

Written evidence submitted by Andrew Sutton and Sophie Williams

Introduction

This submission addresses aspects of all five questions in the committee's [call for evidence](#), with a particular focus on systemic risks and on “general-purpose” AI. In addition, it: (i) encourages the inquiry to consider how AI adoption across society (not just by financial institutions) can affect the financial system, and (ii) proposes ways financial institutions and policymakers can help AI economic transition go well for the UK.

It was authored by Andrew Sutton and Sophie Williams, with input from Hamish Hobbs and others at the [Centre for the Governance of AI](#) (GovAI), and informed by conversations with around 40 other people, including current and recent finance practitioners, and experts in AI technology, regulation and policy.

[Andrew Sutton](#) has twenty years' international experience with a large UK bank, including focuses on innovation and risk. He currently researches AI risks in the financial system at GovAI, and is a trustee of the [London Initiative for Safe AI](#). [Sophie Williams](#) has extensive experience with the UK government and Financial Conduct Authority (FCA), and currently researches frontier AI model regulation and risk management at GovAI. This submission reflects their personal views.

Executive summary

Our key points are as follows:

- **Artificial Intelligence (AI) is poised to reshape the economy and society, both in the UK and globally.** AI is advancing dramatically in capability, adaptability, and autonomy, with an increasing number of experts forecasting that leading models may outperform humans at almost all cognitive tasks by 2030 or sooner¹. Like other major “general-purpose” technologies, such as industry, electricity, and the internet, AI will change both what society needs from the financial system and how that is delivered. While much remains unknown about the future extent and path of AI development, widespread diffusion of AI at its present capability level would already have significant implications for many sectors of the economy. If current trends in AI capabilities and costs continue,² these changes will be still more profound.
- **Although financial institutions have used narrower forms of machine learning for many years, powerful “general-purpose” AI³ creates new challenges and possibilities.** Most AI systems used in finance today are “narrow” models, trained for a specific task with significant human input. By contrast, “general-purpose” AI systems are trained on massive, general datasets, acquiring capabilities across a wide range of domains (to date, this has been done by AI model developers, not by financial institutions). They have impressive new capabilities, and are increasingly able to operate autonomously, planning and executing complex tasks. If risks

¹ This is one formulation of the concept of “Artificial General Intelligence” (AGI). At the time of writing, 2030 is the median estimate of 1518 forecasters using prediction market [Metaculus](#) for when AGI will be reached. The heads of leading AI model development firms have made increasingly confident predictions about how soon this point may be reached. See for example: “[Anthropic CEO says AI could surpass human intelligence by 2027](#)”, Wall Street Journal, January 2025, and: “[AIs that can match humans at any task will be here in five to 10 years, Google DeepMind CEO says](#)”, CNBC.com, March 2025.

² Sam Altman, CEO of leading model developer OpenAI, has stated that the cost to use a given level of AI is reducing by 10x every twelve months. See “[Three Observations](#)”, Sam Altman personal blog, February 2025.

³ “General-purpose” AI refers to models whose ability to perform a given task arises from their general abilities across a wide range of domains. It is broadly synonymous with systems built on foundation models using transformer technology, in which models acquire capabilities from the formation of connections between vast numbers of “neurons”, in response to exposure to data and feedback. The behaviour of such models is not fully predictable, explainable or consistent. Prominent general-purpose models include ChatGPT, Claude, LLaMA, Gemini, and DeepSeek.

can be mitigated, then general-purpose AI systems could help deliver better financial services at lower costs, mitigate barriers to access, and serve vulnerable customers better. However, general-purpose models are opaque in their operations and can make significant mistakes, and there remains an unsolved challenge of how to reliably align increasingly autonomous systems to human goals. To date, general-purpose AI has mainly been deployed in finance for internal functions – such as operations, knowledge retrieval, human resources, and coding – with humans firmly “in the loop”. Use in market- and customer-facing roles has remained more limited for now, in large part due to concerns over transparency and explainability, privacy, and legal and regulatory risk.

- To fully realise the significant potential benefits from AI, it is crucial to manage a broad spectrum of risks in and via the financial system.** Systemic risks include threats from bad actors (e.g. cyber-attacks and market manipulation), risks of AI usage by financial institutions (e.g. resilience and concentration risks in the supply chain; risks to financial markets stability), and ways in which AI’s adoption in wider society may create problems (e.g. increasing credit defaults in disrupted sectors; bank liquidity risks). Consumer risks arise from bad actors (e.g. fraud), from AI’s limitations (e.g. possible biases), from AI’s strengths (e.g. higher insurance costs for high-risk people as risk assessment improves), and from widening digital and economic divides. Managing such risks is essential for promoting widespread and beneficial AI adoption.
- Traditional regulatory tools are helpful in addressing many risks related to general-purpose AI, but there are areas where new approaches are needed.** Existing regulatory principles and tools remain valuable, though may need clarifying or extending (e.g. to make clear what AI uses or features require disclosure and to whom, or what scenarios stress tests should include), especially for systemically important institutions. However, in some scenarios – such as those involving more autonomous AI – existing approaches may become ineffective (e.g. where intention is key to whether an action is permitted). Policymakers will also face new questions, for example, to what extent use of AI agents⁴ by financial consumers requires a novel regulatory approach. Many new tools and techniques to help mitigate risks from AI models are being developed by the UK AI Security Institute (AISI)⁵ and elsewhere, which should be explored further.
- If risks are managed effectively, the UK financial sector is well-placed to competitively leverage AI.** The UK benefits from its position as a global hub for both AI and finance. If, as some studies have suggested,⁶ AI increases rewards for market leaders, this should benefit the UK in various aspects of finance. Meanwhile, finance is rich in cognitively-demanding, high-value activities where AI can, to the extent permitted, complement or replace humans (albeit with unclear employment effects). Challenges include international regulatory divergence (which could complicate operations for global UK firms), and the dominance of large US firms in the AI supply chain potentially capturing much of the value. However, AI complexities also offer commercial opportunities for the UK in AI assurance, legal services, insurance, and consulting.
- The financial sector has a crucial role in facilitating the UK’s AI economic transition.** Widespread AI adoption promises significant economic growth, but may also result in substantial disruption.⁷ The financial sector can help provide investment for growth sectors, and cushion

⁴ “AI agents” are general-purpose AI systems which can autonomously act, plan, and delegate to achieve goals with little to no human oversight (this is the definition used in the [International AI Safety Report 2025](#)).

