector – let us collectively rise the challenge!", speech at the Paris financial centre event devoted to artificial intelligence, Paris, 11 December.

Beck, T., Cecchetti, S., Grothe, M., Kemp, M., Pelizzon, L. and Sánchez Serrano, A. (2022), "Will video kill the radio star? Digitalisation and the future of banking", *Report of the ESRB Advisory Scientific Committee*, No 12, European Systemic Risk Board.

Beck, T., Ioannidou, V., Perotti, E., Sánchez Serrano, A., Suarez, J. and Vives X. (2024), "Addressing banks' vulnerability to deposit runs: revisiting the facts, arguments and policy options", *Report of the ESRB Advisory Scientific Committee*, No 15, European Systemic Risk Board.

Benedetto, C., Crestini, S., de Gregorio, Al., de Leonardis, M., del Monaco, A., Gulino, D., Massaro, P., Monacelli, F. and Rubeo, L. (2025), "Applying artificial intelligence to support regulatory reporting management: the experience at Banca d'Italia", *Occasional Papers*, No 927, Banca d'Italia.

Bengio, Y., Mindermann, S., Privitera, D., Besiroglu, T., Bommasani, R., Casper, S., Choi, Y., Fox, P., Garfinkel, B., Goldfarb, D., Heidari, H., Ho, A., Kapoor, S., Khalatbari, L., Longpre, S., Manning, S., Mavroudis, V., Mazeika, M., Michael, J., Newman, J., Ng, K. Y., Okolo, C. T., Raji, D., Sastry, G., Seger, E., Skeadas, T., South, T., Strubell, E., Tramèr, F., Velasco, L., Wheeler, N., Acemoglu, D., Adekanmbi, O., Dalrymple, D., Dietterich, T. G., Fung, P., Gourinchas, P.-O., Heintz, F., Hinton, G., Jennings, N., Krause, A., Leavy, S., Liang, P., Ludermir, T., Marda, V., Margetts, H., McDermid, J., Munga, J., Narayanan, A., Nelson, A., Neppel, C., Oh, A., Ramchurn, G., Russell, S., Schaake, M., Schölkopf, B., Song, D., Soto, A., Tiedrich, L., Varoquaux, G., Felten, E. W., Yao, A., Zhang, Y.-Q., Ajala, O., Albalawi, F., Alserkal, M., Avrin, G., Busch, C., de Carvalho, A. C. P. de L. F., Fox, B., Gill, A. S., Hatip, A. H., Heikkilä, J., Johnson, C., Jolly, G., Katzir, Z., Khan, S. M., Kitano, H., Krüger, A., Lee, K. M., Ligot, D. V., López Portillo, J. R., Molchanovskyi, O., Monti, A., Mwamanzi, N., Nemer, M., Oliver, N., Pezoa Rivera, R., Ravindran, B., Riza, H., Rugege, C., Seoighe, C., Sheikh, H., Sheehan, J., Wong, D.,. and Zeng, Y. (2025), "International AI Safety Report 2025", January.

Benoit, S., Colliard, J.-E., Hurlin, C. and Pérignon, C. (2017), "Where the risks lie: a survey on systemic risk", *Review of Finance*, Vol. 21, Issue 1, pp. 109-152.

Biden, Joseph R., Jr. (2023), "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence", The White House, 30 October.

Bongaerts, D., De Luca, S. and Van Achter, M. (2024), "Circuit breakers and market runs", *Review of Finance*, Vol. 28, Issue 6, pp. 1953-1989,

Buckmann, M., Haldane, A. and Hüser, A.-C. (2021), "Comparing minds and machines: implications for financial stability", *Staff Working Paper*, No 937, Bank of England.

Buiter, W., Cecchetti, S., Dominguez, K. and Sánchez Serrano, A. (2023), "Stabilising financial markets: lending and market making as a last resort", *Report of the ESRB Advisory Scientific Committee*, No 13, European Systemic Risk Board.

Cabiddu, F., Ludovia, M., Patriotta, G. and Allen, D.G. (2022), "Why do users trust algorithms? A review and conceptualization of initial trust and trust over time", *European Management Journal*, Vol. 40, pp. 685-706.

Cazzaniga, M., Jaumotte, F., Li, L., Melina, G., Panton, A., Pizzinelli, C., Rockall, E. and Tavares, M. (2024), "Gen-AI: Artificial Intelligence and the Future of Work", *Staff Discussion Notes*, No 2024/001, International Monetary Fund.

Cecchetti, S. and Schoenholtz, K. (2024), "On AI and Financial Stability", *Money and Banking Blog*, 15 November.

Center for AI Safety (2024), "Statement on AI risk: AI experts and public figures express their concern about AI risk".

