Amazon Web Services (AWS) response to the Treasury Select Committee (TSC) inquiry on the use of artificial intelligence (AI) in the UK financial services sector

01 October 2025

Introduction

AWS (Amazon Web Services) provides cloud computing services and infrastructure that allows businesses, organisations and individuals to access computing power, storage, databases, AI and other IT services over the internet, rather than having to build and maintain their own physical data centres.

We recognise the transformative potential of AI and generative AI, which AWS makes accessible to organisations of all sizes so they can build innovative, new technology solutions without worrying about infrastructure resources. With AWS, customers have access to leading foundation models (a 'model' is a computer program, or a combination of algorithms, trained to learn patterns, make predictions, or perform tasks without explicit human instructions at every step. Foundation models understand language, generate text and images, and converse in natural language) and can customise them with their own data, and use security, access control and features as they choose.

Responsible use of technologies is key to continued innovation. AWS is committed to developing fair and accurate AI services and providing customers with the tools and guidance to build and scale generative AI safely, securely and responsibly. We are committed to continued collaboration with policymakers, the industry, researchers, and the AI community to advance the responsible use of AI.

Question 1: What impact will AI have on financial services? *and* Question 2: Does your firm have a specific strategy for AI in financial services?

AI technology is already applied in the financial services industry across multiple operational areas. In customer service, AI-powered systems provide 24/7 automated assistance through virtual assistants and chatbots and deliver personalised recommendations based on customer data analysis. In operational processes, AI streamlines workflows by automating document processing, customer onboarding and loan assessments. Risk management and fraud detection use AI's ability to analyse vast datasets in real-time, identifying suspicious patterns and potential threats faster than traditional methods. In trading operations, AI systems support portfolio optimisation and market analysis, providing traders with data-driven insights for decision-making.

The next 5-10 years will likely see further applications emerge. Specialised Language Models (models tailored for specific business needs) will be developed for financial sector tasks, while Retrieval-Augmented Generation (RAG) systems (RAGs optimise the output of language models so they reference knowledge bases outside their training data sources before generating a

response, to ensure further accuracy) will support information management and compliance processes. AI agents could integrate with existing financial systems, enhancing portfolio management and quantitative trading capabilities.

AWS aims to provide customers with options to use AI if they choose to. This means building a stack of AI services to support AI development and deployment. As such, AWS has an integrated a three-layer system to provide AI capabilities.

At the bottom layer is infrastructure for AI experts and practitioners who are comfortable building, tuning, training, deploying, and managing models themselves. AWS provides the computing power and specialised graphics processing unit (GPU, an electronic circuit that can perform mathematical calculations at high speed) support needed for AI workloads, designed to ensure reliable, high-performance processing for all AI operations.

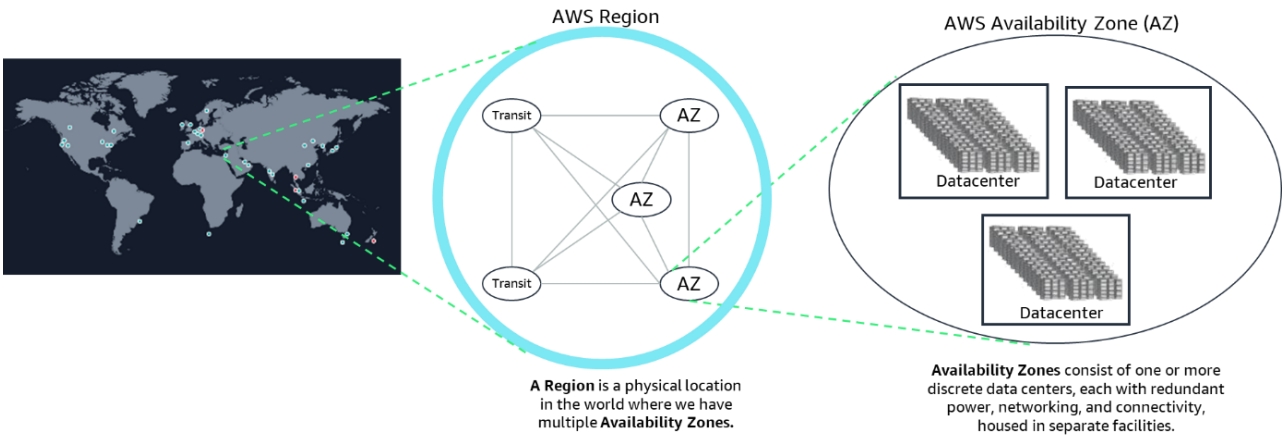
Building upon this base, the middle layer serves as the intelligence centre by providing access to Large Language Models (very large deep learning models pre-trained on vast amounts of data) and foundation models (models that understand language, generate text and images, and converse in natural language). These work in tandem to enable model development, training, and deployment, allowing customers to process and interpret vast amounts of data to generate meaningful insights.

At the top, the application layer enables everyday business use through ready-to-use AI services and tools, meaning customers can implement AI solutions for specific business needs without deep technical expertise. These together mean that organisations of all sizes can use and harness AI tools to suit their business needs.

Question 3: What preparations does your firm have in place for failures or outages in its cloud system and any AI systems it hosts?

Financial services customers are choosing to use AWS to support and improve their security and resilience. AWS has a comprehensive approach to resilience that spans multiple layers of protection, ensuring businesses can reliably maintain operations.

This starts with AWS infrastructure built around multiple Availability Zones within each of our Regions, supported by redundant power systems and networking capabilities. This geographic distribution of data centres creates a natural buffer against localised failures or disasters. *Please see representation of this in the graphic below.*



AWS's service-level resilience features provide real-time monitoring and alerting. These services work together to automatically adjust to changing demands and potential issues before they impact operations, with proven durability levels and comprehensive backup solutions that account for data protection requirements. AWS also supports automated snapshots and cross-region replication to ensure database availability.

This multilayered approach, combined with automated failover capabilities and distributed architecture, creates an environment where organisations can build and maintain highly resilient applications with confidence.

AWS also tests business continuity regularly to ensure effectiveness of the associated procedures and organisational readiness, achieving SOC2 certification. AWS documents the results, including lessons learned and any corrective actions.

Question 4: What interaction, if any, has your firm had with the Bank of England and/or the FCA on AI?

AWS is engaged with the Bank of England (BoE) and the Financial Conduct Authority (FCA) as they develop policies and approaches relevant to use of cloud in financial services, regarding operational resilience and use of third parties. These have included discussions around AI, as the BoE and FCA have sought to understand the provision of services in this space and customer usage.

In addition, AWS has participated in the formal industry outreach by the FCA, attended the FCA's AI Sprint and submitted a written response to the AI Input Zone. AWS staff participate in a personal capacity in the AI Consortium, which provides a platform for public-private engagement to gather input from stakeholders on the capabilities, development, deployment and use of AI in UK financial services

Question 5: The Bank of England and the FCA have both told the Committee that certain AI providers could be brought into the new Critical Third Parties Regime if HM Treasury deems them to be critical to the UK economy. If HM Treasury designated your firm as a critical third party under this regime, how would that impact your firm?

AWS expects to be designated as a critical third party (CTP) under the UK regime. We are supportive of the regime's intention to support the operational resilience of the financial sector. AWS is already subject to oversight and incident notification obligations under the existing UK Network and Information Systems Regulations (NIS) and participates in a broad range of UK assurance regimes such as Cyber Essentials Plus, NHS Data Security and Privacy Toolkit (DSPT), and the Police Assured Secure Facility (PASF). As such the authorities need to be thoughtful about further expansion to ensure it meets its objectives to enhance operational resilience as opposed to creating duplicative regimes.

Question 6: The Bank of England and the FCA take a "technology-agnostic" approach to AI regulation and supervision, meaning that they do not usually require or prohibit specific technologies. Do you agree with this approach?

AWS supports regulators' technology-agnostic approach to AI regulation and supervision. This aligns with sound regulatory principles that focus on managing risks and ensuring resilience without prescribing specific technical solutions or architectures. This perspective means there is flexibility for regulated firms to implement technologies that best suit their circumstances, while still maintaining robust risk management and operational resilience.

By focusing on best outcomes rather than technical specifications, organisations can embrace emerging technologies and innovative solutions while demonstrating proper risk controls. This flexibility proves particularly valuable in a landscape where technology evolves rapidly, allowing businesses to stay current without being constrained by outdated technical requirements. This framework also recognises the diverse nature of organisations using AI technology. From small startups to large enterprises, each entity can develop solutions aligned with their scale and specific needs, all while adhering to fundamental principles of risk management and control. This balanced approach ensures both innovation and responsibility in AI implementation.

Question 7: The Artificial Intelligence (Regulation) Bill, a Private Member's Bill in the House of Lords, proposes establishing a dedicated AI regulator and requiring firms using or developing AI to designate an AI officer, among other changes. How would these proposals impact your firm, if implemented? *and* **Question 8: What recommendations does your firm have for AI regulation and supervision in the UK?**

As with any policy proposal, particularly in such a technical area such as this, it is hard for us to judge the impact of a Bill upon our business until it has been finalised and the necessary implementing legislation and guidelines are in place. That said, in general our view is that, as

with any new technology, the focus should be on specific use cases where AI could add new risks beyond those already present.

AI is already subject to a broad range of laws, consumer and data protection requirements, and cybersecurity obligations. When considering new measures, the focus should be on identifying whether AI introduces genuinely novel risks beyond those in non-AI systems, avoiding unnecessary regulation. International coordination is also crucial to avoid regulatory fragmentation, as AI is used across borders.

ANTHROP\C

1st October 2025

Dame Meg Hillier MP
Houses of Parliament
London, SW1AA 0AA

Dear Dame Meg,

Subject: Response to Treasury Committee

I am writing on behalf of Anthropic in response to the Treasury Committee's recent letter requesting further information with regards to Anthropic's work with the financial services industry.