⁵ AISI works on a range of technical solutions to address risks of deliberate misuse of AI, outlined [here](#).

⁶ See for example: “[Investing in AI boosts growth and increases market dominance, study finds](#)”, Berkeley University news room, October 2020.

impacts on legacy sectors, perhaps by providing transition financing to impacted sectors through public-private mechanisms. As AI is likely to boost returns on capital more than returns on labour, giving more people the opportunity to financially benefit from capital valuation growth (e.g. through pension investment or retail investment, in the private sector) could help to curb rises in inequality, as could sovereign wealth investments aimed at ensuring that public finances directly benefit from rising capital valuations. The finance sector can also provide valuable risk management expertise and AI risk insurance, and promote best practices via institutional investor influence and listing requirements. Many of these represent substantial commercial opportunities.

We therefore make the following recommendations:

- **Broaden the scope of the inquiry to include implications for the financial system (including systemic and consumer risks) arising from AI adoption in wider society** (see [Question 1](#)). Consider how AI adoption in wider society may affect financial stability, consumer protection, and financial institutions' business models, for example, by causing disruption to exposed economic sectors (with employment and credit implications), changing consumer behaviours, or creating other vulnerabilities. The committee could employ targeted horizon-scanning workshops or expert consultations, as well as encourage policymakers and regulators to appropriately prioritise this in their work.
- **Focus primarily on the risks and implications arising from advanced general-purpose AI, both in the inquiry and in the work of policymakers and regulators** (see [Question 1](#)). Advanced general-purpose AI creates novel risks and opportunities significantly beyond prior forms, due to its powerful capabilities, generality, growing autonomy, and challenges around transparency, explainability and alignment to human goals. Clearly distinguish these types of models from narrower forms, and prioritise work on the risks and implications of their use. Challenges relating to narrow AI systems in finance remain real, but are better understood and evolving less rapidly.
- **Consider a wide range of economic scenarios, AI adoption patterns, and downstream effects, recognising these may differ significantly from today** (see [Question 1](#)). Scenarios should consider a broad range of possibilities, including widespread use of highly capable agentic AI, substantial labour displacement (if AI shifts from mainly augmenting to mainly replacing human labour), as well as more moderate outcomes. Adaptable policies that are beneficial across a wide range of scenarios should be designed and iterated as AI's path and impacts become clearer. Regular scenario-planning involving regulators, AI experts, and financial institutions, supported by methods like prediction markets, can assist.
- **Equip regulators and policymakers with the knowledge, capabilities, responsibility, and authority needed to manage AI risks, and require financial firms to show they have done likewise** (see [Question 3](#)). Government should establish AI capacity-building programmes for regulators and policymakers, create industry and technical advisory groups, and clarify regulatory mandates covering AI risks to ensure there are no material gaps. Financial institutions (especially systemically important ones) should demonstrate appropriate governance, systems, and training

⁷ There is no consensus on the labour effects of widespread AI adoption, which depends on many factors, including the balance of labour-complementing and labour-displacing effects. For example, the Tony Blair Institute (TBI) expects the UK unemployment effect to peak "in the low hundreds of thousands and for the effects to unwind over time", empowering workers and improving job-matching ("[The Impact of AI on the Labour Market](#)", TBI, November 2024). Meanwhile, the Institute for Public Policy Research (IPPR) assessed the potential UK jobs impact of 2024-era AI if fully integrated into organisational processes, finding a central case of 4.4m job losses, and depressing wages in sectors that displaced workers move to ("[Transformed by AI: how generative AI could affect work in the UK - and how to manage it](#)", IPPR, March 2024).

to manage the AI systems that they and their suppliers use, and readiness for the impacts of AI adoption among customers, counterparties, and the wider economy. AI challenges should not be treated exclusively, or even primarily, as a data science issue.

- **Consider a combination of traditional and non-traditional tools to mitigate the risks from AI** (see [Question 5](#)). Regulators should continue to review how traditional tools and principles of financial regulation apply to AI, providing guidance and making revisions as needed. They should also explore new interventions, such as AI model audits, requirements for AI explainability and traceability, and real-time monitoring of AI harms, engaging with work at AISI and elsewhere. In many cases it may make most sense to take a cross-regulatory approach (e.g. feeding into development of such tests and leveraging their output, rather than undertaking them directly). Steps should also be taken to proactively identify novel challenges and potential mitigants, as well as new policy trade-offs.
- **Foster international regulatory cooperation and alignment** (see [Question 2](#)). Divergent international AI regulations complicate compliance and risk management for UK financial firms operating globally, and may therefore reduce their competitiveness. Governments and regulators should pursue bilateral and multilateral dialogues, as well as international standards initiatives, to position the UK as a leader in AI governance. They should also collaborate globally on capacity-building, horizon-scanning, and research.
- **Promote and encourage the financial sector's role in helping the AI economic transition go well for the UK** (see [Question 5](#)). Financial firms and policymakers will be central to maximising the economic upside from AI adoption, whilst also mitigating any adverse effects, including labour market dislocations and heightened wealth disparities. Government should support the development of tools (e.g. transition finance, AI risk insurance products, and risk pooling) to aid a successful and equitable economic transition. There may be applicable lessons to draw on from climate transition and climate finance, for example.
- **Study evolving public opinion and expectations around AI** (see [Question 4](#)). Public acceptance and trust in AI is essential to avoid pushback that could disrupt financial services and markets. In addition, AI will create many new policy trade-offs for which public opinion and expectations will be relevant factors. Government should commission ongoing public research and stakeholder engagement initiatives to monitor attitudes towards AI, especially in sensitive areas like financial decision-making, privacy, employment impacts, and systemic stability.

Individual Question Responses

Use of AI in the financial system (Q1)

How is AI currently used in different sectors of financial services and how is this likely to change over the next ten years?

