Acting Comptroller of the Currency Michael J. Hsu
Remarks in Support of the
2024 Conference on Artificial Intelligence and Financial Stability
"AI Tools, Weapons, and Accountability: A Financial Stability Perspective"
June 6, 2024

Thank you for inviting me to speak at the 2024 Conference on Artificial Intelligence and Financial Stability.

I especially want to thank the Financial Stability Oversight Council (FSOC) for organizing this critically important conference. Under Chair Yellen's leadership, the Council has been reinvigorated and has developed stronger capacities to identify and address emerging risks to financial stability. This conference is a testament to that.

Like all technologies, artificial intelligence (AI) can be used as a tool or as a weapon. A lot depends on who is wielding it and for what purpose. Today I would like to discuss the systemic risk implications of AI in banking and finance through this tools-and-weapons lens. Both can create threats to financial stability, but in different ways, and each requires its own analysis.

These threats are exacerbated by the lack of a clear accountability model for AI. AI's ability to "learn" makes it powerful. With this power, however, comes greatly diffused accountability. With AI, it is easier to disclaim responsibility for bad outcomes than with any other technology in recent memory. The implications for trust are significant. Trust not only sits at the heart of banking, it is likely the limiting factor to AI adoption and use more generally.

Currently, the U.S. is the leader in AI innovation. To maintain this role, the U.S. needs to balance technological prowess with trusted use. Developing a robust accountability model can

help. As I will discuss in a bit, in the banking and finance arena, developing a "shared responsibility" model for fraud, scams, and ransomware attacks may provide a useful starting point.

Before getting to that, though, let me first describe the financial stability risks from AI through the lens of tools and weapons.

The Financial Stability Risks From AI's Use as a Tool

Banks, corporations, governments, and others are exploring AI use cases with the intention of using it as a tool. AI holds the promise of doing things better, faster, and more efficiently, yielding benefits for individuals, managers, organizations, and the public.

If the past is any guide, the micro- and macro-prudential risks from such uses will emanate from overly rapid adoption with insufficiently developed controls. What starts off as responsible innovation can quickly snowball into a hyper-competitive race to grow revenues and market share, with a "we'll deal with it later" attitude toward risk management and controls. In time, risks grow undetected or unaddressed until there is an eventual reckoning. We saw this with derivatives and financial engineering leading up to the 2008 financial crisis[1] and with crypto leading up to 2022's crypto winter.[2]

[1] See, e.g., Government Accountability Office, "Confirmation Backlogs Increased Dealers' Operational Risks, but Were Successfully Addressed after Joint Regulatory Action" (June 2007). The growth in trading volumes of over-the-counter derivatives between 2002 and 2005 largely outpaced large banks' and security dealers' processing capabilities. These firms' reliance on manual confirmation processes, as well as contractual provisions allowing unilateral assignments, led to a rapid backlog of trades that had not been formally confirmed with their counterparties, increasing the likelihood of errors going undetected. Industry participants cited, in part, a fear of losing business that prevented them from implementing fixes that could slow down transactions.

[2] Some of the highest-profile bankruptcies in the 2022 crypto winter, including that of FTX, involved shocking lapses of risk management and controls. See *The Wall Street Journal*, "Sam Bankman-Fried 'Wasn't Even Trying' to Manage Risk at FTX, He Says" ("Risk issues weren't seen as a 'core business driver' at FTX, Mr. Bankman-Fried said."); *New York Magazine*, "The Crypto Geniuses Who Vaporized a Trillion Dollars" (noting the "laid back"

How can we manage the risk of innovative tools crossing the line from being helpful to being dangerous? The history of derivatives and crypto suggests that it is extremely difficult to discern that in the moment. The competitive pressure on banks and others to keep up and not be left behind tends to overwhelm any objective sense of when growth needs to slow to allow controls to catch up.