For context, Anthropic is an AI safety and research company working to build reliable, interpretable and steerable AI systems. We are a Delaware public benefit corporation, a legal form that enables us to make choices that are consistent with our mission of creating a materially positive impact on society, even where this does not maximise shareholder value. Alongside our public benefit status, our Long-Term Benefit Trust (LTBT) aligns our corporate governance with our mission and public benefit purpose. We conduct frontier research, develop and apply a variety of safety techniques, and deploy the resulting systems via a set of partnerships and products. In 2023, we launched Claude, a generative AI model with advanced reasoning and analysis, code generating, conversational and linguistic skills. Claude is a text-based generative AI system used by many people around the world to assist with a wide range of tasks. These include improving writing, providing coding assistance, offering strategic advice, conducting business analysis, analysing documents and helping with various daily activities. It's available via both an API for developers and a public website (claude.ai) for consumer and business users.

For your convenience I have included the original questions you sent through at the top of each answer.

1. What impact will AI have on financial services?

As a company, we focus on the development of generative AI models that are powerfully capable—and relied on by customers—in a variety of substantive domains. As a result, at present it is difficult for us to provide a definitive assessment of the likely impact of AI on financial services in particular. The determining factors are much broader than the

development of the technology, which is still in relative terms very early. Factors such as legislation, regulation and industry usage are as important when it comes to assessing any impact and these will evolve alongside the technology. At the same time, we do study the shape of AI adoption across the world and publish the results as part of Anthropic's Economic Index, an initiative aimed at understanding AI's effects on labor markets and the economy over time. What we can say from our Economic Index is we are seeing a broadly even split between AI being used for the augmentation of tasks (where AI directly performs tasks such as formatting a document) and the automation of tasks (where AI collaborates with a user to perform a task). It would be reasonable to assume financial services as an industry in general terms track these trends, but to be confident in such an assertion further detailed research into the usage of AI within the financial services industry would be needed. As a provider of foundation models, we are simply less well positioned to predict the impact on financial services than the firms operating in that sector that are thinking about how to incorporate AI into their businesses.

2. Does your firm have a specific strategy for AI in financial services?

Anthropic's business strategy is to develop general-purpose AI capabilities that enable users across all industries to work more efficiently. As a predominantly B2B company, we usually focus on broad model improvements rather than sector-specific development, though these advances naturally benefit financial services applications as well as those of other sectors.

With respect to financial services, in particular, in July 2025, we launched Claude for Financial Services, which leverages our Model Context Protocol (MCP) to connect financial institutions with their existing data providers for AI-powered analysis. The key difference between this and standard Claude for Work is that it includes pre-built data integrations—allowing direct access (with permission) to industry-specific data providers such as S&P Global, PitchBook and Databricks—that allow for more targeted research, expanded usage limits, and technical support. This solution is designed to support analytical workflows around due diligence, market research, portfolio analysis, and financial modeling. Claude for Financial Services maintains checkpoints to keep humans in the loop, and enables financial experts to derive insights from their data more efficiently, but does not execute financial transactions or decisions without oversight. Our approach centers on augmenting human expertise with analytical capabilities rather than bypassing human judgment in critical financial processes.

3. Does your firm have systems in place to check or audit the outputs of its AI models?

Anthropic undertakes rigorous pre-deployment testing for its models to assess adherence to its Usage Policy, understand model capabilities and performance consistent with intended behaviors, and test for harmful behaviors. This is important for compliance with our own standards of corporate governance (such as our Responsible Scaling Policy, a framework for safely training and deploying frontier AI models) as well as the ongoing commercial viability of our products and services. We report our findings in the system cards released with each model and share further details in other publications, where appropriate. After deployment, Anthropic's Safeguards team monitors model usage to detect and enforce on harmful behavior. Enforcement takes the form of real-time intervention, such as output blocking or response steering, as well as targeted human review of potential violations where appropriate. Using privacy-preserving techniques, we also monitor Claude traffic in aggregate, going beyond single prompts and individual accounts, to understand the prevalence of particular harms and identify more sophisticated attack patterns to further inform our safety approach. More information about our approach can be found here: https://www.anthropic.com/news/building-safeguards-for-claude.

4. The Committee has received evidence warning that AI decision-making lacks transparency, potentially impacting consumers in financial services. How does your firm ensure transparency in the decision-making of its AI models?

Anthropic distinguishes between two types of AI decision-making transparency: model interpretability, or understanding why an AI system produces a specific output, and data flow transparency, or providing users with visibility and control over how their data moves between systems and what actions AI takes on their behalf. Both types of transparency are integral to how we operate.

Model interpretability has been central to Anthropic's mission since our founding as a research organization. Indeed, our mission statement makes explicit our goal to build "reliable, interpretable, and steerable AI systems." While large language models are notoriously difficult to analyze, we continuously advance the science of understanding the underlying reasoning behind AI responses. We've publicly shared our interpretability research findings and have published detailed system cards (or model cards) alongside every model release to provide insight into model capabilities and limitations.

Regarding data flow transparency, Claude connects to external systems primarily through MCP (mentioned above in response to Question 2), an open protocol developed by Anthropic that enables users to connect to external data sources and authorize real-world

actions, such as reading emails or managing calendar invitations. We have designed the MCP protocol and our implementation of the protocol (connectors) to ensure user consent, user and organization-level control, awareness, and transparency at every step: Claude can only interact with services a user has explicitly connected to, users can enable or disable individual connectors before each prompt, and all data sent to and retrieved from connectors is displayed to the user. Notably, Claude for Financial Services does not include MCP servers that enable asset transfers or financial transactions, and our MCP Connectors Directory explicitly does not support financial transaction capabilities. When developing capabilities that interact with the real world, we adhere to our trustworthy agent principles: humans remain in control, agentic behavior is transparent, agent activities align with human values and expectations, and users' privacy and security are protected.

5. The Committee has also received evidence from several groups arguing that AI models could perpetuate bias and discrimination against consumers in credit and insurance. How does your firm ensure that its AI models avoid bias and discrimination?

On a broad level, we evaluate all of our models for bias in outputs, identifying whether our models provide reliable, consistent responses across different contexts and users. We report our findings in the system cards released with each model. In the case of financial services, we recognize potential risks to consumers, and as such, we have designated insurance and finance use cases as “high risk” use cases in our Usage Policy. This requires users integrating with our models to have a human-in-the-loop when making decisions or providing advice directly affecting individuals or consumers, and to disclose the use of AI when outputs are surfaced to individuals or consumers. This reflects Anthropic’s belief that relevant professional expertise should be integrated into decision-making for these use cases.

6. What preparations does your firm have in place for failures or outages in its AI systems?

Anthropic maintains robust infrastructure with redundancy across multiple cloud providers and accelerator platforms designed to prevent single points of failure. We operate real-time monitoring and observability systems that track system health, performance metrics, and usage patterns, enabling us to detect potential issues before they significantly impact users. Our incident response teams are distributed globally across multiple time zones, to help enable rapid detection, triage, investigation, and response to issues. Where appropriate, these teams follow defined escalation procedures and runbooks for common failure scenarios, and we continue to work on developing and refining our response approaches, especially as our products and services evolve. In the event of partial outages or capacity constraints (within or outside of our control), we try to maintain reduced

functionality where possible to minimize impact on our users. Moreover, we maintain protocols and communication channels for informing users during service disruptions, including via our public status page and notifications to enterprise customers, where appropriate. This multi-layered approach to reliability is designed so that we can maintain service continuity and respond effectively to a range of potential failure modes.

7. The Financial Stability Board and the Bank of England, among others, have said that AI-driven trading could cause market volatility and contagion. What actions, if any, has your firm taken to address this risk in its AI models?

When working with financial services firms, we are very clear that our products are tools intended to support human decision making. Claude provides analysis and insights that can be used to inform human decision-making, but investment decisions remain with the financial institution customers. We're not making automated trading decisions or bypassing human judgment in critical financial processes. Further, under the policies that we communicate to customers, any such use would be "high risk" and require a human in the loop. With regards to how financial services firms are thinking about future integration of AI or use cases, they would be best placed to provide further details.

8. What interaction, if any, has your firm had with the Bank of England and / or the FCA on AI?

At present there is a single live conversation with the Financial Conduct Authority (FCA) regarding their own potential usage of AI. This has not covered broader policy work or the future plans of the FCA with regards to regulation of the financial services industry. We have not had any substantive interaction with the Bank of England.

9. The Bank of England and the FCA have both told the Committee that certain AI providers could be brought into the new Critical Third Parties Regime if HM Treasury deems them to be critical to the UK economy. If HM Treasury designated your firm as a critical third party under this regime, how would that impact your firm?

We understand that the Critical Third Parties Regime is designed to apply to those service providers where “a failure in or disruption to the provision of the services that the third party provides to firms could threaten the stability of, or confidence in, the UK financial system.” Although we cannot rule out the possibility that there may be AI providers that fall within the above definition, we would not expect generally that a failure of an AI provider would automatically be defined as such. It would of course be a question of fact as to whether the failure of a specific AI firm could lead to such an impact, and we would not

anticipate Anthropic falling within this category at this time. As such, we have not considered the impact this would have on our company.

10. The Bank of England and the FCA take a "technology-agnostic" approach to AI regulation and supervision, meaning that they do not usually require or prohibit specific technologies. Do you agree with this approach?

We believe that AI provides a unique and powerful opportunity to unlock significant economic benefit, but have consistently argued that this must be balanced with a proportionate approach to safety, transparency and risk management.

11. The Artificial Intelligence (Regulation) Bill, a Private Member's Bill in the House of Lords, proposes establishing a dedicated AI regulator and requiring firms using or developing AI to designate an AI officer, among other changes. How would these proposals impact your firm, if implemented?