- **Most current uses of AI in finance rely on specialised (or “narrow”) forms of AI, and are used mainly in quantitative tasks, operational processing, and anomaly detection.**
 - A recent survey⁸ by the Bank of England and FCA provides useful insight into how AI is used in finance (which roughly accords with the authors’ own experience).
 - “Narrow” forms of AI, including expert systems and specialised machine learning models, have been used since the 1980s, evolving significantly from around 2010 with the emergence of big data and cloud computing. These remain the dominant type of AI used in finance.⁹
 - Such systems are developed for a particular purpose with significant human input. They are used for a range of tasks, typically involving data modelling and prediction, anomaly detection (e.g. in fraud and anti-money laundering), or operational processing, and in algorithmic trading.¹⁰
 - Narrow AI systems have a relatively well-understood set of challenges and limitations in the context of financial services, primarily around data quality and governance, privacy, and challenges of working with legacy systems and business processes.
- **Financial institutions are exploring the use of “general-purpose” AI models, and deploying them cautiously in limited contexts.**
 - General-purpose AI functions differently to narrower systems, in that a model’s ability to perform a given task arises from its general abilities across many domains. These models are trained on vast, general datasets, and can then be fine-tuned to improve performance on specific tasks.
 - General-purpose AI brings powerful new capabilities, including, at the time of writing:
 - Generating **real-time text, speech and video** increasingly indistinguishable from human-created content, including impersonating/portraying specific real people.¹¹
 - **Making sense of vast, structured and unstructured** data streams.
 - **Summarising, drafting and translating text**, and writing high-quality **computer code**.
 - Delivering convincing **human-level chatbots**.¹²
 - Performing complex **cognitive tasks at expert levels**.¹³
 - **Setting and pursuing goals** at increasing levels of sophistication and autonomy.
 - **Conducting financial transactions** (e.g. several cryptocurrency firms offer wallets for AI agents, and there is little in principle to prevent AI agents making transactions on behalf of humans using traditional financial institutions, via APIs, or even through voice channels).

⁸ See “[Artificial Intelligence in UK Financial Services](#)”, Bank of England and FCA, November 2024.

⁹ The BoE/FCA survey found that 75% of institutions use AI in some form, but only 17% of use cases leverage foundation models (loosely equivalent to “general-purpose” AI).

¹⁰ In more detail: in banking, such models are used for anti-money laundering and fraud screening, credit and risk modelling, and back-office automation; in insurance, narrow models are used for risk assessment, underwriting and claims management; in asset management and markets, they are used for risk modelling, market signal analysis, trade fulfilment, and “algorithmic trading”, in which a computer trades automatically based on predefined rules or statistical signals. See “[Risks of an Intelligent Financial System](#)”, Jacobs & Sutton, OMFIF, 2025.

¹¹ See, for example, OpenAI’s [Sora](#) video generation model.

¹² In a recent test, OpenAI’s GPT-4.5 chatbot model was judged to be human more often than actual humans were, when rated by humans who interacted in text-based conversations with both groups. See “[Large Language Models Pass the Turing Test](#)” (pre-print), Jones & Bergan, 2025.

¹³ Including passing expert-level exams in mathematics, science, law, medicine and finance. OpenAI’s o3-mini model scored 96.7% on the American Invitational Mathematics Examination 2024 and 87.7% on GPQA Diamond, a PhD-level science benchmark (see [here](#)). GPT-4 passed the US’s Uniform Bar Exam with a score of 297 (see [here](#)), well above the passing mark for all jurisdictions, scored in the 92nd to 99th percentile in the German medical licencing exam (see [here](#)), and passed Levels 1 and 2 of the Chartered Financial Analyst exam (see [here](#)).

- **Coordinating with other AI systems**, including, in principle, in financial markets (which could include behaviours that would be considered collusive if done by humans).
 - Most of these tasks could previously only be done by skilled humans. With improvements in general-purpose AI these tasks can be done much faster than humans, scalably (and typically at significantly lower ongoing cost), without the need to build a specialised system from scratch.
 - From a labour perspective such models can be thought of as both a powerful tool for human workers, and as, essentially, a highly-skilled, scaleable, synthetic workforce.
 - Financial firms are exploring many use cases for general-purpose AI, but have been understandably cautious in deploying it in customer- or market-facing roles, largely because of concerns about transparency, explainability, privacy and legal and regulatory risk. The survey by the Bank of England and FCA (referenced above) finds the greatest current usage of general-purpose AI is in legal, human resource and research functions. Humans remain firmly “in the loop”, with almost no use of fully autonomous decision-making by such models at this stage.
- **How AI is used by financial institutions over the next decade depends largely on whether challenges related to the transparency, interpretability, and alignment of general-purpose AI systems are resolved, and on how regulation and public sentiment evolve.**
 - Use of narrow AI systems in finance will continue to grow, supported by clear commercial incentives and established use cases (e.g. fraud detection, process automation, risk modelling).
 - General-purpose AI will be more widely deployed, initially for ancillary tasks (e.g., coding support, drafting documents under human supervision, enhancing fraud and AML detection for additional review), and to automate semi-routine tasks with well-defined quality controls (often processes that are currently offshored), delivering significant cost savings.
 - The degree to which general-purpose AI models are entrusted with core financial activities (e.g. lending and trading decisions, financial advice, and financial decision-making), and in direct interactions with customers and markets, will depend on how the following are resolved:
 - **Technical challenges:** Including the transparency and explainability of decisions, concerns about bias, and the need to reduce hallucinations. As general-purpose AI becomes more powerful and autonomous, a more fundamental challenge emerges around how we can ensure the technology is well-aligned with human goals.
 - **Regulatory factors:** Including clarifications around liability, data privacy, and intellectual property obligations, as well as how various international regulations apply in cross-jurisdictional contexts (e.g. where part of a process takes place overseas, or where a model is used for a compliant purpose, but has been developed using techniques or data that would be non-compliant in the jurisdiction where it is deployed).
 - **Public sentiment:** Involves addressing societal acceptance, trust, or resistance toward AI systems. Financial brands enjoy significantly higher trust than tech firms, so trust may be weakened if financial services increasingly rely on such firms. Meanwhile, one US study found that two-thirds of consumers believe AI in financial services increases their risk of fraud or security breaches, with 20% regarding this risk as “extreme”.¹⁴
 - Developing robust AI assurance frameworks should help to accelerate beneficial adoption. Other promising approaches include leveraging general-purpose AI to enhance narrower, interpretable models, or developing specialised AI systems trained on validated financial data.
- **However, the implications and risks of AI for the financial system may depend as much on how it is adopted in wider society as it does on how it is used by financial institutions.**

¹⁴ See “[As New Technology is Integrated in Financial Services Industry, Most Bank Customers in United States Express Distrust for AI](#)”, J.D Power, February 2024.