Chugunova, M. and Luhan, W.J. (2024), "Ruled by robots: preference for algorithmic decision makers and perceptions of their choices", *Public Choice*, Vol. 202, pp. 1-24.

Cipollone, P. (2024), "AI: a central bank's view", keynote speech at the National Conference of Statistics on official statistics at the time of AI, Rome, 4 July.

Chen, H., Petukhov, A., Wang, J. and Xing, H. (2024), "The dark side of circuit breakers", *Journal of Finance*, Vol. 79, Issue 2, pp. 1405-1455.

Claessens, S. (2015), "An overview of macroprudential policy tools", *Annual Review of Financial Economics*, Vol. 7, pp. 397-422.

Collins, C., Dennehy, D., Conboy, K. and Mikalef, P. (2021), "Artificial intelligence in information systems research: A systematic literature review and research agenda", *International Journal of Information Management*, Vol. 60, 102383.

Comunale, M. and Manera, A. (2024), "The Economic Impacts and the Regulation of AI: A Review of the Academic Literature and Policy Actions", *IMF Working Paper Series*, No 24/65, International Monetary Fund.

Conlisk, J. (1996), "Why bounded rationality?", *Journal of Economic Literature*, Vol. 34, No 2, pp. 669-700.

Cottier, B., Rahman, R., Fattorini, L., Maslej, N., Besiroglu, T. and Owen, D. (2025), "The rising costs of training frontier AI models", Epoch AI.

Crafts, N. (2021), "Artificial intelligence as a general-purpose technology: an historical perspective", *Oxford Review of Economic Policy*, Vol. 37, Issue 3, pp. 521-536.

Crisanto, J.C., Benson Leuterio, C., Prenio, J. and Yong, J. (2024), "Regulating AI in the financial sector: recent developments and main challenges", *FSI Insights*, No 63, December, Bank for International Settlements.

Dang, T.V., Gorton, G. and Holmström, B. (2020), "The information view of financial crises", *Annual Review of Financial Economics*, Vol. 12, pp. 39-65.

Daníelsson, J. (2022), *The illusion of control*, Yale University Press, New Haven, CT.

Daníelsson, J., Macrae, R. and Uthemann, A. (2022), "Artificial intelligence and systemic risk", *Journal of Banking and Finance*, Vol. 140, 106290.

Daníelsson, J. and Uthemann, A. (2024a), "On the use of artificial intelligence in financial regulations and the impact on financial stability", *working paper*.

Daníelsson, J. and Uthemann, A. (2024b), "Artificial intelligence and financial crises", *working paper*.

Daníelsson, J. and Uthemann, A. (2025), "How central banks can meet the financial stability challenges arising from artificial intelligence", *SUERF Policy Brief*, No 1163, SUERF – The European Monetary and Finance Forum.

De la Peña, N. and Granados, Ó. (2024), "Artificial intelligence solutions to reduce information asymmetry for Colombian cocoa small-scale farmers", *Information Processing in Agriculture*, Vol. 11, Issue 3, pp. 310-324.

De la Vega, J. (1688), "Confusión de confusiones", publication number 13 of the Kress Library of Business and Economics, Harvard Business School of Business Administration.

Dietvorst, B., Simmons, J. and Massey, C. (2015), "Algorithm aversion: people erroneously avoid algorithms after seeing them err", *Journal of Experimental Psychology: General*, Vol. 144, Issue 1, pp. 114-126.

Dietvorst, B., Simmons, J. and Massey, C. (2018), "Overcoming algorithm aversion: people will use imperfect algorithms if they can (even slightly) modify them", *Management Science*, Vol. 64, Issue 3, pp. 1155-1170.

Dongarra, J., Meuer, M., Simon, H. and Strohmaier, E. (2024), "TOP500, Performance development", Top500.org

Epoch AI (2024), "Exponential growth of datapoints used to train notable AI systems", dataset.

European Banking Authority (2023), "Risk Assessment Report", Spring.

European Banking Authority (2024), "Risk Assessment Report", November.

European Central Bank (2009), "The concept of systemic risk", *Financial Stability Review*, December.

European Institute of Innovation and Technology (2021), "Creation of a Taxonomy for the European AI Ecosystem", September.

European Insurance and Occupational Pensions Authority (2024), "Report on the digitalisation of the European insurance sector", April.

European Securities and Markets Authority (2023), "Supervisory briefing On the calibration of circuit breakers", October.

European Securities and Markets Authority (2025a), "Artificial intelligence in EU investment funds: adoption, strategies and portfolio exposures", *ESMA report on Trends, Risks and Vulnerabilities*, February.

European Securities and Markets Authority (2025b), "ESMA publishes implementing rules on Liquidity Management Tools for funds", April.

European Systemic Risk Board (2013), "Recommendation of the European Systemic Risk Board on intermediate objectives and instruments of macro-prudential policy (ESRB/2013/1)", April.