Based on our preliminary review, we note that the proposals in the draft Bill are very high level, involving significant delegation powers to the Secretary of State and the proposed AI Authority, including the power to make further regulations. It is therefore difficult (if not impossible) to assess the likely impact until those powers have been exercised and relevant explanations provided. That said, this uncertainty in and of itself risks creating burden and expense as Anthropic and others strive to understand what compliance entails. And where the Draft Bill is more concrete, certain proposals—including in relation to certain audits, record-keeping and creation of AI officer duties—will also impose costs and administrative burden. We further note that there is potentially significant overlap with existing laws, as well as ongoing discussions on intellectual property. We believe it would be more efficacious for the ongoing consultation into intellectual property to conclude before offering a further opinion on how these overlaps may manifest. Nonetheless, we believe that appropriate and proportionate regulation is both a necessary and sensible step for governments to be taking as the technology develops and new use cases emerge.

12. What recommendations, if any, does your firm have for AI regulation and supervision in the UK?

We currently work with the UK Government through the Memorandum of Understanding we signed at the beginning of 2025. This work focuses on supporting the government to consider a range of use cases where the technology may be deployed for the benefit or service users, service quality or efficiency. Whilst any decision on regulation or supervision is a matter for the UK government, we have found the work and partnership of the UK AISI to be valuable and it has actively added to the quality of our models, and conversely, we

have added to the AISI's understanding of emergent technology changes. We have signed the EU AI Code of Practice and support its implementation. In a similar vein we have also recently supported the enactment of the California State Legislature Bill SB 53. As a general rule, and in line with our long stated position of supporting responsible development of AI, we are supportive of proportionate regulation and supervision.

I hope these answers are satisfactory. If you require further information or wish to clarify any points, then please don't hesitate to get in touch.

Thank you for your time and consideration.

Sincerely,

Daisy

Daisy McGregor
Head of UK Policy

Google Cloud Response to letter from the UK Treasury Select Committee on AI in Financial Services

October 2025

1. What impact will AI have on financial services?

The financial services industry is currently experiencing a profound transformation, largely propelled by the rapid advancements and increasing adoption of Artificial Intelligence (AI). Some of the largest opportunities include:

- Anti-Money Laundering and Fraud Detection
- Enhancing customer experience
- Advanced risk management and predictive analytics
- Innovation in Fintech and embedded finance
- Modernisation of financial market infrastructure
- We are also seeing continued uses of AI to improve employee productivity and the operational efficiency of both customer facing and internal use cases. There are numerous examples of the use of AI to support document processing including during mortgage origination to provide the consumer with greater flexibility in the evidence they provide to the mortgage lender.

Anti Money Laundering:

The effectiveness of AI in AML is demonstrated by the experience of HSBC working in partnership with Google Cloud to develop and implement AML AI. HSBC's implementation resulted in a 200-400% increase in the identification of actual financial crime risks, coupled with a reduction of over 60% in the volume of alerts generated. HSBC achieved faster detection times, reducing the period to just eight days after an initial alert, and gained the capability to identify complex criminal networks, a task that traditional systems struggled with. This improvement highlights AI's ability to distinguish genuine threats from false positives, allowing investigators to focus their efforts on truly suspicious cases.

Customer experience:

- ING Bank's implementation of a generative AI chatbot is aimed at enhancing consumers' self-service capabilities and ultimately improving their overall satisfaction.
- The Teachers Insurance and Annuity Association's collaboration with Google Cloud to utilize its Contact Center AI (CCAI) solution further illustrates this trend.
- A large UK insurer has used Google Cloud's generative AI to provide real-time suggestions to call center agents and automatically fill out insurance forms during live customer interactions.

Risk management and predictive analytics

The partnership between Google Cloud, Allianz, and Munich Re to revolutionize risk management in the insurance industry through cloud-based AI highlights the potential of these technologies in this sector. AI can help insurers better understand and quantify risks, leading to more accurate pricing, improved underwriting processes, and more effective risk mitigation strategies.

Beyond the insurance industry, AI is also being widely applied in anomaly detection to identify fraudulent transactions, financial crime, and cyber threats. By continuously monitoring and analyzing vast amounts of data, AI algorithms can detect unusual patterns that may indicate illicit activities, enhancing the security and integrity of the financial system.

Innovation/Fintech

The case of Fundwell, which uses Google Cloud AI to connect businesses with funding opportunities, exemplifies how AI can streamline and democratize access to financial services, a core tenet of the way fintech can offer genuine benefits to individuals.

One significant area of potential is in the capacity for AI algorithms to analyze various data points to match businesses with suitable funding options more efficiently than traditional lending processes.

Infrastructure
The collaboration between Bloomberg and Google Cloud to provide real-time data access is an example of this modernization effort. Timely and accurate data is paramount for effective decision-making in financial markets, and cloud infrastructure provides the scalability and speed necessary to process and disseminate vast amounts of information in real-time.

Integrating AI with this real-time data access can further enhance its value by enabling sophisticated analytics and providing actionable insights for event-driven trading and risk management strategies.

BNY Mellon's partnership with Google Cloud to transform the settlement and clearance processes in the US Treasury market demonstrates the potential of cloud technologies, potentially including AI, to improve critical financial infrastructure.

2. Does your firm have a specific strategy for AI in financial services?

Our answer to Question 1 outlines the main areas we are currently working with financial services organisations - our strategy is broadly to make AI solutions that are useful across those domains and future opportunities. As to a specific strategy for financial services, Google's foundation models are designed to address a range of requirements across multiple industries - we don't make a specific version of our products solely for financial services.

To give more strategic context to our answer above, our discussion with financial services customers starts from a recognition of three big operational opportunities:

1. Reimagining distribution and financial advice & guidance: Financial institutions are looking to bring their siloed data together in one system and easily connect it to AI and analytics tools to drive insights at speed and scale. These insights enable data driven, personalized advice and guidance to deliver exceptional experiences across all channels (e.g., tailored wealth management, insurance coverage, and market research, etc.).

2. Automation of middle and back office operations: Transforming efficiency with AI powered automation across heavily manual and complex workflows (e.g., contact center operations, lending

operations, etc.). Financial institutions can put AI agents in the hands of every employee with a unified platform & interface to search documents, databases, and SAAS applications, harvest key insights, and deploy AI agents to significantly improve productivity. They can modernize core systems (mainframe, etc.) in the cloud to enable the agility and resiliency financial institutions need to compete and win.

3. Improving financial and operational risk management: Financial institutions can leverage cloud native database services and tooling to streamline and automate the development of risk models / simulations, financial forecasts, and regulatory reporting, now accelerated by agentic workflows. They can combat financial crime to dynamically identify risk and reduce false positives. And they can leverage an AI-enabled, unified security platform to streamline threat detection and response and protect financial institutions' users, data, and endpoints.

3. Does your firm have systems in place to check or audit the outputs of its AI models?

We have developed an approach to AI governance that focuses on responsibility throughout the AI development lifecycle. Our ongoing work in this area reflects our own AI Principles and key concepts in industry guidelines like the NIST AI Risk Management Framework.

- **Governance and measurement**: our pre- and post-launch processes cover model and application requirements, with a focus on safety, privacy, and security. Post-launch monitoring and assessments enable continuous improvement and risk management. We apply multi-layered red teaming with both internal and external teams proactively testing AI systems for weaknesses and identifying emerging risks. This approach simulates real-world attacks, while content-focused red teaming identifies potential vulnerabilities and issues. Model and application evaluations are central to this measurement approach. These evaluations assess alignment with established frameworks and policies, both before and after launch.
- **Mapping risk**: We map AI risks through research and expert consultation, codifying these inputs into a risk taxonomy. A core component is risk research - we have published in over 300 papers, and this research directly informs our risk taxonomy, launch evaluations, and mitigation techniques.
- **Management:** We deploy and evolve mitigations to manage content safety, privacy, and security, such as safety filters and jailbreak protections. We often phase our launches with audience-specific testing, and conduct post-launch monitoring of user feedback for rapid remediation.

A key part of AI governance is ensuring that deployers of the technology have the controls to govern and manage the intended outputs of their models. Through our developer platform Google Cloud Vertex AI platform, we provide our customers with controls to integrate our technologies while ensuring their own internal and sectoral compliance. The customer defines the guardrails including prompt rewriters, safety filters, and data usage policies. Some of the key controls for developers are:

- **System Instructions:** Customers can provide clear and specific instructions to guide the model's behavior and ensure responses align with their policies. Well-crafted system instructions can be more effective than safety filters at producing safe and neutral outputs. For example, you can instruct the model to avoid certain topics or to adhere to a specific brand voice.

- **Prompt Engineering:** The design of prompts is also a key factor in influencing model responses. By providing clear instructions, few-shot examples, and structured prompts, users can guide the model towards more accurate and neutral outputs.
- **Safety Filters and Content Moderation:** Vertex AI includes built-in content filtering and safety attribute scoring. These allow customers to test Google's safety filters and define confidence thresholds appropriate for their use case.
- **Grounding:** To combat "model hallucinations" and improve factuality, models can be grounded to specific data. This helps to ensure that responses are based on provided information rather than the model's general knowledge.
- **Model Evaluation for Quality, Bias, and Integrity:** A comprehensive service designed to ensure the quality and integrity of all models throughout their lifecycle. Before a model is deployed, the service audits its statistical health, verifying that the model's predictions align accurately with known **ground truth** data. Beyond simple accuracy, this tool is vital for checking the model's **realistic representation** and integrity: it includes specialized checks for **safety and bias evaluation** against sensitive identity groups to ensure the outputs do not reinforce harmful stereotypes. After deployment, the service provides **continuous evaluation** by monitoring real-world production data. By measuring the statistical distance (skew or drift) between the training data and the live data, it detects **incongruencies**. This proactive audit helps practitioners rapidly identify when a model's performance is degrading or when fairness issues emerge, signaling the need for intervention.
- **Output Evaluation Tools:** Automatic Side-by-Side, or **AutoSxS**, is an advanced output evaluation tool that performs automated, large-scale **testing** on models and their outputs. For a given task or prompt, the system analyzes and compares the responses of two competing LLMs (Model A versus Model B) against predefined criteria such as relevance, coherence, and accuracy. The result is summarized in **win-rates**, providing an objective score on which model is qualitatively superior for the specific task. AutoSxS generates a detailed **judgment table** for auditing purposes, which includes the Autorater's decision, a **rationale or explanation** for that decision, and a confidence score allowing customers to select the best-performing model or prompt.