Fortunately, basic risk management and common sense offer an answer: identify in advance the points at which pauses in growth and development are needed to ensure responsible innovation and build trust.[3] Well-designed gates can help strike the right balance between allowing innovation to flourish and having guardrails in place to prevent runaway growth.

The evolution of electronic trading provides a useful case study to consider. Traditionally, trading was manual. Market making eventually migrated to phones with computers providing real-time information, valuations, and forecasts for traders to use. In time, computers did more and more of the work, not just providing information, but also assisting and guiding traders' actions, supporting faster execution and more complex strategies. Eventually, algorithms would do it all—automatically buying and selling securities according to pre-determined instructions without the need for humans to execute trades.

attitude toward risk of the chief risk officer of Three Arrows Capital, the digital assets-focused hedge fund); *The Wall Street Journal*, "Crash of TerraUSD Shakes Crypto. 'There Was a Run on the Bank'" (detailing the risks inherent in the algorithmic stablecoin's business model, despite the founders' belief that they had created a stablecoin less prone to risks). The bankruptcies led to a loss of nearly $2 trillion dollars. See also Federal Reserve Bank of Philadelphia Consumer Finance Institute, "Cryptocurrency Ownership During a Crypto Winter: Effects of a Downturn on Consumer Attitudes to Crypto" (August 2023).

[3] OCC Bulletin 2017-43, "New, Modified, or Expanded Bank Products and Services: Risk Management Principles" (October 20, 2017).

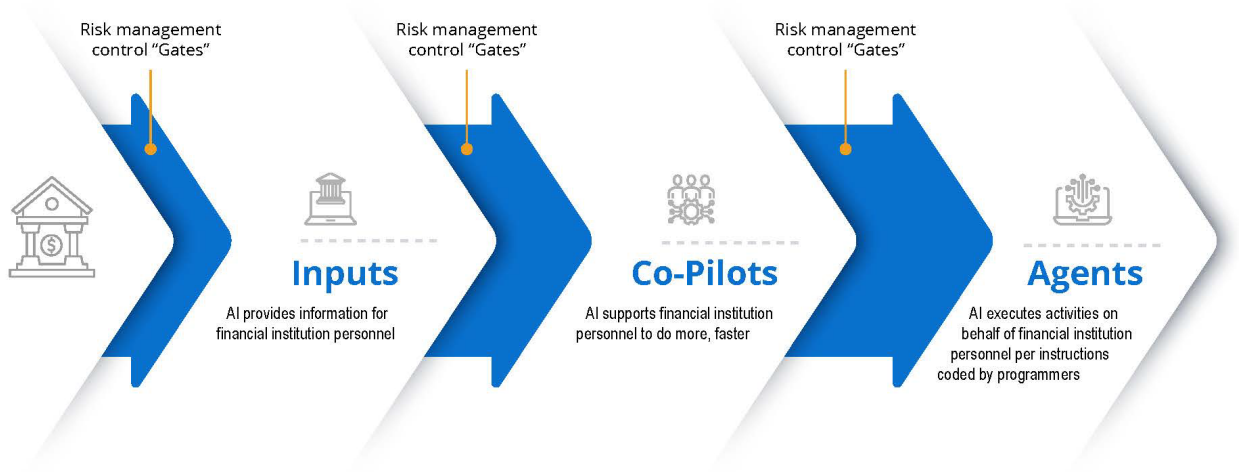
This evolution can be broken down into three phases: (1) *inputs*, where computers provide information for human traders to consider, (2) *co-pilots*, where computers support and enable traders to do more faster, and (3) *agents*, where computers themselves execute trades essentially on behalf of humans according to instructions coded by programmers. The risks and controls needed for each phase differ. For instance, mitigating the risk of flash crashes, which have been greatly enabled by algorithmic trading, requires a much more sophisticated set of controls than those needed when traders are simply provided with information on a screen.[4]

AI appears to be following a similar evolutionary path: where it is used at first to produce inputs to human decision-making, then as a co-pilot to enhance human actions, and finally as an agent executing decisions on its own on behalf of humans. The risks and negative consequences of weak controls increase steeply as one moves from AI as input to AI as co-pilot to AI as agent.