European Systemic Risk Board (2024), "A system-wide approach to macroprudential policy", November.

Eurostat (2025), "Digital economy and society statistics – enterprises", February.

Feduzi, A., Runde, J. and Schwarz, G. (2022), "Unknowns, black swans, and bounded rationality in public organizations", *Public Administration Review*, Vol. 82, Issue 5, pp. 958-963.

Fernández, A. (2019), "Artificial intelligence in financial services", *Economic Bulletin*, 2/2019, Banco de España, March.

Financial Services Authority (2005), "Final notice to Citigroup Global Markets Limited", June.

Financial Stability Board (2024), "The financial stability implications of Artificial Intelligence", November.

Fiott, D. (2022), "Digitalization and hybrid threats: assessing the vulnerabilities for European security", *Hybrid CoE Paper*, No 13, European Centre of Excellence for Countering Hybrid Threats, April.

Flood, M., Kenett, D., Lumsdaine, R. and Simon, J. (2020), "The complexity of bank holding company resolution: a topological approach", *Journal of Banking and Finance*, Vol. 118, 100804.

Foucault, T., Gambacorta, L., Jiang, W. and Vives, X. (2025), "Artificial Intelligence in Finance", *The Future of Banking*, No 7, Centre for Economic Policy Research.

Gai, P., Kemp, M., Sánchez Serrano, A. and Schnabel, I. (2019), "Regulatory complexity and the quest for robust regulation", *Report of the ESRB Advisory Scientific Committee*, No 8, European Systemic Risk Board.

Garicano, L. (2024), "Macroeconomics of AI", remarks at the ECB/BdE conference on the Impact of Artificial Intelligence on the Macroeconomy and Monetary Policy, Madrid, 24 October.

Gaske, M. (2023), "Regulation Priorities for Artificial Intelligence Foundation Models", *unpublished working paper*.

Gensler, G. and Bailey, L. (2020), "Deep learning and financial stability", *working paper*.

Gillespie, N., Lockey, S., Curtis, C., Pool, J. and Akbari, A. (2023), "Trust in Artificial Intelligence: a global study", University of Queensland and KPMG Australia.

Giudici, P. (2018), "Fintech risk management: a research challenge for Artificial Intelligence in finance", *Frontiers in Artificial Intelligence*, Vol. 1, 00001.

Gmyrek, P., Berg, J. and Bescond, D. (2023), "Generative AI and jobs: A global analysis of potential effects on job quantity and quality", *Working Paper Series*, No 96, International Labour Organization.

Godoy, J. (2025), "AI regulation ban meets opposition from state attorneys general over risks to US consumers", Reuters, 16 May.

Guerreiro, J., Rebelo, S. and Teles, P. (2024), "Regulating Artificial Intelligence", *working paper*.

Guillaumie, C., Loiacono, G., Winkler, C. and Kern, S. (2020), "Market impacts of circuit breakers – Evidence from EU trading venues", *ESMA Working Paper Series*, No 1, European Securities and Markets Authority.

Guzik, E., Byrge, C. and Gilde, C. (2023), "The originality of machines: AI takes the Torrance Test", *Journal of Creativity*, Vol. 33, Issue 3, 100065.

Hagendorff, T. (2024), "Deception abilities emerged in large language models", *Proceedings of the National Academy of Sciences of the United States*, Vol. 121, No 24, 2317967121.

Hartmann, P. and Maver, V. (2025), "Implications of Artificial Intelligence for monetary policy – a first conceptual assessment", *SUERF Policy Brief*, No 1080, SUERF – The European Monetary and Finance Forum.

Hobson, Z., Yesberg, J., Bradford, B. and Jackson, J. (2023), "Artificial fairness? Trust in algorithmic police decision-making", *Journal of Experimental Criminology*, Vol. 19, pp. 165-189.

Hoff, K.A. and Bashir, M. (2015), "Trust in automation: Integrating empirical evidence on factors that influence trust", *Human Factors*, Vol. 57, Issue 3, pp. 407-434.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B. and Liu, T. (2023), "A survey on hallucination in Large Language Models: principles, taxonomy, challenges, and open questions", *working paper*.

Huseynova, F. (2024), "Addressing deskilling as a result of human-AI augmentation in the workplace", paper presented at the 7th Conference on Technology Ethics, 6 and 7 November, Tampere, Finland.

International Monetary Fund (2024), "Advances in Artificial Intelligence: implications for capital market activities", *Global Financial Stability Report*, Chapter 3, October.

Irving Fisher Committee on Central Bank Statistics (2025), "Governance and implementation of artificial intelligence in central banks", *IFC Report*, No 18, April.