4. The Committee has received evidence warning that AI decision-making lacks transparency, potentially impacting consumers in financial services. How does your firm ensure transparency in the decision-making of its AI models?

As AI technology and discussions about its development and uses continue to evolve, we will continue to learn from our research and users, and innovate new approaches to responsible development and deployment. As we do, we remain committed to sharing what we learn with the broader ecosystem through public reporting and continuous engagement, discussion, and collaboration with the wider community to help maximize the benefits of AI for everyone.

We foster transparency and accountability throughout our AI governance processes. For our Gemini foundation models, we regularly publish external model cards and technical reports to provide transparency into model creation, function, and intended use. Technical reports provide details about how our most advanced AI models are created and how they function. This includes offering clarity on

the intended use cases, any potential limitations of the models, and how our models are developed in collaboration with safety, privacy, security, and responsibility teams. We are also investing in robust infrastructure to support data and model lineage tracking, enabling us to understand the origins and transformations of data and models used in our AI applications.

Promoting AI transparency with model cards

Model cards were introduced in a Google research paper in 2019 as a way to document and provide transparency about how we evaluate models. These cards offer summaries of technical reports in a "nutrition label" format to surface vital information needed for downstream developers or to help policy leaders assess the safety of a model.

Previous iterations of our model cards, such as one to predict 3D facial surface geometry and one for an object detection model, conveyed important information about those respective models. However, as generative AI models have advanced, we have adapted our most recent model cards, such as the card for our highest quality text-to-image model Imagen 3, to reflect the rapidly evolving landscape of AI development and deployment. While these model cards still contain some of the same categories of metadata we originally proposed in 2019, they also prioritize clarity, practical usability, and include an assessment of a model's intended usage, limitations, risks and mitigations, and ethical and safety considerations.

AI literacy

To complement our transparency measures, it is also critical that governments and industry continue to educate people about how to use AI, and its limitations. We have committed $120 million for AI education and training around the world. We have also launched AI training for businesses, developers, and younger learners.

5. The Committee has also received evidence from several groups arguing that AI models could perpetuate bias and discrimination against consumers in credit and insurance. How does your firm ensure that its AI models avoid bias and discrimination?

Our response to Question 3 sets out our approach to equipping customers with the tools and controls to support them to ensure that bias and discrimination is mitigated and addressed in the outputs of the models they train using our foundational models. This approach is consistent across any use case that our customers choose to address.

6. What preparations does your firm have in place for failures or outages in its cloud and AI systems?

Google Cloud's preparation for failures and outages is not a single plan, but a deeply ingrained cultural and architectural philosophy which covers: **Resilient Global Infrastructure**, **Services Designed for High Availability**, and **Transparent Partnership**.

Resilient-by-Design Global Infrastructure

The physical foundation of our cloud is engineered to withstand a wide array of failures, from minor hardware faults to large-scale regional disruptions.

- **Global Network of Regions and Zones:** Our infrastructure is organized into **Regions**, which are independent geographic areas around the world. Each Region consists of at least three **Zones**. A **Zone** is a fully independent infrastructure deployment with its own power, cooling, networking, and security.
- **Redundancy at Every Layer:** Within each Zone, every critical component is redundant. We have multiple, independent power sources, redundant cooling systems, and multiple network links. Server hardware is designed to ensure workloads automatically and instantly move off failing components.
- **A Private, Global Network:** All our Regions are connected by Google's own high-speed, private fibre optic network. This means customer data, by default, does not traverse the public internet when moving between our services. This private backbone enhances security and provides a more reliable and predictable network performance, which is crucial for service stability.

Services Designed for High Availability and Disaster Recovery

We build our services to leverage this resilient infrastructure, offering customers clear and robust options for continuity.

- **Regional and Multi-Regional Services:** Many of our core services are inherently regional or multi-regional. By hosting services across multiple Zones within a Region, applications can remain online even if an entire Zone fails. We also distribute user traffic across multiple Regions around the globe. If an application in one Region becomes unhealthy or unavailable, traffic is automatically redirected to a healthy Region, ensuring continuous service for end-users.
- **Resilience in AI Systems:** Our AI platforms, like Vertex AI, are built on this same resilient infrastructure.
    - **Model Serving:** AI models are deployed across multiple Zones, just like any other application, to ensure they remain available.
    - **Data and Training Pipelines:** The underlying data storage and processing services used for AI are designed for durability and availability, protecting against data corruption or loss.
    - **Monitoring and Management:** We employ sophisticated monitoring to detect not only if an AI service is offline, but also if its performance degrades.

Transparent Partnership and Shared Responsibility

Finally, we recognise that resilience is a shared responsibility between Google Cloud and our customers. We are committed to being a transparent and supportive partner:

- **Public Status Dashboard and Communication:** In the event of an outage, we provide prompt and transparent communication through our public Status Dashboard. We detail which services

are affected, the geographic scope of the impact, and provide regular updates on our progress toward resolution.

- **Architectural Guidance:** We provide extensive documentation, best practices, and expert support to help customers design and build applications that take full advantage of our resilient infrastructure.
- **Compliance and Certification:** We adhere to a broad range of global and regional compliance standards and certifications. Our operations are regularly audited by independent third parties, providing external validation of our security and resilience.

7. The Financial Stability Board and the Bank of England, among others, have said that AI-driven trading could cause market volatility and contagion. What actions, if any, has your firm taken to address this risk in its AI models?

We refer the Committee again to our response to Question 3. Google provides foundation models to our customers that enables them to train their own models and applications to address particular use cases, and controls to ensure the outputs of those models align with their intentions. Google doesn't have expertise in developing trading platforms or specific controls to address any related risks.

8. What interaction, if any, has your firm had with the Bank of England and / or the FCA on AI?

Google has engaged with the Bank of England and the FCA on a regular but ad hoc basis to support their understanding and consideration of AI. We remain open and committed to ongoing engagement and have offered to support initiatives such as the AI Consortium should this be of use to the FCA and Bank of England.

9. The Bank of England and the FCA have both told the Committee that certain AI providers could be brought into the new Critical Third Parties Regime if HM Treasury deems them to be critical to the UK economy. If HM Treasury designated your firm as a critical third party under this regime, how would that impact your firm?

We welcome the introduction of CTP to support the continued work by the Bank of England and the FCA to manage systemic risk. Financial institutions' reliance on Critical Third Parties (CTP) is deepening. Ensuring operational resilience of the sector and its increasingly complex ecosystem is of paramount importance. Whilst existing regulations include a range of measures for financial entities to manage operational resilience and third party risk, UK CTP provides a set of more third party-centric obligations that complement existing regulatory requirements and oversight.

As we await the HM Treasury's CTP designation process, Google Cloud has been preparing to comply with these regulations for some time. We're working to implement new obligations, complement our existing compliance program across security, resilience, and third-party risk, and proactively and regularly engage with the Bank of England, the Prudential Regulation Authority and the Financial Conduct Authority on the key aspects of the regime.

10.The Bank of England and the FCA take a "technology-agnostic" approach to AI regulation and supervision, meaning that they do not usually require or prohibit specific technologies. Do you agree with this approach?

Google supports supervisory authorities in adopting principle-based, proportionate, and technology-agnostic approaches to risk management. This stance is rooted in the belief that it fosters innovation by avoiding prescriptive rules that can quickly become obsolete and stifle the adoption of new, more efficient technologies. It allows for flexible, principles-based, and outcomes-focused frameworks that can accommodate the rapid pace of technological change and the diverse operating models of technology companies, which differ significantly from traditional financial institutions.

By focusing on outcomes rather than methods, this approach creates a durable and flexible framework that promotes resilience, encourages the adoption of advanced technologies, and accommodates the diverse operating models of modern ICT providers. The ultimate success of this regulatory philosophy rests on its execution. It requires a significant commitment from regulators to develop deep technical expertise and invest in the necessary resources for effective supervision. It also demands a new level of collaboration and communication between industry and supervisors to ensure clarity and consistency.

11.The Artificial Intelligence (Regulation) Bill, a Private Member's Bill in the House of Lords, proposes establishing a dedicated AI regulator and requiring firms using or developing AI to designate an AI officer, among other changes. How would these proposals impact your firm, if implemented?

AI is too important not to regulate — and too important not to regulate well. Regulation must be grounded in evidence and understanding of the full potential of this rapidly evolving technology. That's why Google supports well-designed, evidence-based regulation that fosters certainty and predictability within the AI ecosystem — for developers, for deployers, and for users. Smart regulation can promote responsible AI practices while ensuring that innovation can continue to flourish.

We recommend a hub-and-spoke AI governance architecture, where a "hub" of centralized AI technical expertise is maintained to support sectoral regulators in their regulation of specific AI applications — the "spokes". The hub of centralized AI expertise can lay out sector-agnostic AI risk management best practices and frameworks, and provide consulting expertise within government for technical advice on the capabilities and implications of evolving AI technologies. This function should complement the "spokes" — sectoral regulators tasked with regulating the use of AI in their specific areas of expertise. The risks of AI are highly context-dependent — issues in finance will differ from those in healthcare or transportation. Sectoral regulators are best placed to assess context-specific uses and impacts of AI in their respective domains, and whether and how best to regulate them. For instance, health-focused agencies are best positioned to evaluate the use of AI in medical devices. Similarly, transportation agencies have expertise in evaluating the deployment of autonomous vehicles.

The UK's current approach - that is context-specific and empowers existing regulators - is right. To ensure society reaps the full benefits of AI, we require clear, context-based, and flexible regulation that keeps pace with the development of technology, while also mitigating potential risks.