- Many credible risk stories rely on AI being adopted outside the financial system and then creating risks within it. These are explored in [Question 3](#) and [Question 4](#).
- It may be useful to think about the effects of AI on the financial system via the following vectors:
 - Use of AI by **financial institutions**, including existing firms and new entrants, and the effects of consequent changes to the structure of the industry.
 - Use of AI by **bad actors** (e.g. fraud, cyber-attacks, market manipulation) attacking financial sector participants, and knock-on effects on trust.
 - Use of AI by **consumers** (increasingly unmediated by any financial firm), e.g. if consumers were to use AI agents to manage personal finances, and how financial firms may respond to this.
 - Effects of AI within the **wider economy** (e.g. disrupting or augmenting different sectors; employment effects, and structural changes in key parameters such as interest rates¹⁵).
 - Development of an economic ecosystem comprising largely-autonomous **AI agents**, doing activities previously reserved to humans and engaging with financial services.¹⁶
- Anecdotally, attention within financial institutions and regulators is focused mainly on the earlier items in the list above, whereas the later items may prove to be at least as consequential.
- **Modelling these wider effects is difficult, but important if the risks and implications for the financial system are to be properly identified and managed.**
 - Economic and broader effects of AI may include:¹⁷
 - A significant **increase in the returns on capital**, relative to labour.
 - A large **increase in interest rates** (returns on capital, and risk, both increase).
 - **Unclear effects on employment**, with potential for both job displacement and creation.
 - In many scenarios, a significant **increase in productivity and economic growth**.
 - Related effects on **public finances** (including possible shrinking income tax base as returns accrue to capital, higher debt interest, possible increased unemployment costs).
 - **Increasing the power of AI firms** and potential market concentration.
 - Shifting **geopolitical dynamics** shaped by an AI “arms race”.
 - There is substantial expert disagreement on the timing, degree, and in some cases the direction, of these effects. Some economists are working to consider economic questions that may be relevant in a “post-AGI” future,¹⁸ but this is a under-developed area of study.
 - This implies a very wide range of plausible futures that policymakers must consider, as well as a need for further work in these areas.

Recommendations (these are repeated from the earlier summary list)

- **Broaden the scope of the inquiry to include implications for the financial system (including systemic and consumer risks) arising from AI adoption in wider society.** Consider how AI adoption in wider society may affect financial stability, consumer protection, and financial institutions' business models, for example, by causing disruption to exposed economic sectors (with

¹⁵ On AI and interest rates see, for example: “[AI and government bond yields - onwards and upwards?](#)”, Robin Marshall, London Stock Exchange Group, June 2024, and: “[Transformative AI, existential risk, and real interest rates](#)”, Chow, Halperin and Mazlish, December 2024.

¹⁶ See this article describing the possible development of an “agent-to-agent” economy: “[The three stages of the Agent-to-Agent \(A2A\) economy](#)”, John S. Kim, Sendbird, February 2025. Such an economy could use both traditional and decentralised financial services: for example, Coinbase, a US based cryptocurrency exchange, has launched “AgentKit” which allows developers to build AI agents that can perform a variety of operations. See “[Introducing AgentKit](#)”, November 2024.

¹⁷ Two resources that set out a range of possible future economic scenarios are here (note that the points above draw on a range of sources, not only the following, which differ in various respects): “[Threshold 2023: Modelling AI Economic Futures](#)”, Deric Cheng et al, February 2025, and “[GATE: Modelling the trajectory of AI and Automation](#)”, Epoch AI, March 2025.

¹⁸ For example, see “[Economic Policy Challenges for the Age of AI](#)”, forthcoming in Rethinking Economic Policy: Steering Structural Change, MIT Press, Korinek, 2025.

employment and credit implications), changing consumer behaviours, or creating other vulnerabilities. The committee could employ targeted horizon-scanning workshops or expert consultations, as well as encourage policymakers and regulators to appropriately prioritise this in their work.

- **Focus primarily on the risks and implications arising from advanced general-purpose AI, both in the inquiry and in the work of policymakers and regulators.** Advanced general-purpose AI creates novel risks and opportunities significantly beyond prior forms, due to its powerful capabilities, generality, growing autonomy, and challenges around transparency, explainability and alignment to human goals. Clearly distinguish these types of models from narrower forms, and prioritise work on the risks and implications of their use. Challenges relating to narrow AI systems in finance remain real, but are better understood and evolving less rapidly.
- **Consider a wide range of economic scenarios, AI adoption patterns, and downstream effects, recognising these may differ significantly from today.** Scenarios should consider a broad range of possibilities, including widespread use of highly capable agentic AI, substantial labour displacement (if AI shifts from mainly augmenting to mainly replacing human labour), as well as more moderate outcomes. Adaptable policies that are beneficial across a wide range of scenarios should be designed and iterated as AI's path and impacts become clearer. Regular scenario-planning involving regulators, AI experts, and financial institutions, supported by methods like prediction markets, can assist.

Productivity and UK financial sector competitiveness (Q2)

To what extent can AI improve productivity in financial services?

Note: AI offers various opportunities for financial services. We touch on some of these, though this paper mainly addresses how to identify and mitigate risks, reflecting the authors' current research focus.

- **AI has significant scope to increase productivity in financial services, with unclear implications for employment and wages in the sector.**
 - In principle, there is great potential to automate work in the financial sector, and strong economic incentives to do so. Finance has a high concentration of highly-paid, cognitively demanding roles of the type that can be complemented or displaced with AI, including many tasks that could not be done by earlier iterations of AI, but which now can be. For example:
 - Goldman Sachs's CEO, David Solomon, said in January that AI can now draft 95% of an IPO prospectus in minutes, versus a six-person team working for two weeks, commenting that "the last 5 per cent matters because the rest is now a commodity".¹⁹
 - A 2023 report by the UK's Department for Education found that jobs in the finance and insurance sector are more exposed to (i.e. automatable by) general-purpose AI than any other sector.²⁰
 - This implies that very high potential productivity gains are available to the finance sector from deploying AI, albeit with potentially significant employment implications.
- **The UK's financial sector is well-placed to take advantage of AI, leveraging its strong position in both the financial services and AI sectors.**

¹⁹ See "[Goldman Sachs chief David Solomon questions start-ups' need to list](#)", Financial Times, January 2025.

