Basel Committee
on Banking Supervision

Managing AI in banking: are we ready to cooperate?

Speech by Pablo Hernández de Cos
Chair of the Basel Committee on Banking Supervision and Governor of the Bank of Spain

Keynote speech at the Institute of International Finance Global Outlook Forum
Washington DC, 17 April 2024

Introduction

Good morning, and thank you for inviting me to speak at this conference today.

In 1986, the historian Melvin Kranzberg published his six laws of technology.[1] At the top of his list was his view that "technology is neither good nor bad; nor is it neutral".[2] Fast forward to today, and this edict could seamlessly apply to the debate about the use of artificial intelligence (AI) and machine learning (ML) in banking.

Discussions about the promises and pitfalls of the use of generative AI and large language models in banking are becoming increasingly common. That talk is also being turned to action, with banks starting to use and invest in AI/ML. More boldly, some are already talking about the potential impact of the yet-to-be-established artificial general intelligence on banking.[3]

There is now an emerging narrative that lauds the purported benefits of AI in banking – in terms of operational efficiencies and improved risk management – while also cautioning about challenges, ranging from data privacy to model hallucinations to reputational risk.

Yet we're still left with an unanswered question: is the use of AI/ML in banking a net positive or negative to global financial stability, and perhaps society more generally? Have we thought through all of the potential scenarios that could play out in a world where AI/ML plays a prominent role in banking? Are we at risk of our own "consensual hallucinations" about AI/ML if we fail to take a step back and consider the bigger picture?

I shall not attempt to provide a definitive answer to these questions today. Instead, I will follow Kranzberg's framing and try to bring together both the micro- and macroprudential and financial stability considerations.

My main message is that the use of AI in banking raises important prudential and financial stability challenges. We've yet to see how AI/ML performs across a full financial cycle – and this

[1] Kranzberg (1986).

[2] Ibid.

[3] See, for example, Silva (2024).

could be some time off. Left unchecked, such models could potentially amplify future banking crises. But these challenges and limitations are not insurmountable, provided that central banks and supervisory authorities adjust to this new reality and collaborate effectively.

AI/ML in banking is neither bad...

Let me start with the potential benefits of AI/ML in banking.

In principle, AI/ML could help increase banks' operational efficiency, risk management capabilities and product offering. More specifically, AI/ML applications could include:

- superior pattern recognition ability and predictive power compared with more traditional approaches (eg in improving investment performance or expanding credit access);
- cost efficiencies (ie AI/ML approaches may be able to arrive at outcomes more cheaply, with no reduction in performance) such as enabling the development of multi-channel customer access and increasing self-service by customers;
- greater accuracy and consistency in processing compared with approaches that have more human input and higher "operator error" (such as detection of anomalies in anti-money laundering monitoring);
- improving the client experience, as the technology can help streamline customer interactions (such as applying for a loan) by removing manual steps; and
- better capability to accommodate very large and less structured data sets, and to process those data more efficiently and effectively (eg the ability to gain greater insight into customer needs and the provision of more tailored or customised services).

In practice, we are already seeing the emergence of many real-world use cases of these applications. One notable area is the use of AI/ML in "scientific" applications, such as the use of graph neural networks in detecting money laundering. An example is Project Aurora, developed by the BIS Innovation Hub, which uses AI to enhance money laundering transaction monitoring by identifying anomalies in transaction data that traditional methods cannot detect.[4] In that context, AI can help reduce risks to banks and the banking system.

We are also seeing some banks testing the use of natural language processing or machine learning methods for digital assistants / robo-advisers, summarising documents, code generation and testing, fraud and money laundering detection, and algorithmic trading strategies.

In a similar vein, supervisory authorities have also been exploring the use of AI/ML and suptech in particular to help fulfil their mandates. To name just a few (global) examples, the BIS

[4] BIS Innovation Hub (2023).

Innovation Hub has developed various proofs of concept that rely on AI/ML to enhance regulatory reporting and data analytics (Project Ellipse), and to facilitate climate risk analysis (Project Gaia).[5]

So there are no shortages of examples of how AI/ML could transform the banking system landscape. All the more reason to make sure that we carefully assess its associated limitations and risks.