12. What recommendations, if any, does your firm have for AI regulation and supervision in the UK?

1. **Interpret Existing Frameworks.** We encourage policymakers to consider how existing rules and guidance already apply in this space, and recognise in guidance how existing privacy principles can be applied to AI systems. Many existing regulations, such as the UK GDPR, already contain robust provisions for legal basis for processing, data minimization, and data subject rights, etc.

2. **Transparency / Encourage Accessible and High-Level Action Explanations.** To build user trust and accountability, policymakers should encourage AI systems to provide clear, high-level explanations for their actions involving personal data. This does not mean these explanations must be surfaced in real-time for every single action. Instead, they should be accessible to the user in a privacy policy, or at a more granular level, in a history or a dedicated log. The explanations themselves should be concise and avoid excessive detail that could overwhelm the user.

3. **Data Limitations / Flexibility** Existing data protection laws place limitations on the collection and use of personal data, and these principles are already a guiding lens for AI development. The policy focus should be on ensuring that an AI's declared goal, particularly in the agentic space (e.g., "find and book a restaurant") is unequivocal and that its actions and data access are made transparent to the user. A system narrowly designed to manage your calendar, for instance, should not access your financial data unless instructed to do so. At the same time, a general agent that is designed to work with multiple datasets for multiple purposes should be allowed to do so when that is the expectation of the consumer. The declared goal and user expectations should govern the activity.

4. **Training** Responsible development of AI requires access to a diverse range of training data to learn from. Training a model requires billions of data points to provide information, culturally relevant content, varied perspectives, and different forms and use of language to help guard against bias and take into account societal norms. Without access to the open web, our training of AI models would be limited and consequently the output would be skewed, which could lead to perverse results.

09 October 2025

Meta

Dear Dame Hillier

Thank you for your recent letter regarding the use of AI in the financial services sector. I have set out answers to your specific questions below. I hope this information is of use in your ongoing inquiry.

1. What impact will AI have on financial services?

As with many sectors across the economy AI is already having a meaningful impact on financial services. The financial sector has been using machine learning and AI models for fraud monitoring, risk management, financial crimes prevention, and other use cases for decades. We are now seeing providers using more advanced models to improve their fraud detection, offer better customer service and improve credit risk assessments. We also believe that AI will have profound impacts on the broader economy in ways that will be relevant to the financial sector. For example, payments and commerce conducted by AI agents are likely to provide financial institutions with new business opportunities and customer experiences.

2. Does your firm have a specific strategy for AI in financial services?

Meta is not a financial institution and does not have a specific strategy for AI in financial services. Meta open sources our foundational large language models and they can therefore be downloaded, fine tuned and used by multiple sectors.

3. Does your firm have systems in place to check or audit the outputs of its AI models?

Meta has built an end-to-end governance processes to ensure accountability for how our AI models and systems are built. I have set out the core components of this below.

- Meta's AI governance program is coordinated with our existing privacy review program, and is driven by a comprehensive AI Risk Assessment process. This process includes mapping out the risks associated with launching new AI models, and documenting the mitigations that are applied and the residual risks that

remain.

- The use of the term "AI Risk Assessment" in the context of our AI governance program refers to an essential process for systematically identifying and evaluating a comprehensive list of risks associated with AI offerings, and documenting relevant mitigations and residual risks.

- The risk assessment process involves multi-disciplinary engagement, including internal and, where appropriate, external experts from various disciplines (which could include engineering, product management, compliance and privacy, legal, and policy) and company leaders from multiple disciplines.

- At a high level, Meta's AI Risk Assessment process generally includes the documentation of:
    - Model, data, and release details
        - A description of the model, its capabilities
        - Technical characteristics, including modality/ies, open sourcing, training data.
        - Release plans/narrative (Open Source vs closed, Internal vs External Release, accompanying research paper, etc.)
        - Details on data used for training
    - Risk Categories
        - Meta evaluates risk and mitigation strategies across a number of risk categories, including: cybersecurity and chemical and biological.
    - Risk mitigations/controls - Information about the mitigations/controls in place, and metrics demonstrating their efficacy. Examples of potential controls include limitations on training data, retention policies, and transparency measures.
    - Evaluations and Red teaming
        - Evaluations: Assess performance using benchmarking standards to evaluate performance of the model on relevant benchmarks, and to measure the efficacy of mitigations
        - Red teaming reports against relevant risk categories
    - Evaluation of residual risk

- In Meta's standalone Meta AI app, we apply safeguards at both the prompt and response level. To encourage Meta AI to share helpful and safer responses that are in line with its guidelines, we implement filters on both the prompts that users submit and on responses after they're generated by the model, but before they're shown to a user.

- We also engage in a number of fora - including the Frontier Model Forum and the Partnership on AI - to share our expertise and learn from fellow experts across numerous topics related to AI governance. We also actively engaged with the National Institute of Standards and Technology (NIST) in the creation of the NIST AI Risk Management Framework, and will continue to engage on key issues related to AI governance.

Frontier AI Framework

- For the development of frontier models, which potentially have heightened risks that require deeper evaluation and mitigation, we have established additional assessment and documentation requirements. These additional assessment requirements may include heightened impact assessments, testing, and documentation (i.e., internal and external Red teaming results and benchmark evaluations), incident reporting, and AI lifecycle planning (including risk evaluation across the lifecycle of the model) as applicable.

- Our Frontier AI Framework focuses on the most critical risks in the areas of cybersecurity threats and risks from chemical and biological weapons. By prioritizing these areas, we can work to protect national security while promoting innovation. Our framework outlines a number of processes we follow to anticipate and mitigate risk when developing frontier AI system, for example:
    - Identifying catastrophic outcomes to prevent: Our framework identifies potential catastrophic outcomes related to cyber, chemical and biological risks that we strive to prevent. It focuses on evaluating whether these catastrophic outcomes are enabled by technological advances and, if so, identifying ways to mitigate those risks.
    - Threat modeling exercises: We conduct threat modeling exercises to anticipate how different actors might seek to misuse frontier AI to produce those catastrophic outcomes, working with external experts as necessary. These exercises are fundamental to our outcomes-led approach.
    - Establishing risk thresholds: We define risk thresholds based on the extent to which our models facilitate the threat scenarios. We have processes in place to keep risks within acceptable levels, including applying mitigations.
    - While the focus on this framework is to help mitigate severe risks, it's important to remember that the primary reason we're developing these technologies is because they have the potential to greatly benefit society.

4. The Committee has received evidence warning that AI decision-making lacks transparency, potentially impacting consumers in financial services. How does your firm ensure transparency in the decision-making of its AI models?

At Meta we open source our large language foundation (Llama) models because we believe that providing transparency about how this technology is built, trained and how it works is vital to both building trust and building the most robust models possible.

We are also committed to providing greater transparency to people through measures that explain how our AI-powered products work and enable users to understand when they are engaging with AI-generated content.

As the capability and ubiquity of AI has grown, so have calls for greater transparency and control over how the technology is developed and used - from developers and end users to governments and civil society. If people cannot understand how or why an AI model or product behaves the way it does, or feel that they or others have insufficient control over it, it will be difficult to build the trust necessary for the widespread uptake of these technologies.

We've developed and implemented several transparency measures aimed at addressing these concerns and increasing trust in our AI technologies, including:

- Our commitment to providing users with greater transparency and control in relation to our AI technologies is reflected in our belief that research and innovation in this field - much of which underpins our products - should be open, so that others can evaluate, improve, and build off each other's efforts.

- For example, we open source our Llama foundation models together with a number of artefacts - including a Responsible Use Guide - to help those developing and using the model to understand how it was trained, how performance and potential risks were measures, the mitigations taken to reduce identified risks, and the mitigations that others can take when fine-tuning the model for their specific applications. We believe that this open source approach enhances accountability, improves model performance, and ultimately helps to build trust in how AI is developed and used.

- System Cards. An artefact that contains information and actionable insights anyone can use to understand and customise their specific AI-powered experiences in our products.

- Our Generative AI guide in Privacy Center and other transparency resources are there for people to understand how we built our AI models, how our AI features work, and what choices and data privacy rights they have.

- Meta has joined the Coalition for Content Provenance and Authenticity (C2PA) underscoring its commitment to content provenance by joining the organisation as a new steering committee member. As a global standards body advancing transparency online, the C2PA publishes and maintains Content Credentials, a rapidly growing digital standard that provides information about how and when digital content was created or modified. Meta joins a group of industry leaders on the C2PA steering committee, including Adobe, BBC, Google, Intel, Microsoft, OpenAI, Publicis Groupe, Sony, and Truepic.

5. The Committee has also received evidence from several groups arguing that AI models could perpetuate bias and discrimination against consumers in credit and insurance. How does your firm ensure that its AI models avoid bias and discrimination?

Meta does not develop AI models specifically for the credit and insurance industries. More generally Meta employs a comprehensive, multi-layered approach to prevent and mitigate bias and discrimination in its AI systems informed by both internal research and external stakeholder engagement.

One way we are addressing AI fairness through research is the creation and distribution of more diverse datasets. Datasets that are used to train AI models can reflect biases, which are then passed on to the system. But biases might also be due to what isn't the training data. A lack of diverse data — or data that represents a wide range of people and experiences — can lead to AI-powered outcomes that reflect problematic stereotypes or fail to work equally well for everyone.

By fostering the development of more inclusive datasets, we can create AI systems with the potential to bring the world closer together, helping people communicate across languages and cultures and creating experiences that reflect the diversity of the people who use Meta's platforms.

Because AI systems are complex, it is important that we develop documentation that explains how systems work in a way that experts and nonexperts alike can understand. One way we've done this is by open sourcing our models and publishing documentation

like model and system cards which provide insight into an AI system's underlying architecture and helps better explain how the system operates. Because our models are open sourced this also means that organisations are able to train and fine tune applications they build themselves and better understand the inner workings of the model.

By sharing these sorts of details we aim to accelerate research around language generation systems so the broader field can work toward making these systems safer, more useful, and more robust.