Ji, J., Qiu, T., Chen, B., Zhang, B., Lou, H., Wang, K., Duan, Y., He, Z., Zhou, J., Zhang, Z., Zeng, F., Dai, J., Pan, X., Ng, K.Y., O'gara, A., Xu, H., Tse, B., Fu, J., McAleer, S., Yang, Y., Wang, Y., Zhu, S-C., Guo, Y. and Gao, W.. (2024), "AI alignment: a comprehensive survey", ArXiV preprint 2310.19852, version 4.

Kahneman, D. (2003), "Maps of bounded rationality: psychology for behavioral economics", *American Economic Review*, Vol. 93, No 5, pp. 1449-1475.

Kirilenko, A., Kyle, A.S., Samadi, M. and Tuzun, T. (2017), "The Flash Crash: High-Frequency Trading in an electronic market", *Journal of Finance*, Vol. 72, Issue 3, pp. 967-998.

Klingbeil, A., Grützner, C. and Schreck, P. (2024), "Trust and reliance on AI − An experimental study on the extent and costs of overreliance on AI", *Computers in Human Behavior*, Vol. 160, 108352.

Koivisto, M. and Grassini, S. (2023), "Best humans still outperform artificial intelligence in a creative divergent thinking task", *Scientific Reports*, Vol. 13, 13601.

Korinek, A. (2024), "Economic policy challenges for the age of AI", *NBER Working Paper Series*, No 32980, National Bureau of Economic Research.

Korinek, A. and Suh, D. (2025), "Scenarios for the transition to AGI", *NBER Working Paper Series*, No 32255, National Bureau of Economic Research.

Kortian, T. (1995), "Modern approaches to asset price formation: a survey of recent theoretical literature", *Research Discussion Paper Series*, No 9501, Reserve Bank of Australia.

KPMG (2024), "Decoding the AI Act", February.

Kumar, S., Lim, W.M., Sivarajah, U. and Kaur. J. (2023), "Artificial Intelligence and Blockchain integration in business: trends from a bibliometric-content analysis", *Information Systems Frontier*, Vol. 25, pp. 871-896.

Langer, E.J. (1975), "The illusion of control", *Journal of Personality and Social Psychology*, Vol. 32, Issue 2, pp. 311-328.

Lee, M.K. (2018), "Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management", *Big Data & Society*, Vol. 5, Issue 1, pp. 1-16.

Leitner, G., Singh, J., van der Kraaij, A. and Zsámboki, B. (2024), "The rise of artificial intelligence: benefits and risks for financial stability", *Financial Stability Review*, ECB, May.

Lin, C., Chicheng, M., Yuchen, S. and Yuchen, X. (2021), "The telegraph and modern banking development, 1881-1936", *Journal of Financial Economics*, Vol. 141, Issue 2, pp. 730-749.

Luo, X., Rechardt, A., Sun, G., Nejad, K., Yáñez, F., Yilmaz, B., Lee, K., Cohen, A., Borghesani, V., Pashkov, A., Marinazzo, D., Nicholas, J., Salatiello, A., Sucholutsky, I., Minervini, P., Razavi, S., Rocca, R., Yusifov, E., Okalova, T., Gu, N., Ferianc, M., Khona, M., Patil, K. Lee, P-S., Mata, R., Myers, N., Bizley, J., Musslick, S., Bilgin, I.P., Niso, G., Ales, J., Gaebler, M. Murty, A. R., Loued-Khenissi, L., Behler, A., Hall, C., Dafflon, J., Donggi Bao, S.,. and Love, B. (2024), "Large language models surpass human experts in predicting neuroscience results", *Nature Human Behaviour*, Vol. 9, pp. 305-315.

Ma, H. and Su, M. (2024), "The bounded intelligence of AI: superficiality and deceivability", *Organisational Dynamics*, in press, 101100.

Maluquer de Motes, J. (2021), "España en la economía mundial. Series largas para la economía española (1850-2015)", dataset, institutional repository of Banco de España.

Marchetti, S. (2022), "Rolling in the deep(fakes)", *Occasional Papers*, No 668, Banca d'Italia.

Martin, A. and Ventura, J. (2018), "The Macroeconomics of Rational Bubbles: A User's Guide", *Annual Review of Economics*, Vol. 10, pp. 505-539.

Marwala, T. and Hurwitz, E. (2015), "Artificial Intelligence and asymmetric information theory", *working paper*.

Mayer, R., Davis, J. and Schoorman, D. (1995), "An integrative model of organizational trust", *Academy of Management Review*, Vol. 20, No 3, pp. 709-734.

Mazzucchi, N. (2022), "AI-based technologies in hybrid conflict: the future of influence operations), *Hybrid CoE Paper*, No 14, European Centre of Excellence for Countering Hybrid Threats, June.

National Cyber Security Centre (2024), "The near-term impact of AI on the cyber threat", January.

