Bank of England

Dame Meg Hillier MP
Chair, Treasury Committee
House of Commons
Westminster
London
SW1A 0AA

Andrew Bailey
Governor

9 July 2026

Dear Chair,

I'm writing in response to your letter of 25th June 2026 regarding frontier AI and cybersecurity. I welcome your interest in this important topic. As you will see from the **Financial Policy Committee's Record** of discussion and the **Financial Stability Report** published earlier this week, this is an area of keen interest for the Bank and PRA.

Recent advancement in AI capabilities and their relevance to cybersecurity

You rightly note the cybersecurity risks posed by frontier AI. I will return to the potential impact we see on the financial services industry, and the specific questions you ask, but first, I want to set out our broader assessment of the recent developments in this space.

Recent advances in frontier AI models have shown marked capability gains, with progress in some areas exceeding experts' expectations. Whilst Anthropic's Mythos model was the first to demonstrate this level of cyber capability, it is no longer the only model of its kind. OpenAI's GPT 5.5 also demonstrated a step change in cyber capability[1]. The evidence points particularly to improvements in the length and complexity of software tasks these models can complete. This is relevant to cybersecurity because the ability to reason, write code, use tools, recover from errors

[1] See AISI's **April 2026 evaluation of OpenAI's GPT-5.5 cyber capabilities**

and carry out longer task sequences is closely linked to the ability to identify and exploit software vulnerabilities.

Frontier AI has now demonstrated, in testing environments, the ability to plan and execute cyber exploits autonomously. This means malicious actors could, in theory, use frontier AI to aid attempts to execute cyber attacks on institutions. For the financial sector, this matters because important business services depend heavily on complex digital infrastructure and common third-party technology, so a faster pace of vulnerability discovery and exploitation could have system-wide consequences, including through correlated exposures at shared suppliers, software components or critical infrastructures on which finance depends.

However, there are currently limiting factors. Success in these simulated environments does not, alone, show that frontier models could yet carry out reliable attacks against well-defended real-world targets. Other evidence suggests that the most advanced models still struggle to operate reliably, persistently and covertly in more demanding real-world conditions. They can be costly to use at scale, which could limit sustained use by lower-resourced malicious actors. Presently, access to such models, at least in their most powerful forms, is restricted to trusted partners. To date, these models would seem to have been put to primarily positive uses, namely, finding and remedying software vulnerabilities in a defensive capacity. Frontier models have been used by organisations to identify large numbers of weaknesses in existing software, demonstrating the potential power of frontier AI for cyber security[2].

However, we cannot assume that these limitations will persist indefinitely. As already noted, in several areas, frontier models have advanced faster than many experts expected, and have overcome constraints that previously appeared material. Further progress could narrow some of these remaining constraints quickly[3]. In terms of access, as models are more widely rolled out, restrictions on using models for cyber offensive activities are controlled by guardrails[4], which are known to not be universally reliable. Further, expert evidence suggests it will only be a matter of time before such capability proliferates, including via more widely available ‘open weight’ models[5] which have far weaker safeguards on their use.

[2] As of 22 May 2026, **Anthropic** claimed to have identified 23,019 candidate software vulnerabilities in open-source software using its Claude Mythos Preview model.

[3] This uncertainty is reinforced by emerging evidence that frontier AI models are already being used to accelerate parts of AI research and development, particularly coding and evaluation workflows.

[4] These are safeguards designed to detect and block harmful use.

[5] These are models of which the core parameters are made available for others to run and adapt.

Financial stability implications

Turning to the potential financial sector impacts, a key issue from here is the balance of advantage between malicious actors and defenders. We see this depending on three things: how quickly and comprehensively firms strengthen their processes to identify and fix vulnerabilities, how specifically the technology improves from here (as some capability gains will be more valuable for offensive vs defensive use cases); and how quickly malicious actors can access frontier models. There is uncertainty around these factors, and so the Financial Policy Committee, as part of its recent assessment of frontier AI's cyber risks, has used three indicative scenarios to understand how these dynamics might evolve and impact financial stability.

Across these scenarios, the level of cyber risk posed to the financial sector is variable, but across them all it remains the case that frontier AI may materially increase the flow of vulnerabilities requiring assessment and remediation by firms. Absent malicious use, this still could become a source of systemic risk, as a materially higher volume of identified vulnerabilities would require firms and suppliers to patch and validate changes at much greater speed and frequency, increasing the risk of errors, outages, and disruption across interconnected systems. The systemic significance of these developments will therefore depend not only on how material those vulnerabilities are, but also on how effectively firms and suppliers can absorb a persistently higher operational burden of triage, patching, testing, and recovery.