The rapid advance of emerging technologies makes it difficult to fully understand and anticipate how they might eventually impact communities around the world. To help develop forward-looking policies around the development and use of new technology Meta supports Open Loop, a global experimental governance program. By involving governments, tech companies, academia, and civil society, Open Loop aims to connect tech and policy innovation for closer collaboration between those building emerging technologies and those regulating them.

These Open Loop programs are emblematic of Meta's evidence-based, broadly collaborative approach to establishing standards around the future of AI governance, and we look forward to continuing these programs in partnership with a broad selection of participating regulators, companies, and expert stakeholders.

6. What preparations does your firm have in place for failures or outages in its AI systems?

Meta's approach to AI failures and outages is comprehensive, combining technical infrastructure, automated and manual recovery, robust incident response, disaster recovery planning, and a culture of continuous improvement. This ensures that AI systems are resilient, recover quickly from failures, and that learnings from incidents are systematically incorporated into future operations. Meta's Llama models are also open source, and are therefore hosted by users on their own infrastructure.

7. The Financial Stability Board and the Bank of England, among others, have said that AI-driven trading could cause market volatility and contagion. What actions, if any, has your firm taken to address this risk in its AI models?

Meta does not develop models for AI driven trading but when designing and training our models we test and red team them against relevant risk categories and to test the effectiveness, performance and efficiency of mitigations.

Meta has also released Llama guard, our latest safeguard model with improved inference for detecting problematic prompts and responses. Amongst other categories Llama guard is trained to filter "specialised advice" so that AI models should not create content that contains specialised financial, medical or legal advice.

8. What interaction, if any, has your firm had with the Bank of England and / or the FCA on AI?

We meet regularly with the FCA to discuss a range of issues, primarily these are centred around the work we do to prevent and disrupt fraud and this includes the use of AI. We have also met with the Bank of England on a number of occasions to discuss AI and financial services.

9. The Bank of England and the FCA have both told the Committee that certain AI providers could be brought into the new Critical Third Parties Regime if HM Treasury deems them to be critical to the UK economy. If HM Treasury designated your firm as a critical third party under this regime, how would that impact your firm?

Meta has not assessed the impact of being designated a critical third party and is not aware of any current proposals to introduce such a designation.

10.The Bank of England and the FCA take a "technology-agnostic" approach to AI regulation and supervision, meaning that they do not usually require or prohibit specific technologies. Do you agree with this approach?

Yes. Tech neutrality in financial regulation is a time-tested approach that appropriately balances risk management and innovation. The only way to future proof this kind of regulation is to regulate the impact rather than the tech itself. We applaud UK regulators' stance in this area and believe it remains a key to enabling innovation in the UK financial services sector.

11.The Artificial Intelligence (Regulation) Bill, a Private Member's Bill in the House of Lords, proposes establishing a dedicated AI regulator and requiring firms using or developing AI to designate an AI officer, among other changes. How would

these proposals impact your firm, if implemented?

Meta has not assessed the impact of the changes proposed in the Private Members Bill referenced but would note that there are already a number of different regulators who already have powers covering a number of issues the committee is referencing including but not limited to Ofcom, the Information Commissioner's Office, the Financial Conduct Authority and the Competition and Markets Authority. Meta regularly interacts with each of these regulatory bodies.

12. What recommendations, if any, does your firm have for AI regulation and supervision in the UK?

Meta supports responsible, proportionate, and innovation-friendly regulation that is focussed on preventing harm and mitigating risk whilst also maximising the potential of AI to drive growth. We believe the emphasis of AI regulation should be on the outcomes sought (e.g., harm prevention, risk mitigation) rather than prescriptive rules that could stifle innovation or be quickly outdated by technological advances.

As AI innovation increases at pace it is important to ensure any AI regulation is consistent with international frameworks (such as the OECD AI Principles) to avoid a fragmented regulatory landscape and to promote innovation globally. It is also important to acknowledge that AI may already fall under a number of existing regulatory regimes - including the Online Safety Act so it will be important to ensure any future regulation avoids duplication or cutting across those existing regimes.

Meta believes that the UK has a real opportunity to position itself between the US's market-led model and the EU's prescriptive AI Act. By adopting proportionate, practical rules that balance real risks with innovation, the UK can become a global leader in AI governance while attracting investment. To achieve this, the UK should establish a pro-growth regulatory framework that:

- Drives innovation while protecting against tangible harms.
- Uses an evidence-first approach, reviewing existing laws before creating new burdens.
- Applies risk-based, flexible rules, particularly for SMEs.
- Keeps copyright and AI safety legislation separate — avoiding an omnibus Bill that stifles clarity.
- Coordinates across regulators through a Digital Regulation Cooperation Forum-style body to ensure coherence.

Sincerely,

Rebecca Stimson
Director of UK Public Policy,
Meta

Microsoft Limited
Microsoft Campus
Thames Valley Park
Reading RG6 1WG
Tel 0344 800 2400

Dame Meg Hillier MP
Chair, Treasury Select Committee
Houe of Commons
London SW1A 2AA

09 October 2025

Dear Dame Meg,

Ref. Artificial Intelligence in Financial Services

Thank you for your letter of 17 September to Darren Hardman on the use of artificial intelligence in financial services. Darren has asked me to respond in his absence. I have discussed the questions you raise with our teams at our headquarters in Redmond covering our responsible AI, our product development and engineering and our specialists supporting the financial services industry. I am pleased to attach a memo which provides full answers to your questions.

Please do not hesitate to contact me should you want clarification on any of these topics.

Yours sincerely,

Hugh Milward
Vice President, External Affairs, Microsoft UK

Microsoft

Microsoft Response to the Treasury Select Committee Inquiry into AI in the Financial Services Sector

Microsoft would like to thank the Treasury Select Committee for the opportunity to provide input on its inquiry into the use of artificial intelligence (AI) in the UK financial services sector.

As a technology provider of AI-enabled products and services, Microsoft is not a financial institution. Responses in this document are based on Microsoft's experience in supporting financial institutions as they adopt and implement AI technologies. Microsoft's approach is centred on enabling the responsible use of AI, guided by its established Responsible AI Principles, which govern the development and deployment of its AI solutions. Microsoft is committed to equipping financial services industry participants with the tools and expertise to innovate safely and responsibly while ensuring compliance with regulatory frameworks and ethical AI standards.

EXECUTIVE SUMMARY:

Microsoft's response to the UK Parliament's Treasury Committee letter on AI in financial services outlines the company's commitment to responsible AI use. As a technology provider, Microsoft supports financial institutions in adopting AI technologies while enabling compliance with regulatory frameworks and ethical standards. The response emphasises the importance of responsible AI principles in the development and deployment of AI solutions. Microsoft highlights several key themes, including the transformative impact of generative AI on the industry. Additionally, the response discusses the challenges and barriers to AI adoption, including legacy systems and the need for employee upskilling, and also underscores the importance of collaboration between financial institutions, regulators, and technology providers to ensure the safe and effective use of AI. Microsoft remains committed to supporting the financial services industry in navigating the complexities of AI implementation and achieving compliance with regulatory requirements.

DETAILED RESPONSE:

There are 12 questions posed in the inquiry into the use of artificial intelligence (AI) in the UK financial services sector. While Microsoft has addressed the specific questions in the inquiry letter, it is important to note that the response is intended to offer general input on the subject. The links used throughout the document are referenced at the end for ease of reference.

RESPONSES TO THE QUESTIONS

1. What impact will AI have on financial services?

The power of generative AI combined with rich industry data is transforming financial services by enabling professionals to quickly access and synthesise critical insights for faster, more efficient, and better-informed decision-making. Innovation is happening in every corner of financial services. According to the Gartner 2025 CIO Agenda: Top Priorities and Technology Plans for Banking (available subject to registration), "the biggest expected changes in

technology investments are generative AI (39%), cyber security/information security (34%), and AI (33%)".

Across the industry, Microsoft's customers are realising important productivity benefits in AI innovation. For example, firms are reshaping how they deliver experiences, products, and services through generative AI:

- Moody's launched a custom enterprise copilot that is enhancing productivity and innovation for its 14,000 employees.

- ERGO Insurance has revolutionised its customer service in just four months using an AI Virtual Agent powered by Azure, developed by EBO.

- Virgin Money has developed an award-winning virtual assistant using Microsoft Copilot Studio, integrated with Microsoft Dynamics 365 Customer Service to enhance the omnichannel customer experience.

- First National Bank has improved customer communications with Microsoft Copilot for Sales, ensuring client responses reflect understanding of their requests.

- CommBank is using generative AI to deliver personalised customer experiences and help protect against fraud, scams, and financial abuse.

Authorities in the financial services sector are also increasingly leveraging AI across various use cases, particularly in supervisory technology (SupTech). The adoption of SupTech by supervisory authorities has grown significantly, with the recent Financial Stability Board (FSB) AI paper noting that 59% of surveyed authorities reported its use in 2023 - a 5% increase from 2022. Key developments include faster indicators for real-time economic monitoring, the use of alternative data for supervisory assessments, and experimentation with Large Language Models (LLMs) for research, statistical analysis, and drafting. Generative AI is also being applied during inspections, streamlining information collection and allowing regulators to extract relevant information from extensive inspection documents and generate summaries or draft reports. Additionally, LLMs have potential applications in improving data quality assessments and tailoring regulatory reporting error messages to specific jurisdictions and scenarios.

Generative AI is also enabling broad access to safe and effective use of AI for smaller to mid-sized firms, as well as the largest enterprises. For example, Microsoft's AI products are available to financial institution customers of all sizes, from the largest global banks and insurers to local community banks and state insurers focused on serving the needs of their local customers. In general, most financial institutions do not find it cost-effective or practical to develop AI models internally. As a result, they typically use AI models provided through Software as a Service (SaaS) from AI application vendors or cloud providers such as Microsoft. This delivery model allows institutions to access AI services without the significant investment of time and resources required for in-house development, making various AI products available to organizations of different sizes.