For banks interested in adopting AI, establishing clear and effective gates between each phase could help ensure that innovations are helpful and not dangerous. Before opening a gate and pursuing the next phase of development, banks should ensure that proper controls are in place and accountability is clearly established.

[4] *The Wall Street Journal*, "Dow Takes a Harrowing 1,010.14-Point Trip."

Risk Management Control Gates



The three phases noted here are conceptual; in practice, banks use a host of methods to ensure that new products and processes are rolled out in a safe and sound manner. Other factors may also feature prominently in banks' approaches and risk management, such as whether a new product or process is customer-facing or the degree to which it impacts a critical operation or service. We expect banks to use controls commensurate with "a bank's risk exposures, its business activities, and the complexity and extent of its model use."[5] Strong frameworks can help with this.

The Financial Stability Risks From AI's Use as a Weapon

Like any technology, AI can be used as a weapon just as easily as it can be used as a tool. In the wrong hands, AI can facilitate fraud, scams, cyberattacks, and operational disruptions. These threats require different responses than the tools-based risks noted above.

[5] OCC Bulletin 2011-12, "Supervisory Guidance on Model Risk Management" (April 4, 2011). See also *Comptroller's Handbook*, "Model Risk Management" (August 2021).

AI-enabled fraud is a top concern. As noted in a recent report issued by the Treasury Department, the ease with which AI tools can be accessed and used is rapidly lowering the barriers to entry for nefarious activities.[6] For instance, impersonating people's voices can now be done cheaply, easily, and at sufficiently high quality to fool not just family and friends, but also biometric systems.[7] Deepfakes abound[8] and have advanced since the early days of simple voice tricks[9] to more sophisticated and higher-dollar heists.[10]

To date, these types of incidents generally have been manageable from a financial impact standpoint. However, as criminals become more adept at using AI—see for example the launch of FraudGPT on the dark web—we should expect an increase in the scale and scope of fraud and scams. This could result in much larger financial impacts for banks and their customers.

More importantly, an increase in AI-powered fraud could sow the seeds of distrust more broadly in payments and banking.

In some ways, this is already happening. A recent survey of American consumers showed that many users, particularly younger consumers, welcome frictions in their digital financial products and services to ensure their digital identity is protected.[11] This outlook cuts against the broadly held notion that friction is bad for business. That sense applied in an environment where trust could be presumed to be high across all platforms and payment rails. With fraud on the

---

[6] U.S. Department of the Treasury, "Managing Artificial Intelligence-Specific Cybersecurity Risks in the Financial Sector" (March 2024).

[7] *The Wall Street Journal*, "I Cloned Myself With AI. She Fooled My Bank and My Family."

[8] *The Wall Street Journal*, "Deepfakes Are Coming for the Financial Sector."

[9] *The Wall Street Journal*, "Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case."

[10] CNN, "Finance Worker Pays Out $25 Million After Video Call With Deepfake 'Chief Financial Officer.'"

[11] Plaid, *The Fintech Effect: 2023 Consumer Trend Report.*

rise, however, consistent, high-trust platforms—rather than seamless user interfaces—are likely to win and retain customers in the long term.

AI-enabled cyberattacks are another threat vector warranting close attention. Criminals are using AI to generate code quickly to launch sophisticated cybercrimes. As a result, the frequency and scale of ransomware attacks are likely to increase, as are nation-state attempts to penetrate, disrupt, vandalize, or disable critical infrastructure.[12] The cascading risks from such attacks can be hard to foresee, but they warrant our full attention and highlight the importance of banks' operational resilience capabilities and investments.[13]

Finally, we need to prepare for an increase in AI-enabled disinformation. Last year a fake Bloomberg social media account posted an image of a bombing at the Pentagon. It went viral, spread in part by Russian government-sponsored media organizations, and caused a brief drop in the stock market before being confirmed as fake news.[14] The image was AI-generated.