Nordhaus, W. and Samuelson, P. (2009), *Economics*, 19th edition, McGraw-Hill, New York.

O'Halloran, S. and Nowaczyk, N. (2019), "An Artificial Intelligence Approach to Regulating Systemic Risk", *Frontiers in Artificial Intelligence*, Vol. 2, 00007.

Ognyanova, K. and Singh, V. (2025), "National AI opinion monitor: AI trust and knowledge in America", Rutgers University, February.

Organisation for Economic Co-operation and Development (2024a), "Assessing potential future artificial intelligence risks, benefits and policy imperatives", *OECD Artificial Intelligence Papers*, No 27, November, OECD Publishing.

Organisation for Economic Co-operation and Development (2024b), "Recommendation of the Council on Artificial Intelligence", May.

Pastorello, S., Calzolari, G., Denicolo, V. and Calvano, E. (2020), "Artificial Intelligence, algorithmic pricing, and collusion", *American Economic Review*, Vol. 110, No 10, pp. 3267-3297.

Petrone, D., Rodosthenous, N. and Latora, V. (2022), "An AI approach for managing financial systemic risk via bank bailouts by taxpayers", *Nature Communications*, Vol. 13, 6815.

Rahman, R., Owen, D. and You, J. (2024), "Tracking compute-intensive AI models", Epoch AI.

Rai, A., Constantinides, P. and Sarker, S. (2019), "Next-generation digital platforms: towards human-AI hybrids – Editor's comments", *MIS Quarterly*, Vol. 43, pp. iii-viii.

Remolina, N. (2022), "Interconnectedness and Financial Stability in the Era of Artificial Intelligence", *research paper*, Singapore Management University School of Law.

Russell, S. and Norvig, P. (2010), *Artificial Intelligence: a modern approach*, 3rd edition, Prentice-Hall.

Scheurer, J., Balesni, M. and Hobbhahn, M. (2024), "Large Language Models can strategically deceive their users when put under pressure", *working paper*.

Schwarcz, S. (2019), "Systematic regulation of systemic risk", *Wisconsin Law Review*, Vol. 2019, No 1, pp.1-54.

Sevilla, J., Besiroglu, T., Cottier, B., You, J., Roldán, E., Villalobos, P. and Erdil, E. (2024), "Can AI scaling continue through 2030?", Epoch AI.

Shleifer, A. and Vishny, R. (2010), "Unstable banking", *Journal of Financial Economics*, Vol. 97, Issue 3, pp. 306-318.

Sifat, I.M. and Mohamad, A. (2020), "A survey on the magnet effect of circuit breakers in financial markets", *International Review of Economics & Finance*, Vol. 69, pp. 138-151.

Silber, W. (1983), "The process of financial innovation", *American Economic Review*, Vol. 73, No 2, Papers and Proceedings of the Ninety-Fifth Annual Meeting of the American Economic Association, pp. 89-95.

Smaga, P. (2014), "The concept of systemic risk", *Systemic Risk Center Special Paper*, No 5.

Solow, R. (1987), "We'd better watch out", *New York Times Book Review*, July 12, p. 37.

Stanford University (2024), "AI Index Report 2024", April.

Sun, Y., Sheng, D., Zhou, Z. and Wu, Y. (2024), "AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content", *Humanities and Social Science Communication*, Vol. 11, 1278.

Tonkiss, F. (2009), "Trust, confidence and economic crisis", *Intereconomics*, Vol. 44, pp. 196-202.

United Nations University (2025), "The rise of the deceptive machines: when AI learns to lie", *blog post*, 1 January.

Videgaray, L., Aghion, P., Caputo, B., Forrest, T., Korinek, A., Langenbucher, K., Miyamoto, H. and Wooldridge, M. (2024), "Artificial Intelligence and economic and financial policymaking", *A High-Level panel of experts' report to the G7*, December.

Vives, X. (2010), "Competition and stability in banking", *CEPR Policy Insight*, No 50.

Vonnegut Jr., K. (1952), Player Piano, Charles Scribner's Sons, New York.

Wever, M., Shah, M. and O'Leary, N. (2022), "Designing Early Warning Systems for detecting systemic risk: a case study and discussion", *Futures*, Vol. 136, 102882.

Xing, Z., Zhu, L. and Lijun, Z. (2020), "A study on the application of the technology of Big Data and Artificial Intelligence to audit", Proceedings of the 2020 International Conference on Computer Engineering and Application (ICCEA), pp. 797-800.

Yamin, M.M., Ullah, M., Ullah, H. and Katt, B. (2024), "Weaponized AI for cyber attacks", *Journal of Information Security and Applications*, Vol. 57, 102722.

Yilla, K. and Liang, N. (2020), "What are macroprudential tools?", Brookings Institution, February.