²⁰ See "[The Impact of AI on UK jobs and training](#)", Department for Education, November 2023.

- London is a global financial hub, the second-largest AI centre after the Bay Area, and has a significant fintech ecosystem. The UK also benefits from proactive, principles-first regulation, and the government has been a global leader in engaging both with the promise and risks of AI.
- This puts the UK financial sector (and ancillary businesses) in a strong position to reap competitive advantages, through the development and sale of finance AI technology and services, as well as by leveraging the technology to gain a competitive advantage globally.
- Furthermore if AI has market-concentrating effects (e.g. rewarding firms with the greatest data assets or R&D budgets), that should benefit various UK firms or markets with leading international positions.
- There is also scope for the UK to develop leadership in AI assurance, which could further enhance the UK's position, both by accelerating implementation of beneficial AI applications in the UK, and by providing it with more services to export. Additionally, the UK could develop and export AI insurance, legal, and consulting services.
- **International regulatory divergence, especially with the EU and US, presents complexities.**
 - Three main models for the regulation of AI's use by business are emerging: a more laissez faire US model (which recently rolled back federal regulation, leaving rule-making largely with states and courts²¹), a rights-based and rules-based EU model (expressed mainly through the EU AI Act, which establishes a risk-based classification for uses of AI), and a statist Chinese approach (emphasising national security and stability).
 - The UK's principles-based, sector-led regulation provides important flexibility to adapt to a fast-changing context. Yet, divergent international regulations may force financial institutions to build separate systems or limit AI adoption by complying with the most stringent rules (for example, EU reporting requirements and bans on certain "high-risk" activities), and/or to face competitive pressure from more lightly-regulated foreign markets and institutions.

Recommendations

- **Foster international regulatory cooperation and alignment.** Divergent international AI regulations complicate compliance and risk management for UK financial firms operating globally, and may therefore reduce their competitiveness. Governments and regulators should pursue bilateral and multilateral dialogues, as well as international standards initiatives, to position the UK as a leader in AI governance. They should also collaborate globally on capacity-building, horizon-scanning, and research.

Financial stability risks of AI (Q3)

What are the risks to financial stability arising from AI and how can they be mitigated?

- **Advanced general-purpose AI presents a range of risks to financial stability, both from its use by financial institutions and customers, and as a consequence of its use in wider society.**
 - These risks arise via various mechanisms, and range from well-evidenced present threats to more speculative ones that depend on how, and how fast, AI is developed and deployed.
 - Some more notable risks highlighted by researchers and regulators²² are as follows:

²¹ This describes the US's regulatory stance domestically. AI is also seen through the lens of security and geopolitical competition, especially with China, with significant export controls and other restrictions aimed at maintaining the US's lead in AI technology.

²² See, for example: "[The financial stability implications of artificial intelligence](#)", Financial Stability Board, 2024, and: "[Risks of an Intelligent Financial System](#)", Jacobs & Sutton, OMFIF, January 2025.

- **Increased risk of fraud and cyber-attacks (systemic):** AI is increasing the risk of fraud and cyber-attacks by introducing new techniques that help attackers (e.g. the use of deepfakes), and also undermining some existing controls (e.g. voice passwords).²³ These pose a direct risk of losses, as well as a systemic threat if such attacks take down critical infrastructure or institutions. However, it should be noted that advanced AI can also be a powerful tool to detect and defeat fraud and cyber crime.
- **Potential increase in financial market fragility:** There are concerns that advanced AI agents trading in financial markets may make crashes more frequent or severe.²⁴ Risks arise from the speed at which AI bots ingest information from a wide range of sources, the potentially-correlated ways they may act on such information, and the unpredictable nature of interactions between AI models. These may amplify market volatility, or reduce market liquidity during periods of stress.²⁵ AI agents have been shown in experiments to discover market abuse strategies, and can communicate without detection. AI trading agents may be vulnerable to covert manipulation by other market participants,²⁶ or may themselves propagate misinformation to move markets. Consideration is needed around the use of AI agents, especially by more lightly-regulated parties (e.g. hedge funds).²⁷
- **Concentration risks and operational resilience:** There is high concentration through most of the AI supply chain, in terms of firms, models, techniques and geographies (including high exposure to the US and Taiwan). Meanwhile, financial firms' reliance on third-party providers is rising. This presents concerns about: i) correlated model behaviours or vulnerabilities (which may lead many models to make the same mistake or respond in the same way to certain contexts; ii) risks of correlated service failures. In 2023, Gary Gensler, then chair of the US Securities and Exchange Commission, described it as "nearly unavoidable" that AI will cause a financial crisis within a decade, primarily because of concentration risks.²⁸
- **Challenges to lending posed by economic transition:** There is uncertainty about how lending will respond to economic transformation driven by AI. If AI diffuses rapidly through the economy it will create many winners and losers (individuals, companies, sectors, countries) in ways that are hard to predict. To lenders, this represents a material increase in risk that is not fully captured in existing credit models. This heightens risks of a credit crisis, triggered either by bad debts (if lenders do not respond to a changed outlook) or by a sharp reduction in credit availability (if they do). This appears under-explored in the literature. At a sovereign level, while AI-driven economic growth may strengthen some countries' public finances, countries highly reliant on income tax, especially those unlikely to capture much upside from AI capital growth, may instead face worsening public finances and thus increased default risk, risking contagion.
- **Challenges to bank liquidity:** There are at least two mechanisms here. If customers use AI agents to manage their finances more effectively (e.g. by quickly moving deposits to the institution paying the highest rates, or moving money out of institutions in response to rumours of weakness, for example on social media) that may reduce deposit stickiness,

²³ 72% of respondents to a recent survey reported an increase in organisational cyber risks, with 47% citing adversarial advances powered by generative AI as their primary concern. Firms reported a sharp increase in phishing and social engineering attacks in 2024. See "[Global Cybersecurity Outlook](#)", World Economic Forum, January 2025.

²⁴ "[Monsters in the Deep?](#)" (speech by Jonathan Hall), Bank of England, May 2024.