The financial system's vulnerability to disinformation attacks seems to be increasing. Speed and decentralized information networks have been contributing factors. AI is likely to be an amplifier.

---

[12] Federal Bureau of Investigation, "Chinese Government Poses 'Broad and Unrelenting' Threat to U.S. Critical Infrastructure, FBI Director Says" (April 18, 2024); Microsoft Threat Intelligence, "Volt Typhoon Targets US Critical Infrastructure With Living-off-the-Land Techniques" (May 24, 2023); Microsoft Threat Intelligence, "Staying Ahead of Threat Actors in the Age of AI" (February 14, 2024) (detailing state-sponsored actors' current use of large language models in intelligence and criminal activities).

[13] Acting Comptroller of the Currency Michael J. Hsu, "Thoughts on Operational Resilience," Remarks at the Institute of International Bankers Annual Washington Conference (March 12, 2024).

[14] Associated Press, "FACT FOCUS: Fake Image of Pentagon Explosion Briefly Sends Jitters Through Stock Market."

The nightmare paperclip/Skynet scenario[15] for financial stability does not require big leaps of the imagination. Say an AI agent is programmed to maximize stock returns. It ingests the history of the stock market and identifies a pattern: the most severe stock market crashes are associated with bank runs. Bank runs are associated with high-profile bad news about a bank. Bad news about a bank can be easily spread via viral posts. The AI agent concludes that to maximize stock returns, it should take short positions in a set of banks and spread information to prompt runs and destabilize them.

Compared to the paperclip and Skynet scenarios, this financial scenario seems uncomfortably plausible given the state of today's markets and technology.

The Accountability Challenge with AI

The risks discussed thus far are complicated by the unique accountability challenge posed by AI. Let's start with a simple example.

When Jake Moffat's grandmother passed away in November last year, he went to Air Canada's website to get a flight to Toronto. He didn't know how to find the bereavement rate, so he asked the chatbot. The chatbot encouraged him to book the flight and ask for a refund retroactively within 90 days. Air Canada's policy, it turns out, prohibits retroactive refunds for bereavement flights. The chatbot was wrong.

[15] See Nick Bostrom, "Ethical Issues in Advanced Artificial Intelligence" (the famous hypothetical imagines a superintelligence tasked with manufacturing as many paperclips as possible, "with the consequence that it starts transforming first all of earth and then increasing portions of space into paperclip manufacturing facilities"). Skynet refers to the fictional AI network from the *Terminator* movie franchise that turns against humanity.

Later, when Moffat asked for a refund as per the chatbot's advice, Air Canada refused. Moffat sued. Air Canada argued that it "cannot be liable for the information provided by a chatbot."[16] The Canadian Civil Resolution Tribunal disagreed and ruled in Moffat's favor.

This outcome is intuitive and straightforward. What's more interesting is the revealed logic of Air Canada. The company believed that the chatbot was more akin to a "separate legal entity that is responsible for its own actions" than to one of its web pages or an employee.[17] This argument sounded ridiculous to the tribunal judge (as it does to most people). But consider the situation from a management perspective. With a faulty web page or incompetent employee, a company can identify who is at fault and then put in place controls to mitigate a repeat of that problem in the future. With a black box chatbot that is powered by third parties, most companies are likely to struggle to identify whom to hold accountable for what or how to fix it.

Consider another example: AI-powered credit underwriting. Last year a bank CEO noted to me that his team had been analyzing AI approaches to underwriting those who had been denied a credit card under his bank's standard underwriting techniques. The team determined that a significant portion of those denied could be safely extended credit using an AI algorithm. The problem was that the AI's decisions could not easily be explained.[18]

---

[16] *Moffatt v. Air Canada*, 2024 British Columbia Civil Resolution Tribunal 149, ¶ 27 (Can.).