Financial institutions can face a range of challenges in adopting AI. The most pressing challenges are driven by the maturity and sophistication of the institution's existing technology infrastructure. Legacy systems often need upgrading or replacement before AI can be implemented, regardless of institution size. Employee upskilling is also necessary. Despite these hurdles, financial institutions are generally equipped to manage AI within existing governance frameworks; the sector has already developed mature approaches to implementing traditional machine learning and predictive AI models to ensure safe and effective use. The adoption of generative AI, supported by cloud-based controls, allows smaller enterprises to scale responsibly and benefit from advanced AI solutions.

2. Does your firm have a specific strategy for AI in financial services?

Microsoft appreciates that, with all AI systems, it is critical to understand their capabilities as well as their limitations and potential risks. As a leader in responsible development and use of AI in financial services for nearly a decade, Microsoft aims to be a trusted partner in helping financial services unlock the benefits of AI as well as identifying and developing solutions to mitigate risks – so that financial institutions and their regulators can be confident in the use of AI in financial services. These efforts are informed by Microsoft's deep knowledge of financial institutions' technology needs, the ways financial institutions explore and contract with technology providers, and its interaction with financial regulators across the world.

The lessons Microsoft has learned from this outreach informs its Responsible AI Initiative, a programme that helps ensure its AI systems are responsible from the design phase upwards. At the heart of Microsoft's Responsible AI Initiative are six foundational principles to guide AI development and use: (1) Fairness, (2) Reliability and Safety, (3) Privacy and Security, (4) Inclusiveness, (5) Transparency, and (6) Accountability. Based on these principles, Microsoft launched the Microsoft Responsible AI Standard, a multi-year effort to define product development requirements for responsible AI. The Microsoft Responsible AI Standard consists of goals and requirements for each of the six principles and is intended to function as a checklist or scorecard for companies developing and implementing AI systems.

3. Does your firm have systems in place to check or audit the outputs of its AI models?

Microsoft's AI products and services are not developed for specific use cases, therefore Microsoft supports a 'shared responsibility' model. This means assuming responsibility for certain elements of AI risk management, however sector-specific compliance and other risks are driven by how AI models are implemented, deployed, and used; decisions that remain solely within the domain of financial institutions. Nevertheless, Microsoft provides broad information on certain AI risks based on Microsoft's experience as developers of the underlying technology.

At Microsoft, earning trust in the age of AI begins with keeping people at the centre of how Microsoft designs, develops, and deploys technology, a practice rooted in its first draft of AI principles in 2016. Formally adopted in 2018, these principles, fairness, transparency, accountability, reliability and safety, privacy and security, and inclusiveness, continue to serve as Microsoft's enduring North Star. The foundation of Microsoft's governance efforts is

the Microsoft Responsible AI Standard, first developed in 2019 and publicly released in 2022. The Standard is Microsoft's internal playbook for translating these principles into practice, guiding teams across the company to build AI systems responsibly and consistently. To ensure its AI models perform safely and as intended, Microsoft applies a multi-layered assurance approach that combines red teaming, independent audits, quantitative evaluations, and external expert review. Red teaming focuses on risk identification through adversarial testing, while systematic measurement and evaluation help quantify and mitigate risks at scale. Together, these mechanisms enable continuous validation of model outputs and strengthen accountability across the AI lifecycle.

4. **The Committee has received evidence warning that AI decision-making lacks transparency, potentially impacting consumers in financial services. How does your firm ensure transparency in the decision-making of its AI models?**

Microsoft AI technologies are designed to help financial institutions implement AI responsibly, grounded in the three pillars of transparency: traceability, communication, and intelligibility. Microsoft's platforms provide integrated tools and governance practices that enable institutions to document, monitor, and communicate AI system behaviour, helping them meet both regulatory and ethical expectations. While research confirms that ungrounded outputs are an inherent feature of generative AI models, these can be significantly reduced through continuous monitoring and targeted mitigations.

Doing so effectively requires scalable detection mechanisms that go beyond manual review. Delivering such transparency in practice requires a structured approach built on traceability, communication, and intelligibility, the three pillars that guide Microsoft's responsible AI practices.

- **Traceability:** Transparency begins with robust traceability. Microsoft documents goals, design choices, and assumptions throughout the AI lifecycle to ensure that decisions are auditable and comprehensible. Microsoft's documentation artefacts, such as model cards, transparency notes, and technical reports, record system capabilities, limitations, and intended uses in a consistent and interpretable manner.

- **Communication:** Transparency also depends on open and proactive communication. Microsoft is committed to ensuring that users, customers, and regulators understand when, why, and how AI systems are developed and deployed, as well as their limitations. Product-level disclosures, in-product notices, and regular updates to transparency documentation support informed oversight and responsible deployment. Annual transparency reports and FAQs further keep stakeholders informed of significant developments and emerging risks.

- **Intelligibility:** Transparency finally depends on AI systems remaining understandable, empowering developers and decision-makers to evaluate trustworthiness and performance. Microsoft's system cards and capability evaluations make model behaviour interpretable for engineers, compliance teams, and executives. To strengthen this pillar, Microsoft provides groundedness detection capabilities within Azure AI Content Safety, designed to identify ungrounded

statements in generative AI outputs when using grounded documents, such as in Q&A Copilots or document summarisation applications.

Microsoft continuously refines its governance and documentation practices to reflect evolving capabilities, risks, and regulatory expectations, ensuring that transparency remains a dynamic and evolving discipline across the AI lifecycle.

5. **The Committee has also received evidence from several groups arguing that AI models could perpetuate bias and discrimination against consumers in credit and insurance. How does your firm ensure that its AI models avoid bias and discrimination?**

Microsoft mitigates unfairness by assessing how bias may be introduced across the AI lifecycle—from data collection to model design and system integration, and by implementing safeguards to detect, measure, and reduce it. Microsoft applies assessment tools, diverse evaluation datasets, and structured review processes to identify potential sources of bias and improve model performance across groups. Microsoft continually monitors AI models for drift, deterioration, or emerging risks, and conducts regular testing not only at launch but throughout the lifecycle of the system, with specific attention paid to identifying and mitigating bias in outputs.

For sensitive use cases, such as those involving consequential impacts on people's rights, opportunities, or access to services, Microsoft applies additional layers of governance through Microsoft's Sensitive Uses Review Program (as detailed in the Microsoft Responsible AI Standard), ensuring that these systems are subject to heightened scrutiny, multidisciplinary review, and human-in-the-loop decision-making to promote fair, transparent, and accountable outcomes.

Together, these safeguards, supported by ongoing monitoring, evaluation, and transparent documentation, help ensure that Microsoft's AI systems remain fair, reliable, and aligned with evolving standards and expectations. Microsoft encourages institutions to implement strong governance controls, including human-in-the-loop decision-making for high-impact scenarios, to ensure that consequential outcomes are not determined solely by automated systems.

6. **What preparations does your firm have in place for failures or outages in its cloud and AI systems?**

Consistent with Microsoft's principles-driven and risk-based approach, Microsoft applies existing, technology-neutral operational risk management frameworks to prepare for and mitigate failures or outages in both cloud and AI systems. These frameworks provide robust mechanisms for resilience, recovery, and continuity planning, and Microsoft finds they are equally effective across technologies without the need for AI- or cloud-specific requirements.

For example, Microsoft's approach to operational resilience and business continuity in AI solutions, such as Azure AI, is aligned with that used its broader cloud services. This ensures consistent oversight and accountability across all critical systems. At this time, Microsoft is

not aware of material use cases where firms rely solely on AI to ensure service continuity. Microsoft's Responsible AI Standard also requires all teams to leverage dedicated tools and escalation processes for incident management. In the event of a failure or outage, incidents are triaged by dedicated response specialists, who coordinate root-cause analysis, mitigation, and communication. Post-incident reviews are conducted to capture lessons learned and drive continuous improvement, with repair items tracked by impact severity and time frame. All employees also have access to multiple avenues for raising concerns, including an anonymous reporting channel and direct escalation to Microsoft's Office of Responsible AI.

7. The Financial Stability Board and the Bank of England, among others, have said that AI-driven trading could cause market volatility and contagion. What actions, if any, has your firm taken to address this risk in its AI models?

This question is highly application and market-specific. AI-driven trading relates specifically to regulatory requirements for investment services, including both retail and wholesale solutions - for example, the EU requirements regarding algorithmic trading, including high-frequency algorithmic trading, are set out under MiFID II. As a technology provider serving financial institutions, Microsoft applies a principle-driven, risk-based operational frameworks to its AI models that help financial institutions using its AI systems to understand model behaviour and take informed decisions, reducing the likelihood of unintended market impacts. However, financial institutions, as the deployers of the technology, are ultimately responsible for taking appropriate AI-literacy and risk management steps on the use of AI in trading or its impact on market volatility and contagion (such as those highlighted by ESMA in its guidance on using AI for investing).

8. What interaction, if any, has your firm had with the Bank of England and / or the FCA on AI?

Microsoft's previous interactions with the Bank of England and / or the FCA are related to sharing of its experience in supporting financial institutions as they adopt and implement AI technologies. Microsoft regularly responds to regulatory consultations and participates in industry dialogue.

9. The Bank of England and the FCA have both told the Committee that certain AI providers could be brought into the new Critical Third Parties Regime if HM Treasury deems them to be critical to the UK economy. If HM Treasury designated your firm as a critical third party under this regime, how would that impact your firm?

Microsoft welcomes the UK Critical Third Parties (UK CTP) regime upon designation. With increased focus on operational resilience and third-party dependencies, Microsoft is well-prepared for regulations such as the EU's Digital Operation and Resilience Act (DORA). Microsoft sees this as an opportunity to strengthen operational resilience further, contribute to the sector, and learn from other participants.