We see these developments as material. We emphasised in our recent Financial Stability Report that firms – including PRA-regulated firms and financial market infrastructures - should not treat frontier AI as a minor extension of the existing cyber threat, nor assume that the change in risk profile is static from here on. The issue is not that cyber risk is new, but that frontier AI may materially change its speed, scale and economics. Firms should also plan on the basis that a higher volume of vulnerabilities and a faster pace of patching may become a sustained feature of the threat environment and that this could itself create operational and resilience risks if change, testing and recovery processes are not able to keep pace safely.

The Bank's response

The Bank and PRA have been focussed on developments in frontier AI and cyber since the release of Anthropic's Mythos model and we have responded swiftly and comprehensively. We have drawn on a range of sources in forming this assessment, including supervisory engagement with regulated firms. Some of the information relevant to individual firms' preparedness, use of AI tools, and cyber resilience arrangements is firm-specific and supervisory in nature, and therefore cannot be disclosed publicly. I have therefore set out our assessment at an aggregate level.

Firm level resilience

The extent to which this situation increases risks to financial stability depends heavily on how quickly and thoroughly firms respond. Our approach to operational resilience focusses on managing the impact of disruption, meaning it is already predicated on the assumption of disruption. But frontier AI is likely to make cyber attacks faster, more scalable and more complex. Together, this raises the bar for what constitutes a "severe but plausible" scenario, meaning firms need to ensure that their controls, governance and recovery capabilities are sufficiently robust to operate under such severe and time compressed scenarios.

In May, together with the FCA and HM Treasury we issued a **joint statement**, which sets out practical steps that banks, insurers and Financial Market Infrastructure ("firms") should actively take now to prepare for and reduce frontier AI-related cyber risks within the existing operational resilience framework. This covers important areas of: governance and strategy, identification and risk management of vulnerabilities, managing risks from third parties, and protection, response and recovery. We have reinforced our existing expectations through supervisory and sector engagement; as we emphasise in the joint statement, it is essential that senior leadership of firms take ownership of cyber risk including the implications of frontier AI. We are engaging firms through the Cross Market Operational Resilience Group (CMORG), which the Bank jointly co-chairs with UK Finance, and which in May 2026 convened firms, authorities, the NCSC and AISI to discuss the sector's response to emerging frontier AI models and subsequently ran a **frontier AI risk mitigation webinar** for UK regulated financial services firms. CMORG also issued **guidance on frontier AI and cyber resilience** in June. In addition, the Bank has actively engaged with non-systemic firms on frontier AI risks. The Bank and PRA's upcoming consultation on Cyber risk management and resilience will consider issues relevant to frontier AI cyber risks.

Through our regular supervisory engagement and industry forums, we are assessing firms' preparedness to respond to frontier AI, including improving vulnerability management, testing their response to more severe cyber scenarios, and ensuring they can detect, respond to and recover from incidents at greater speed.

There remains restriction around the access to models at the frontier. Given the potential offensive capabilities of such models, some level of restricted access is an important component of managing the risks. However, advances in artificial intelligence pose global challenges, given the global financial system is interconnected with significant cross-border dependencies. Its risks, similarly, will not respect geographical borders. We see international co-operation as crucial to managing what is a very serious set of risks in this space. We are working with our colleagues internationally on these matters. More progress on this front will be important.

We have been engaging closely with our regulated firms, AI providers, and domestic authorities such as AISI and the NCSC, to understand developments and emerging risks related to frontier AI and cyber. As a regulator, we seek to ensure firms are able to build and maintain effective cyber resilience capabilities overall, including through a combination of tools. This includes leveraging the capabilities of latest accessible AI models where appropriate, for example to strengthen cyber defence, improve vulnerability identification, and support faster response to emerging threats. Many firms are already doing this.

The Bank has taken the same approach to responding to these risks in relation to its own estate. We are continually ensuring that our cyber defences are resilient to new developments and learning from the advanced capabilities of these new models. We are engaging directly with leading AI providers and working in collaboration with firms across the financial sector as well as international counterparts to develop our own approach.

It is also important to bear in mind that reliance on third party technology providers is a longstanding feature of the financial system. Activity to identify and respond to vulnerabilities within a firm's technology estate can only go so far, as cyber risks may also arise from third parties, impacting operations. This is why we expect firms to identify and appropriately manage the risks from material third party dependencies, which support their important business services. Firms will need to strengthen not just their cyber recovery capabilities but also their management of third party dependencies. This includes planning for scenarios in which third party infrastructures are compromised and ensuring they can restore service.