[17] Ibid.

[18] As the CEO acknowledged, the explainability challenge raised consumer protection and compliance risk flags, specifically with regard to adverse action notices. For example, the Consumer Financial Protection Bureau recently issued a circular noting that the Equal Credit Opportunity Act requires companies to provide specific and accurate explanations for denying applications, regardless of whether the creditor is relying on opaque AI-driven credit models. CFPB, *Consumer Financial Protection Circular* 2023-03: "Adverse Action Notification Requirements and the Proper Use of the CFPB's Sample Forms Provided in Regulation B" (September 19, 2023).

For those who would have been denied by the AI algorithm, there is a question of fairness. *Why was I denied?* Data sets can be biased, algorithms can hallucinate, and reinforcement learning from human feedback can yield mistakes. How can one trust that the decisions reached by an AI algorithm are fair?

These consumer protection examples may seem far afield from systemic risk, but they illustrate two challenges with AI adoption. The first relates to the black box nature of AI and what that means for accountability and risk governance.

Second, and just as concerning, the immediate benefits of AI can quickly push accountability and governance questions to the background. Air Canada, for instance, was quick to deploy a chatbot even though it gave clearly wrong answers. And as the bank CEO noted to me, expanding credit access to those who are traditionally denied can be very compelling from both a business and a policy perspective. But can it—should it—compensate for the uncertain fairness that comes from an unexplainable model?

Sharing Responsibility

Accountability at its best aligns responsibility with capability. Put another way, when those on the hook for outcomes are most able to affect them, outcomes improve.

Today with AI, however, the companies most capable of affecting outcomes have limited responsibility for them. For instance, the ability of Air Canada to fix its chatbot pales in comparison to the ability of the AI platform, which runs the large language model upon which the chatbot was built. Yet Air Canada was the one held responsible for its chatbot misinforming Jake Moffat.

This is suboptimal and unsustainable. In theory, contracts and tort law could solve this problem. Companies could negotiate terms with their AI partners to ensure that the liability for bad outcomes was shared. Or companies could sue their AI partners under tort theories of strict liability, product liability, or negligence.[19] But such efforts are unlikely to be successful in changing the landscape more broadly. The history of networks, platforms, and utilities strongly suggests that private causes of action alone are unlikely to be effective in establishing safe, competitive, and fair outcomes.[20]

Fortunately, one does not need to look far for better approaches.

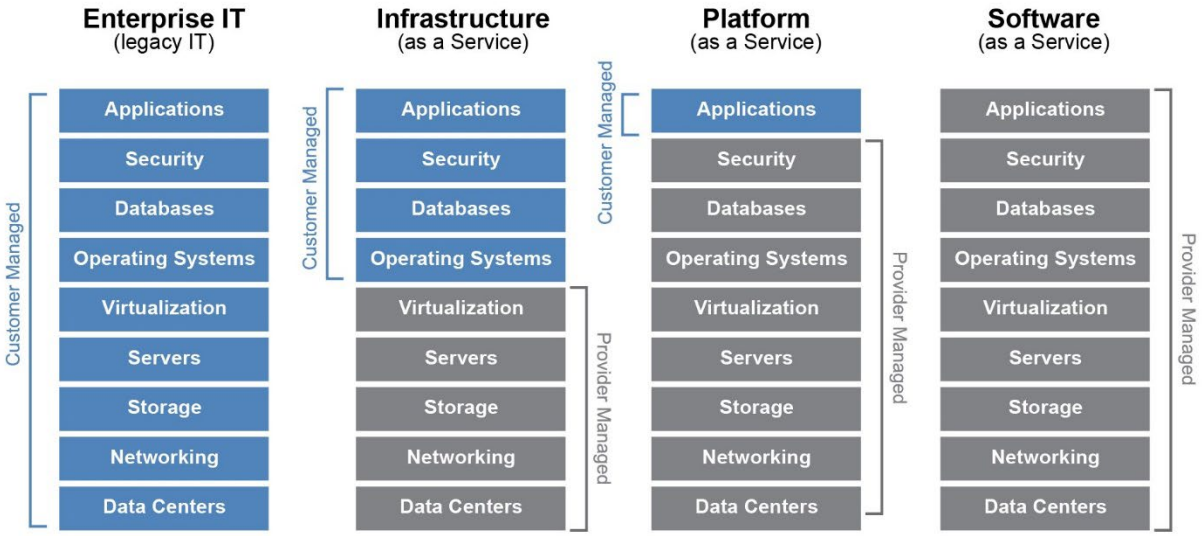
In the cloud computing context, the "shared responsibility model" allocates operations, maintenance, and security responsibilities to customers and cloud service providers depending on the service a customer selects.[21]