²⁵ Note that there are also plausible scenarios under which the use of advanced AI trading bots may reduce volatility. This may be the case if they price assets more closely to their true value, leading markets to move more smoothly, or if they respond in more sophisticated ways to shocks in the market (such as by coordinating to avoid contagion).

²⁶ "[Technological and economic threats to the US financial system](#)", Sytsma et al., RAND, 2024.

²⁷ "[AI in the real world: hedge funds' use of Artificial Intelligence in trading](#)" (Majority Staff Report), US Senate Committee on Homeland Security and Governmental Affairs, June 2024.

²⁸ "[Gary Gensler urges regulators to tame AI risks to financial stability](#)", Financial Times, October 2023.

increasing the chance of bank runs.²⁹ In addition, a recent study³⁰ suggested that AI-generated misinformation could be used by bad actors to trigger a run on a UK bank (e.g. for financial gain or political motives).

- **Loss of control risks:** More speculatively, the UK's AI Security Institute (AISI) highlights a potential future risk of humans losing control of powerful AI systems, either generally or in limited but high-risk contexts. Some scenarios in the literature involve uncontrolled AI systems using money to control other resources including labour.³¹ Financial institutions could help mitigate this risk by monitoring and controlling AI systems' access to financial services. This relies on financial institutions being aware of when their services are being used by AI agents, for which there are currently few controls. Decentralised assets (cryptocurrencies)³² and "smart contracts" (i.e. self-executing contracts on a blockchain) may enable AI systems to transact and contract with other parties independently, undermining these controls.
- In addition, AI systems may behave differently during crises than normal periods, while previous experience that informs regulators may no longer be predictive, complicating risk management.

● **The following chart provides a visualisation of our assessment of risks in this area:**

Risks shown (not in any order)

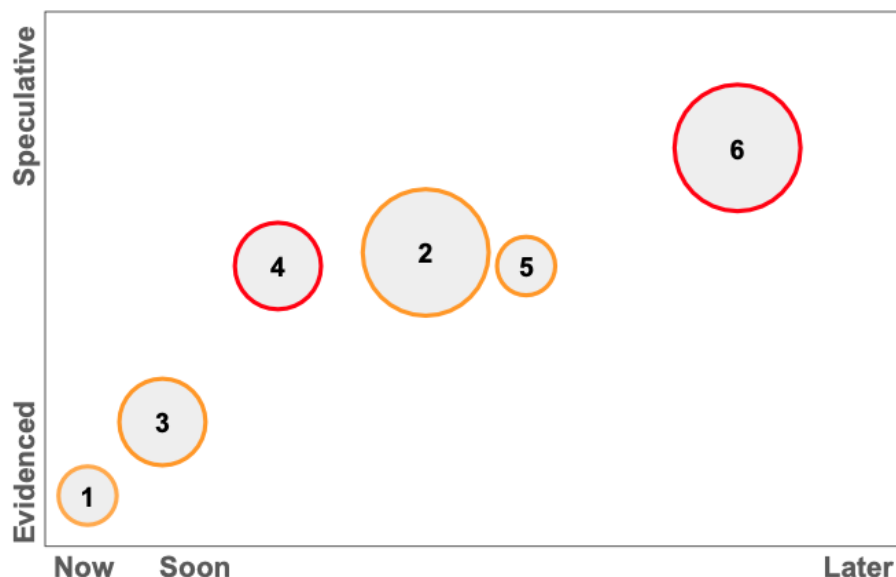
1. Cyber Attacks and Fraud
2. Financial Markets Fragility
3. Concentration & Resilience
4. Challenges to Lending
5. Challenges to Bank Liquidity
6. Loss of Control Scenarios

Potential harm (systemic)



Neglectedness

- Somewhat neglected
- Very neglected



Recommendations

- **Equip regulators and policymakers with the knowledge, capabilities, responsibility, and authority needed to manage AI risks, and require financial firms to show they have done likewise.** Government should establish AI capacity-building programmes for regulators and

²⁹ A somewhat analogous issue contributed to the collapse of Silicon Valley Bank, where a tightly networked, digitally connected depositor base - coordinating in real time via platforms like WhatsApp - enabled a rapid withdrawal cascade before any institutional response could intervene. The depositor base in that case was unusually well-networked, attuned to rumours, and fast acting (properties they may share with AI agents).

³⁰ "Can AI trigger a Bank Run?", Say No to Disinfo and Fenimore Harper Communications.

³¹ The idea of AI agents hiring humans is not quite as far-fetched as it might sound: there are already many digital platforms where workers can be hired for specific tasks, and many "gig economy" business models where humans are directed, essentially, by computer. A 2023 paper from OpenAI on its GPT-4 model relates that the model was asked to complete a "CAPTCHA" test (i.e. a type of simple online test designed to ensure that a website user is a human and not a computer). The model found itself unable to complete the CAPTCHA, but successfully hired a human on an online labour platform to solve the test for it, with the AI model pretending to be a visually impaired human to explain to the worker why it needed help to complete the test. This illustrates both models' ability to hire workers, and to engage in deception to achieve its aims. See "GPT-4 Technical Report", OpenAI, 2023.

³² The US government's failed attempts to shut down Tornado DAO illustrate the legal and technical barriers to a governments' ability to shut off access to a blockchain-based system. See "Victory for Tornado Cash as Court Rules Sanctions Were Unlawful", Baker Hostetler, March 2024.

policymakers, create industry and technical advisory groups, and clarify regulatory mandates covering AI risks to ensure there are no material gaps. Financial institutions (especially systemically important ones) should demonstrate appropriate governance, systems, and training to manage the AI systems that they and their suppliers use, and readiness for the impacts of AI adoption among customers, counterparties, and the wider economy. AI challenges should not be treated exclusively, or even primarily, as a data science issue.

Benefits and risks to consumers (Q4)

What are the benefits and risks to consumers arising from AI, particularly for vulnerable consumers?