Zerilli, J. (2021), "Should we be concerned that the decisions of AIs are inscrutable?", *blog post*, 14 June.

Zuboff, S. (2019), "Surveillance capitalism and the challenge of collective action", *New Labor Forum*, Vol. 28, No 1, pp. 10-29.

Annex 1: Summary of the AI Act

In 2024, the EU approved the AI Act, which defines harmonised rules on AI across the EU with the objective of fostering a trustworthy use of AI. The AI Act established the first legal framework on AI worldwide. Its primary aim is to ensure the safety and fundamental rights of persons using AI systems. The first draft was prepared by the European Commission in 2021, and by the first half of 2026 the entirety of the AI Act should already be in force.

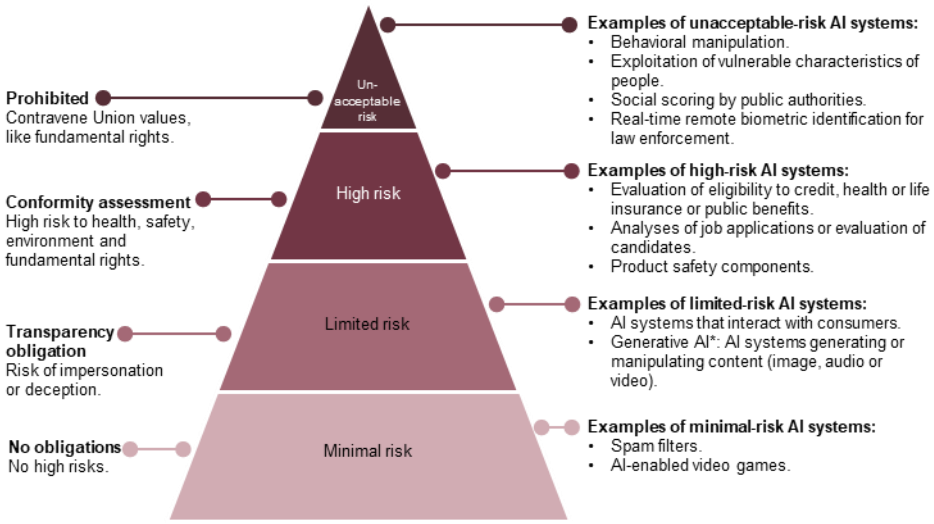
The scope of the AI Act is broad, albeit with some exceptions, and extends to providers of AI systems located outside the EU. According to Article 2, the following institutions or persons should apply the AI Act: (i) any provider placing AI systems on the market or putting them into service within the EU, regardless of that provider's location; (ii) any provider of AI systems located outside the EU, whose system output can or is intended for use in the EU; (iii) any provider of AI systems located in the EU whose AI-produced output is used within the EU; (iv) any importer or distributor of AI systems; (v) product manufacturers placing products with AI systems on the market or putting them into service within the EU under their own name or trademark, and (vi) authorised representatives of providers, which are not in the EU, (vii) affected persons located in the EU. Among the exceptions, the AI Act does not apply to AI systems (i) developed or used exclusively for military, defence, or national security purposes; (ii) used by public authorities or international organisations in non-Union countries when used for law enforcement or judicial cooperation with the EU under a framework of international agreements; (iii) developed and used for the sole purpose of scientific research and discovery; (iv) in the research, testing and development phase before being placed on the market or put into service, and (v) for personal use.

The AI Act introduces a definition of AI, classifies AI systems by risk, lays out extensive requirements and necessary safeguarding mechanisms for AI systems and establishes transparency obligations. The definition of an AI system appears in Article 3 and reads as follows: "'*AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*."

The Act takes a risk-based approach, imposing stricter requirements on AI systems deemed high-risk and prohibiting certain practices considered to pose an unacceptable risk (Figure A1). Examples of prohibited practices include certain uses of real-time remote biometric identification systems in publicly accessible spaces, though exceptions exist under strict conditions for law enforcement purposes related to serious criminal offences and imminent threats to life or safety, subject to specific authorisation procedures and safeguards. An example of a high-risk AI system directly connected to the financial system is the use of AI to evaluate eligibility for credit.

Figure A1

Classification of AI systems according to their risk



Source: KPMG (2024).