[19] E.g., Weil, Gabriel, "Tort Law as a Tool for Mitigating Catastrophic Risk from Artificial Intelligence" (January 13, 2024), available at SSRN: https://ssrn.com/abstract=4694006; Ryan Abbott, *The Reasonable Robot: Artificial Intelligence and the Law* (Cambridge University Press, 2020).

[20] See generally Morgan Ricks, et al., *Networks, Platforms, and Utilities: Law and Policy* (Faculty Books, 2022).

[21] U.S. Department of the Treasury, "The Financial Services Sector's Adoption of Cloud Services," p.22 (February 8, 2023).

Shared Responsibilities Model



Source: General Services Administration, Cloud Information Center, available at https://cic.gsa.gov/basics/cloud-security.

A similar framework could be developed for AI. The high-level components of the "AI stack" are fairly intuitive—that is, there is an infrastructure layer, a model layer, and an application layer. But for the framework to be actionable, consensus on the sub-components within each layer and on the types of third-party arrangements would be needed.

The recently established U.S. Artificial Intelligence Safety Institute (AISI), which is situated within the National Institute of Standards and Technology (NIST), may be well positioned to take on this task. The Institute could leverage its AI Safety Institute Consortium which consists of over 280 stakeholder organizations, from the largest AI platforms to academic AI safety research teams. Notably, a consortium model was used in the 1980s to develop the internet protocols that we take for granted today.[22]

[22] The Internet Engineering Task Force (IETF), founded in 1986, developed the technical standards for the internet protocol suite (TCP/IP), the set of communication protocols used for the internet. Originating as a quarterly meeting of U.S. government-funded researchers, the IETF expanded to its current open, non-membership model. Working groups develop "technical documents . . . that define how internet technology works in detail, and can be operated and managed at scale." Larger group meetings help guide the work of these groups. IETF, "Introduction to the IETF" (last accessed May 29, 2024).

Assuming for a moment that a shared responsibility framework for AI safety could be developed, the natural question is how it would be enforced. As noted earlier, I am skeptical that contracts and torts alone can be effective in the long term. Other models warrant consideration, for example, self-regulatory organizations (such as the Financial Industry Regulatory Authority), network membership organizations (like Nacha or the Clearing House), and split reimbursement liability (as the United Kingdom does for authorized push payment fraud).

The FSOC is uniquely positioned to contribute to this, given its role and ability to coordinate among agencies, organize research, seek industry feedback, and make recommendations to Congress.

Conclusion

The real power of AI stems from its ability to learn. With this learning, however, comes novel challenges for accountability and governance.

From a financial stability perspective, AI holds promise and peril from its use as a tool and as a weapon. The controls and defenses needed to mitigate those risks vary depending on how AI is being used.

At a high level, though, I believe having clear gates and a shared responsibility model for AI safety can help. Agencies like the OCC and bodies like the FSOC and the U.S. AI Safety Institute can play a positive role in facilitating the discussions and engagement needed to build trust in order to maintain U.S. leadership on AI.