Future work on systemic resilience

Given the potential financial stability risks that we see, beyond individual firms, we are also placing priority on system-wide monitoring and preparedness. As we noted in our **letter** to you of 1st April 2026, the Bank is working to incorporate AI scenarios into different forms of cyber and operational testing for the financial sector, including the FPC's system-wide cyber stress tests. In this threat environment, we judge that this work is now likely to need to place greater weight on simultaneous multi-firm disruption, correlated disruption across technology supply chains, and to consider whether existing expectations, tools and coordination arrangements remain sufficient if frontier AI compresses the time available to identify, contain and recover from cyber incidents.

In addition to this, and given the scale of the challenge, we need to work together with industry to consider further options to bolster systemwide resilience and recovery. We are considering whether the present standard of recovery capability across regulated firms remains adequate, given the scope for more severe and fast-moving disruption, or

whether enhanced response and recovery options should be expected of certain systems supporting important business services that are critical to UK financial stability. For example, this could involve 'bare metal rebuild' in an isolated environment or a 'stand-in facility' to deliver a minimum level of service. We will also consider whether greater coordination and industry level co-operation, including on information sharing, cyber-protection infrastructures and coordinated response and recovery options for important business services, may be warranted. This also reinforces the importance of implementing the Critical Third Party regime, as the Committee emphasised in its recent AI inquiry and, on which, designations from HM Treasury are expected in the next few days. We must also be mindful of the resilience expected of other material third party technology providers that may not meet the threshold for designation as CTPs, particularly given some of these may provide important services to multiple sectors. As I noted above, it therefore remains key that firms strengthen their individual third party risk management capabilities in line with existing regulatory requirements and expectations.

Collaboration domestically and internationally

Deepening international coordination here is also critical. As noted above, the financial system is globally interconnected, and cyber disruption can spread across borders through common technology dependencies, financial institutions and market infrastructure. Differences in cyber capability, resilience and recovery capacity across jurisdictions could therefore have consequences beyond the jurisdiction in which an incident begins. We will need to continue working through international forums, including the FSB and G7, to build common understanding of how frontier AI is changing cyber risk and support consistent approaches to monitoring, testing and preparedness.

The G7 Cyber Experts Group which the Bank co-Chairs alongside the U.S Treasury, plans to accelerate significantly and expand its AI workplan over coming months. The Bank was also a key contributor to the FSB's draft Sound Practices for Responsible Adoption of Artificial Intelligence in the financial sector, which are out for consultation at the time of writing[6]. The FSB consultation proposes twelve sound practices to strengthen AI risk management in the financial sector, including a sound practice on cyber/ICT risk management that incorporates initial lessons learnt from recent developments in frontier AI. All of this work requires close coordination between authorities.

From a domestic perspective, you asked which part of the system is in the lead for financial services, and whether we have received support from Government

[6] **Sound Practices for Responsible Adoption of Artificial Intelligence (AI): Consultation report - Financial Stability Board**

colleagues. Cyber resilience in the financial system is inherently a shared responsibility between regulators, Government and national security authorities. Our work on cyber risks in the financial sector is undertaken actively and collaboratively between the Bank/PRA, HM Treasury[7], and FCA. These authorities continue to regularly engage on frontier AI developments and risks, and collectively agreed to issue the joint statement I referenced earlier. The Bank/PRA and FCA have also discussed their respective work on frontier AI to ensure co-ordination and with a view to ensure our approaches remain proportionate. We work closely with the National Cyber Security Centre, receiving the benefit of their threat intelligence, technical expertise and national-level coordination, which are essential to understanding and responding to the evolving cyber threat landscape. The Bank/PRA, HM Treasury and NCSC also participate in CMORG which has ensured a collective approach to supporting the financial sector's response to frontier AI. In particular, the NCSC actively supported development of CMORG's guidance to the financial sector on building resilience to the cyber risk from frontier AI. We are also fortunate in being able to draw on the guidance and expertise of the UK AI Security Institute, as part of the Department for Science Innovation and Technology.

We are mindful that the resilience of the financial system depends on the effectiveness of the preparedness agenda beyond the financial sector itself, such as critical sectors including energy and telecoms on which the financial sector depends. If frontier AI compresses the time between vulnerability discovery, exploitation and disruption across interconnected sectors, resilience in those sectors becomes more directly relevant to financial stability. Our engagement and collaboration with Government and other UK authorities is mindful of this, and we are keen to support a coordinated response across sectors. We are optimistic that the Government's Cyber Security and Resilience Bill, if effectively implemented, will help improve the resilience of critical sectors.

Conclusion

I welcome the Committee's engagement on this most important of matters, which I agree poses a significant risk. I hope my response assures you that the Bank is taking this very seriously. I would be happy to expand upon this matter further at our hearing at the Treasury Committee.

Yours sincerely,

[7] In the UK, the financial services sector is classed as a critical national infrastructure, and HM Treasury acts as the lead department responsible within Government for the sector's resilience.