Microsoft's approach is threefold:

(i) Engage with regulators and customers to build trust and help shape policy.
(ii) Align Microsoft's cloud services and contractual terms with applicable regulatory requirements.
(iii) Embed compliance capabilities directly into Microsoft's offerings to ensure ongoing adherence and resilience.

10. The Bank of England and the FCA take a "technology-agnostic" approach to AI regulation and supervision, meaning that they do not usually require or prohibit specific technologies. Do you agree with this approach?

Microsoft supports a technology-agnostic approach to AI regulation and supervision. It is essential that regulatory frameworks remain principles and outcomes based, ensuring they are flexible enough to accommodate ongoing technological advances while maintaining robust oversight. Microsoft encourages regulators to continue fostering global interoperability by aligning with international counterparts and leveraging mechanisms such as the Financial Stability Board (FSB) and the Bank for International Settlements (BIS) to promote consistent standards and supervisory expectations across jurisdictions.

Rather than introducing entirely new, prescriptive requirements, Microsoft believes that financial services regulators should build upon existing frameworks including those related to model risk management (MRM), outsourcing, and operational resilience to ensure AI is deployed safely and responsibly. This approach will help avoid regulatory fragmentation, support innovation, and provide firms with the clarity and consistency needed to operate at a global scale.

11. The Artificial Intelligence (Regulation) Bill, a Private Member's Bill in the House of Lords, proposes establishing a dedicated AI regulator and requiring firms using or developing AI to designate an AI officer, among other changes. How would these proposals impact your firm, if implemented?

Microsoft welcomes ongoing discussions and further dialogue across regulatory and industry stakeholders to ensure a balanced approach that fosters innovation while safeguarding the financial services sector, consumers, and technology providers alike.

As a technology provider, Microsoft believes that effective supervision of AI-enabled services and products requires a risk-based regulatory framework that is flexible enough to apply to different use cases while remaining predictable. The real-world impact of AI depends not only on the developer of the model or system, but also on how it is implemented in specific services or applications. For example, in the financial services sector, this means that every financial institution using AI must manage not only the technology itself, but also its broader legal, compliance, and risk management responsibilities.

Financial services already have mature risk management frameworks. Financial services regulators should prioritise existing principles and rules for risk management. For example, risk-based guidance on ethical AI use - including fairness, bias, and explainability - may be more effective than imposing additional prescriptive rules. Such considerations should be assessed through broad industry engagement and regulatory dialogue, including forums such as the G7, Financial Stability Board, BIS, IOSCO, and IAIS.

Microsoft supports governance mechanisms that enable clear accountability and coordination. This could include responsibility for sharing best practices, ensuring transparency and liaising with regulators and internal stakeholders on AI risk management. Microsoft has also established its internal governance structure accordingly to reflect these principles. At the same time, it is important that such a role complements existing governance and risk management functions rather than imposing duplicative or prescriptive requirements, to avoid creating unnecessary operational burden.

12. What recommendations, if any, does your firm have for AI regulation and supervision in the UK?

Microsoft supports a model-level, frontier-focused approach that sets clear, technology-neutral expectations for developers of highly capable systems. Government can incentivise adoption of responsible scaling policies and rigorous pre-deployment and continuous evaluations complemented by independent testing. Proportionate transparency should support supervisory review and comparability across providers. By centring supervision on evidence from evaluations, independent testing, and verified disclosures, the UK can strengthen oversight and accountability in a risk-based, proportionate, and interoperable way.

Microsoft

List of References and Sources – Last Access: 8 October 2025

Responsible AI at Microsoft

https://www.microsoft.com/en-us/ai/responsible-ai?msockid=0f6adfe057646aa124e9cc2156bb6bc0

Gartner 2025 CIO Agenda: Top Priorities and Technology Plans for Banking (available subject to registration)

https://www.gartner.com/account/signin?method=initialize&TARGET=http%3A%2F%2Fwww.gartner.com%2Fdocument%2F5765215%3Fref%3DsolrAll%26refval%3D436204119

Leading AI Transformation: Moody's | Microsoft

https://www.linkedin.com/pulse/leading-ai-transformation-moodys-microsoft-judson-althoff-mdi3c/?trackingId=fyFgrr5BTsSTs2DGmx1XKA%3D%3D

Revolutionizing customer service: ERGO Insurance uses AI and Microsoft Azure to provide 24/7 personalized support

https://www.microsoft.com/en/customers/story/17040327631956448l4-ergo-azure-insurance-en-greece

Azure

https://azure.microsoft.com/en-us

Redi, set, go: Virgin Money delivers exceptional customer experiences with Microsoft Copilot Studio

https://www.microsoft.com/en/customers/story/17958361410960130l8-virgin-money-dynamics-365-customer-service-banking-and-capital-markets-en-united-kingdom

Microsoft Copilot Studio

https://www.microsoft.com/en-us/microsoft-365-copilot/microsoft-copilot-studio

Dynamics 365 Customer Service

https://www.microsoft.com/en-us/dynamics-365/products/customer-service

First National Bank enhances customer communications with Microsoft Copilot for Sales

https://www.microsoft.com/en/customers/story/17619315882309838l5-first-national-bank-dynamics-365-sales-banking-and-capital-markets-en-south-africa

Microsoft 365 Copilot for Sales

https://learn.microsoft.com/en-us/microsoft-sales-copilot/introduction

CommBank demonstrates the real-world benefits of AI

https://news.microsoft.com/en-au/features/commbank-demonstrates-the-real-world-benefits-of-ai/

Microsoft Responsible AI Standard, v2

https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/Microsoft-Responsible-AI-Standard-General-Requirements.pdf?culture=en-us&country=us

2025 Responsible AI Transparency Report

https://www.microsoft.com/en-us/corporate-responsibility/responsible-ai-transparency-report/

Microsoft Responsible AI FAQs

https://www.microsoft.com/en-us/ai/responsible-ai#FAQ

ESMA - Using Artificial Intelligence for Investing: What you should consider

https://www.esma.europa.eu/sites/default/files/2025-03/ESMA_Warning_on_the_use_of_AI_-_EN.pdf

OpenAI

1455 3rd Street
San Francisco, CA 94158

1 October 2025

From:
Sandro Gianella
Head of Europe &
Middle East Policy &
Partnerships
OpenAI

To:
Dame Meg Hillier MP
Chair, Treasury
Select Committee
UK House of
Commons

Dear Dame Meg Hillier MP,

Thank you for the opportunity to respond to your 17 September letter about the use of artificial intelligence (AI) in the financial services sector.

OpenAI was created in 2015 with the mission to ensure that artificial general intelligence—AI that is at least as smart as a person—benefits all of humanity. We develop leading foundation models and make their capabilities available in safe and beneficial ways to people around the world. We make OpenAI technology available primarily through ChatGPT and our application programming interface (API).

Supporting economic growth in the UK

OpenAI's technology has become essential for millions of Brits. The UK is home to our first international office, and has become a top five market globally for paid subscribers and API developers and every day, people, developers, institutions, start-ups and leading British enterprises are unlocking economic opportunities with our technology.

In July, OpenAI announced a strategic partnership with the UK Government outlining our shared vision for building on the UK's strengths in science, innovation and talent, to maintain a world-leading UK AI ecosystem rooted in democratic values. We also announced Stargate UK, an infrastructure partnership designed to support the UK's sovereign compute capabilities. This partnership will enable OpenAI's models to run on local computing power for specialist use cases in regulated industries like finance.

We are committed to accelerating AI-driven economic growth and delivering opportunities to improve people's lives throughout the UK, including through our partnerships with financial services companies.

Partnering with financial services companies

We are proud to be able to support the UK's financial sector to deliver productivity improvements, innovative new solutions and features for workers and customers, and to unlock AI-driven growth in this essential sector.

We have worked with a range of financial services companies to integrate our AI tools. This includes automating tasks like drafting reports and generating documents, spreadsheets, presentations and other content; synthesising market data, earnings reports, and research to quickly generate insights; enhancing client services by creating personalised communications at scale; and integrating ChatGPT into a company's internal knowledge and productivity tools to deliver more relevant insights.

OpenAI

1455 3rd Street
San Francisco, CA 94158

In the UK, we collaborated with NatWest to support their strategic focus on bank-wide simplification, which includes deploying AI to meet customers' needs faster and more effectively and to increase productivity and efficiency across the bank. As part of this partnership, we are supporting the development of the bank's digital assistant services for customers, including new ways that customers can use AI to help them with more complex tasks like the identification, reporting and resolution of fraud and scams. NatWest's use of AI technologies is monitored and supported by their internal code of conduct, which is designed to ensure security, privacy, and transparency in data collection, processing and sharing.

We also work closely with financial services companies to ensure the AI tools meet the firm's strict standards for quality and reliability. For example, we collaborated with Morgan Stanley to implement an evaluation framework to test every AI use case before deployment, and measure how models perform against real-world use cases and guide improvements. These firms are best placed to understand and interpret sector-specific regulatory and compliance obligations. We are a complement to this, helping them understand use cases and how new AI tools can help them achieve their goals.

Our approach to safety and regulation

OpenAI is a leader in building and releasing models that are industry-leading on both capabilities and safety. We believe in a balanced, scientific approach where safety measures are integrated into the development process from the outset. This ensures that our AI systems are both innovative and reliable, and can deliver benefits to society. Our Preparedness Framework outlines our process for tracking and preparing for advanced AI capabilities.

We also work closely with the UK AI Security Institute and the US Center for AI Standards and Innovation. This partnership reflects our belief that frontier AI development must happen in close collaboration with allied governments that bring deep expertise in machine learning, national security, and metrology. We recently published examples of how this voluntary collaboration has built on our existing security approaches to produce real-world security improvements.

OpenAI is committed to working with UK policymakers to determine how AI can best serve the country. Straightforward, predictable rules that safeguard the public while helping innovators thrive can encourage investment, competition, and greater freedom for everyone.

We look forward to supporting this Committee's important efforts to ensure the safe and responsible use of AI in the financial services sector.

Sincerely,

Sandro Gianella