- **Advanced general-purpose AI presents a wide range of risks to consumers, with vulnerable and disadvantaged people at particular risk.**
 - Systemic failures all have consumer harm implications. In addition, AI may create harm to consumers directly, through misuse, accident, or structural features:
 - **Increasing fraud and cyber-attacks (via direct harms):** AI is driving an increase in fraud and cyber attacks (as noted). Meanwhile, public awareness of new types of fraud remains low (e.g. frauds involving convincing personalised phishing messages, or involving deep-faked voice or video).³³ As financial institutions respond to this threat, authentication processes may become more onerous, creating challenges, in particular, for some classes of vulnerable consumer.
 - **Bias or discrimination in lending, pricing or service decisions:** AI systems used in decision-making often reflect the patterns seen in their training data, which may embed biases.³⁴ In addition, if there are legal constraints that prevent specific factors being used in decision-making, then AI may recreate these prohibited factors in ways that are hard to detect. Models may also suffer from a recency bias because most data used in training was created in recent years.³⁵
 - **Challenges to insurance provision:** As insurers use AI to improve risk assessment, high-risk people may increasingly be priced out of insurance provision, creating a policy challenge. Conversely, if AI becomes widely used by consumers to assess their own risk (e.g. via personalised medical diagnostic tests, one area in which AI has led to dramatic technical advances, albeit with regulatory approval lagging), then adverse selection effects³⁶ may make some conditions uninsurable to the general market.
 - **Exacerbating digital and economic divides:** Customers who use AI tools effectively to manage their finances could secure better terms or avoid penalties, disadvantaging those without access. Additionally, unsophisticated customers may face unfavourable terms if AI identifies them as such. More broadly, AI is expected to increase returns to capital relative to labour, disproportionately harming consumers without investments.
- **There are also many ways in which AI will improve financial services for consumers.**
 - Although this is not the main focus of our submission, the authors recognise many ways in

³³ See, for example, "[UK engineering firm Arup falls victim to £20m deepfake scam](#)", The Guardian, May 2024.

³⁴ Note that it is possible for AI to also reduce bias, relative to human assessors.

³⁵ The amount of digital data created and stored each year grows exponentially, driven by many factors (cheaper storage, more people creating more types of content, profusion of digital devices and sensors, etc.). Thus, AI models have hugely more data from 2024 than from 2014 or 2004 or earlier years. This may cause them to overlook durable historical patterns in favour of more recent anomalies, biasing their output.

³⁶ Adverse selection: if insurers know that consumers can identify their own risk level more accurately than the insurer can (because consumers now have access to cheap and accurate tests), they will expect mainly higher-risk customers to seek insurance. To price in this higher risk, they must increase rates, driving away lower-risk customers. This cycle continues until insurance is either prohibitively expensive or unavailable.

which advanced AI can directly benefit consumers, for example:

- **Better and more personalised products and services**, and more “human-like” interaction via remote channels.
- **Breaking down barriers to access**, for example by explaining complex topics in an accessible way, and providing tailored financial advice.³⁷
- **Lowering costs of provision**, enabling fee reductions and expansion of financial services to more marginal customers and cases.
- **Reducing bias**, relative to human assessors, in some scenarios.
- **Improved risk pricing and forecasting**, opening up more products for many consumers.
- **Improved fraud detection** (notwithstanding a concurrent increase in fraud threat).

Recommendation(s)

- **Study evolving public opinion and expectations around AI.** Public acceptance and trust in AI is essential to avoid pushback that could disrupt financial services and markets. In addition, AI will create many new policy trade-offs for which public opinion and expectations will be relevant factors. Government should commission ongoing public research and stakeholder engagement initiatives to monitor attitudes towards AI, especially in sensitive areas like financial decision-making, privacy, employment impacts, and systemic stability.

How should policymakers and regulators proceed? (Q5)

How can Government and financial regulators strike the right balance between seizing the opportunities of AI but at the same time protecting consumers and mitigating against any threats to financial stability?

- **Regulators should consider a range of traditional and novel approaches to mitigate AI risks.**
 - As a general objective, policymakers and regulators must have appropriate knowledge, capabilities, responsibilities, and authority to manage AI risks effectively.
 - Policymakers should collaborate across departments and with the industry and researchers to:
 - i. **Conduct horizon scanning and scenario modelling.** On a rolling basis, explore a wide range of plausible scenarios for AI technical development, adoption and effects, using a range of tools including expert panels and prediction markets, using this to inform work prioritisation. Identify the most important drivers of financial system AI risk, and develop leading indicators for these items (e.g. speed of AI adoption, use of general purpose models for trading in markets, model supplier concentration, etc.). Develop rapid-response capabilities to identify, assess, and mitigate emergent or unexpected AI-related risks.
 - ii. **Study specific risk drivers and prioritise**, including of individual AI applications, emergent properties of AIs interacting, and customer interactions with AI. Conduct experimental, observational and theoretical work to assess key risks, for example on behaviour of advanced AI trading in markets, or collusion. Collaborate with industry using regulatory sandboxes and targeted compliance advice to explore safe AI adoption.
 - Which AI techniques, or which applications of AI in finance are highest risk, and should there be formal designation or additional processes for these?³⁸

³⁷ The first author of this paper has successfully obtained complex multi-jurisdictional tax advice by asking a general purpose AI chatbot (later validated with a professional advisor), illustrating the potential of such systems for such work. The challenges are that a non-trivial chance of hallucination or error remains, as well as challenges around auditability and liability if this were to be deployed in a regulated context.

³⁸ i.e. should an approach similar to the EU AI Act be taken, whereby certain uses are designated as Prohibited (not permitted in any circumstances) or High Risk (carrying additional obligations). The “High Risk” functions identified in finance are: i) credit scoring of individuals, ii) risk assessment and pricing for life and health insurance.