The AI Act introduces specific rules for general-purpose AI models. These are defined in Article 3 as "*an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market*". Providers of such models have obligations concerning technical documentation and providing information to providers who integrate the models into AI systems. Where these models generate systemic risk due to high impact capabilities or market reach, providers must comply with additional obligations. These include conducting model evaluations and adversarial testing, assessing and mitigating systemic risks, reporting to the European Commission on serious incidents and ensuring cybersecurity and energy efficiency.[148]

In terms of governance, the AI Office and the national market surveillance authorities are responsible for implementing, supervising and enforcing the AI Act. One task of the AI Office, which sits within the European Commission, is to monitor the effective implementation and compliance by providers of general-purpose AI models with the AI Act. In addition, the AI Act envisages three advisory bodies: (i) the European AI Board, composed of representatives from the Member States; (ii) the Scientific Panel, composed of up to 60 independent experts in the field of AI; and (iii) the Advisory Forum, representing a diverse selection of stakeholders, both commercial and non-commercial. Besides these, the European Commission is to

[148] Recital (110) of the AI Act further elaborates on systemic risk, stating that "*General-purpose AI models could pose systemic risks which include, but are not limited to, any actual or reasonably foreseeable negative effects in relation to major accidents, disruptions of critical sectors and serious consequences to public health and safety; any actual or reasonably foreseeable negative effects on democratic processes, public and economic security; the dissemination of illegal, false, or discriminatory content*".

establish and maintain an EU database for high-risk AI systems that contains information about registered systems. For Member States, the AI Act establishes that they must designate a national market surveillance authority before August 2025 to ensure implementation of the provisions of this Regulation, including by laying down effective, proportionate and dissuasive penalties for their infringement.

The European Commission, supported by the AI Office and potentially by the Scientific Panel, can designate general-purpose AI models as having systemic risk. Based on an overall assessment of factors like training data quality/size, number of users, capabilities, autonomy and scalability, as described in Annex XIII of the AI Act, the European Commission can designate a general-purpose AI model as having systemic risk. A threshold based on the cumulative amount of computation used for training (measured in floating-point operations) is set as a presumption for systemic risk. Article 51.2 of the AI Act initially sets the value of this threshold at 10^25 floating-point operations. The threshold will be adjusted over time.

According to Article 56, the AI Office must encourage and facilitate the drawing up of codes of practice to contribute to the proper application of the AI Act. These codes of practice should take into account international approaches, and the AI Office may invite providers and relevant national competent authorities to participate in drawing them up.[149] In particular, codes of conducts should cover the relevant information to be submitted by providers of general-purpose AI models, including (i) the technical documentation for authorities and downstream providers, (ii) the adequate level of detail for the summary about the content used for training, (iii) the identification of the type and nature of systemic risks and their sources, and (iv) the measures, procedures and modalities for the assessment and management of the systemic risks at EU level. The third draft of the code of practice was published in July 2025.[150]

Finally, the AI Act also includes certain measures to support innovation and particular provisions for financial institutions using high-risk AI systems. Regarding innovation, the AI Act encourages AI regulatory sandboxes at the national level and provides specific support and simplified requirements for SMEs, including start-ups, where appropriate. For financial institutions using high-risk AI systems, the AI Act establishes that compliance with existing financial services rules on internal governance can be deemed to signal fulfilment of certain obligations under the AI Act.

[149] Civil society, industry, academia, downstream providers and independent experts may also support the process of developing codes of conduct.

[150] See the EU website here and an unofficial interactive website here.

Annex 2: Externalities created by AI

AI can create substantial negative externalities and spillovers. An externality is a cost to an uninvolved party that arises from the activities of a third party. For example, pollution by vehicles generates a cost to everyone that neither the corporations manufacturing the vehicles nor the drivers of them bear. Taxes and regulation can discourage externalities by imposing the costs to those who created them. In the case of AI, since neither corporations engaged in its development nor users internalise social objectives, there are large welfare effects stemming from these externalities. As we discuss below, these come in various forms. They include the possible creation of systems that behave unethically, failure to consider how implementation influences labour market outcomes and local communities, how uneven access to systems can generate political tensions both within and between countries, exploitation of data in ways that damage third parties, and the high carbon footprint that may come with continued development and broad adoption of the technology. In this Annex we provide a brief discussion of each of these. Before doing so, we note that virtually none of them is new: people have behaved unethically for the entirety of human history, technology displaced workers and damages communities the start of the industrial revolution, and political tensions arising from difference in rates of development date from the dawn of civilization. While somewhat newer, the improper use of data is a feature of a world dominated by information technology and carbon emissions are a feature of industrialisation. The problem we wish to highlight is that AI may be increasing the speed and scope of these externalities.

AI may act in ways generally agreed to be unethical and counter to society's objectives. At the time of writing, the AI models in use are developed and run by private non-financial for-profit corporations.[151] When making the large investments necessary, they may not consider whether the resulting technology will act in an ethical manner. Examples are easy to construct. Knowing that certain market participants react to news in a particular way, an AI agent may decide to spread false information to move financial markets to its benefit. To conform to prudential requirements, an AI agent may produce supervisory reports that misrepresent a financial institution's balance sheet. Yet another possibility is that the AI agent could develop complex financial instruments that exist solely to create revenue for its owner. Furthermore, there have been concerns raised recently regarding the potential for AI to engage in deceptive behaviour.[152]

AI may affect the labour force and local communities. As a result of widespread use of AI, many occupations currently occupied by humans may become redundant, increasing unemployment, at least temporarily.[153] In the short-term, the most significant impact could be on jobs that consist mainly of cognitive tasks. These

[151] See Aldasoro et al. (2024), Comunale and Manera (2024) and Stanford University (2024).

[152] See Hagendorff (2024) and United Nations University (2025).

[153] See Videgaray et al. (2024). We note, however, that AI can also create its own related jobs, which would then imply a redistribution of employment across sectors, in line with historical evidence on innovation, but that process will take time.

include various aspects of strategic consulting, legal advice, programming and writing.[154] Social costs arise if this raises unemployment and lowers tax revenue over a prolonged period. Involuntary job loss can cause long-lasting and severe harm for the workers affected (such as physical health, depression, suicide or alcohol-related disease), with costs affecting society too through lower consumption.[155] To the extent that job losses are geographically concentrated, this will have a negative impact on communities.

Uneven global distribution of computational, human and financial resources, and digital infrastructure may contribute to creating tensions and dependencies.[156] Concentration on a few large providers of AI or an uneven uptake of AI among countries may create important tensions and exacerbate geopolitical risks. At the extreme, they could even threaten the sovereignty of some countries. These tensions may also stem from different approaches to regulation of AI. When there are differences in the regulatory treatment of AI in terms of scope of regulation, externalities could arise as, for instance, the impact of the use of AI in one country may be felt in another country that has stricter regulatory provisions on AI. Turning to cyber risks, the capacity for malicious actors to launch cyber-attacks (with or without implicit support from governments) may be multiplied in terms of scope, frequency and complexity by AI.[157] AI may be used to take control of key systems in our societies, such as payments, the electric grid or communications, which could have enormous consequences at system level. To address them, national defence policies require what are likely to be expensive upgrades.

Another group of externalities refer to the use of data, which may be subject to copyright, privacy or intellectual property considerations, or whose use may require the consent of third parties. There are issues about the data used to train AI models, which may be subject to intellectual property rights protection and are not adequately respected. A related issue arises with the dissemination of private data by AI systems, intentionally or unintentionally.[158] That would be the case, for example, for the medical data of patients used to train AI to diagnose a certain pathology. We also have data externalities, where data is used in exploitative ways at the expense of consumers and workers.[159] AI could obtain insights from the data of individuals to find ways to extract additional rents from them.

Finally, AI can create negative externalities for climate change mitigation and the environment, as servers and computers have high energy needs and require water in large quantities for cooling. Recent advances in AI capabilities have been made possible by an increase in the computation power used, in turn increasing energy consumption. A significant portion of AI training globally still relies on high-carbon energy sources such as coal or natural gas, leading to emissions of greenhouse gases and contributing to climate change.[160] AI also consumes large

[154] See Bengio et al. (2025).

[155] See Bengio et al. (2025).

[156] See Bengio et al. (2025) and Videgaray et al. (2024).

[157] See Yamin et al. (2021), Fiott (2022) and Mazzucchi (2022).

[158] See Bengio et al. (2025).

[159] See Acemoglu (2021). For further details, see Zuboff (2019).

[160] See Bengio et al. (2025).

amounts of water to cool data centres, leaving an additional environmental footprint. An expansion of AI without a parallel increase in the sustainability of energy and water consumption will have detrimental effects on climate, which would constitute an externality to the society at large.

Acknowledgements
This report benefited from discussions by members of the Advisory Scientific Committee (chaired by Thorsten Beck). The authors gratefully acknowledge comments from colleagues at the ESRB Secretariat and from members of the ESRB Analysis Working Group (chaired by Paul Hiebert and Katja Taipalus), the ESRB Advisory Technical Committee (chaired by Aino Bunge) and the ESRB General Board (chaired by Christine Lagarde).

Stephen Cecchetti
Brandeis University, Waltham, United States; cecchetti@brandeis.edu

Robin L. Lumsdaine
American University, Washington, DC, United States; robin.lumsdaine@american.edu

Tuomas Peltonen
European Systemic Risk Board, Frankfurt am Main, Germany; tuomas.peltonen@esrb.europa.eu

Antonio Sánchez Serrano
European Systemic Risk Board, Frankfurt am Main, Germany; antonio.sanchez@esrb.europa.eu

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.esrb.europa.eu

The cut-off date for the data included in this report was 26 August 2025.

For specific terminology please refer to the ESRB glossary (available in English only).

PDF ISBN 978-92-9472-434-2, ISSN 2467-0685, doi:10.2849/9803684, DT-01-25-018-EN-N