- Should concentration risks and resilience be a higher priority for regulators, given very high concentration and third-party reliance in the AI supply chain?
- iii. **Consider the implications for traditional regulatory tools, principles, and controls, and whether these need to be revised, clarified or supplemented.** For example:
 - What AI capabilities and skills should boards, senior managers, and staff demonstrate? What governance processes should institutions implement?
 - How can executives demonstrate compliance with Senior Manager or other obligations, if deploying systems whose operations are not fully understood?
 - What culture or values regarding AI should institutions foster, and how?
 - Which AI uses and system features should require disclosure (to regulators, customers or markets)? Should AI risk assessments be standardised?
 - Should risk-absorbing features of the financial system be adapted? (e.g. stress test scenarios in bank capital testing, “circuit-breaker” mechanisms in markets)?
 - Should certain AI applications be banned or require regulatory pre-approval? Could “safe harbour” frameworks speed adoption of AI in low-risk scenarios?
 - What KYC or monitoring controls are needed to distinguish AI and human activity?
 - How should record-keeping and auditing standards apply to AI systems? What AI model audits should be mandated?
 - Are new regulatory mechanisms needed to address risks of a more powerful new generation of AI trading bots?
- iv. **Consider tools beyond the typical financial services regulation toolkit,** working with cross-government colleagues in DSIT and elsewhere. For example:
 - Should (and how should) obligations arising from financial regulation extend to AI developers, cloud providers, or others in the AI value chain? Would rules on “downstream AI developers” (e.g. financial institutions, or others, who fine-tune or modify a commercially available foundation model) help address risks?³⁹
 - How can rules, standards or tests being developed by AISI and others for frontier AI address risks in finance? (e.g. model audits, risk assessments, removal or detection of capabilities relevant to fraud or market manipulation)
 - What standards of transparency and explainability should apply to AI in finance?
 - Should mandatory human or automatic “kill switches” be required for AI systems deployed in high risk finance applications? What “fail-safe” standards are required?
 - What economic structures and financial mechanisms are required to support large-scale economic transition? What is the proper role of government in this?
 - Harness AI tools for regulators (e.g. surveillance, anomaly detection).
- v. **Consider where AI creates new policy trade-offs or requires re-consideration of provision of essential financial services such as insurance.** Study evolving public opinion and societal values as they relate to AI, and integrate these into policy decisions, aiding transparency and public legitimacy.
 - One example might be where consumers use AI agents to manage their deposits more efficiently, increasing liquidity risk. Is it right to discourage this use of AI agents, or deliberately introduce friction into this process?
 - Similarly, insurers having better risk assessment tools may lead more people to become uninsurable (at an affordable rate). Should insurers be limited in how they use such tools? If not, are there social provision gaps that need to be filled?
 - Should requirements for transparency and explainability remain, even if this means forgoing AI use for tasks where it clearly outperforms humans in other respects?

³⁹ See for example: [“On regulating downstream AI developers”](#), Williams, Schuett & Anderljung, March 2025.

- vi. **Provide evidence to feed into wider policy debates on liability, copyright, competition law and other issues.**
 - How should anti-collusion rules apply to AI-driven decisions, including ones made without explicit communication? How should rules that turn on intention apply?
 - Should AI systems be expected to perform the standard of a typical human, or some other standard?
 - Should there be liability requirements and/or insurance mandates for third-party providers of critical AI services to finance?
- vii. **Work with international regulators, seeking alignment where possible**
 - Collaborate on capacity-building, upskilling, horizon-scanning and risk research.
 - Monitor international regulatory approaches to AI and their effectiveness, seeking to adopt best practice in balancing risk-management and growth objectives.
 - Seek to harmonise new requirements where possible.
- **The financial sector also has an important role in helping AI go well for the UK, and the UK has an important global role too.**
 - Finance has an important role to play in supporting beneficial AI adoption more broadly while mitigating risks, and the government has a role in supporting these. For example:
 - i. **Financing an orderly economic transition to boost growth and reduce labour impacts:** The shift to an economy increasingly driven by AI will require significant investment, both in AI infrastructure and broader economic restructuring. This will involve the decline of many sectors and, it is to be hoped, the flourishing of many others. The financial sector can play a pivotal role by creating large pools of risk-tolerant types of capital (equity markets, debt-equity structures, public-private partnerships and policy banks) capable of supporting industries in transition and unlocking new opportunities. Private-sector capital will be very important but probably insufficient, both because of capacity limitations and externalities (the government carries many costs of labour dislocation).
 - ii. **Sharing financial upsides to mitigate wealth inequality (via investment products, pensions, and sovereign wealth funds):** AI is likely to accelerate economic growth, while also increasing returns to capital relative to labour. This has implications for wealth distribution and the tax base. Measures to encourage broader equity ownership (directly, or through pension investment policy) can help mitigate this, as could sovereign wealth structures whereby growth in capital valuations directly benefits the public finances.
 - iii. **Risk mitigation expertise and markets:** Financial firms and markets have extensive experience in identifying, pricing, and managing novel risks. Insurance mechanisms can help reduce and align incentive to risk-taking, encourage data-sharing, and develop best practice to manage AI risks effectively. Risk-pooling mechanisms may be valuable for addressing catastrophic AI risks. This is a commercial opportunity for the UK, given its strong position in insurance.
 - iv. **Promoting transparency and corporate governance through listing codes and institutional shareholder codes of practice and board seats:** The financial sector can help develop and popularise codes of good practice for AI in various ways, e.g. through listing requirements for public markets, influence of board seats held by institutional shareholders, or development of financial disclosures, in some ways analogous to how sustainability standards and practice have developed.
 - Considering the financial and professional services sectors together, the above represents a big commercial opportunity.
 - Finally, finance provides relevant techniques that could be used to improve wider AI security

(e.g. digital know-your-customer regimes, suspicion reporting, and export control enforcement), while international financial regulation models show how countries can cooperate to maintain stability despite competing in other regards. Additionally, climate risk may offer useful lessons for incorporating complex, forward-looking AI risks into assessments and reporting.

Recommendations

- **Consider a combination of traditional and non-traditional tools to mitigate the risks from AI.** Regulators should continue to review how traditional tools and principles of financial regulation apply to AI, providing guidance and making revisions as needed. They should also explore new interventions, such as AI model audits, requirements for AI explainability and traceability, and real-time monitoring of AI harms, engaging with work at AISI and elsewhere. In many cases it may make most sense to take a cross-regulatory approach (e.g. feeding into development of such tests and leveraging their output, rather than undertaking them directly). Steps should also be taken to proactively identify novel challenges and potential mitigants, as well as new policy trade-offs.
- **Promote and encourage the financial sector's role in helping the AI economic transition go well for the UK.** Financial firms and policymakers will be central to maximising the economic upside from AI adoption, whilst also mitigating any adverse effects, including labour market dislocations and heightened wealth disparities. Government should support the development of tools (e.g. transition finance, AI risk insurance products, and risk pooling) to aid a successful and equitable economic transition. There may be applicable lessons to draw on from climate transition and climate finance, for example.

April 2025