…nor good…

In some respects, AI/ML models in banking are essentially more sophisticated and powerful versions of existing modelling techniques. But their potential scale and scope raise important prudential risks and limitations. Let me mention three in particular.

Let's start with the inputs used in such models. Higher volumes of data, at higher velocities and with different varieties, may present greater data governance challenges in ensuring data quality, relevance, security and confidentiality. Users may not know the sources used to produce output or how such sources were weighted, and a financial institution may not have full understanding of or control over the data set being used. There are fundamental questions on data governance and ethical questions that we need to answer before relying on such models extensively.

There is also a question about the scope for further improving AI/ML models. To date, such models have relied on bigger and bigger data sets to enhance their accuracy. Yet how much unused data is left to feed future generations of these models? One estimate suggests that the stock of high-quality language data available could be exhausted as early as next year.[6] There is therefore likely to be a greater need to rely on other techniques to enhance these models in the future and enhance their usefulness in banking.

Second, as with traditional models, AI/ML models can reflect biases and inaccuracies in the data on which they are trained. Hallucinations – a term which in my view unhelpfully personifies models and downplays the gravity of inaccuracies – produce erroneous and flawed outputs, at times in a convincing manner. To be clear, humans are also prone to bias in their decision-making. In many instances, these models cannot be "explained" – put differently, we don't always understand why and how a model generated a specific output. This lack of explainability adds an extra layer of questions about models relative to existing ones (where the focus may lie mostly on statistical errors and residuals). So banks and supervisors must be sufficiently comfortable and confident about the robustness of their potential use in critical banking services.

What should our risk tolerance level be for such models? Are 10,000 "correct" answers by a model worth the cost of a single hallucination, especially for systemically important banking activities? Do we currently have sufficiently robust mitigants to address the risks associated with

---

[5] BIS Innovation Hub (2022, 2024).

[6] Villalobos et al (2022) and *The Economist* (2023).

incorrect, biased and/or unethical outputs produced by AI/ML models? Does it matter if we don't know why we get such results? These are important questions that banks and supervisors should be contemplating.

Third, the current AI/ML landscape can heighten third-party dependency by banks and exacerbate concentration risks. To date, most foundation models are developed by a small number of providers. This leads to convergence and reliance by banks when it comes to their use of "off-the-shelf" or adapted models developed by such providers. AI/ML deployment also often involves the use of cloud technologies, further entrenching this network of interconnections, which can test banks' operational resilience. The network-type effect and economies of scale could further increase the role of these service providers and further accentuate the dependency risks for banks. In my view, this calls for a simple principle when it comes to overseeing such third-party service providers: they should be required to have the same level and robustness of governance, risk management and operational resilience practices as those expected from banks.

…nor neutral

But beyond these mostly micro-level considerations, there is perhaps an even more fundamental financial stability challenge when it comes to the use of AI/ML in banking. Are AI/ML models "neutral" by nature, or do they by design magnify and amplify human behaviour?

What makes AI/ML models stand out from the existing generation of models is their:

- automaticity, namely their ability to operate with no/little human involvement;
- speed, as a result of their sheer processing power and reliance on big data; and
- potential ubiquity resulting from a widespread, cross-sectoral adoption and integration (ie a "network effect").

To borrow from the electronic music duo Daft Punk, AI/ML models may be harder, faster and stronger than existing ones, but are they necessarily better for financial stability?[7]

When viewed from a system-wide perspective, banks' reliance on AI/ML models can exacerbate the self-reinforcing features of the financial cycle, including most notably the procyclical evolution and amplification of risks at an aggregate level.[8] This is because AI/ML models may not be robust to endogenous risk over time.[9] These models rely heavily on historical patterns and fail to capture human behavioural dynamics in times of stress.

We learnt that painful lesson during the Great Financial Crisis with the existing generation of bank internal models. The ability of internal models to robustly see "through the cycle" was

---

[7] Daft Punk (2001).

[8] Hernández de Cos (2023).

[9] See, for example, Danielsson and Uthemann (2024), Danielsson and Shin (2002) and Danielsson et al (2012).

clearly put into question, as these models imploded during stress events – recall the incredible observation by a chief financial officer of a global systemically important bank in August 2007 that they were "seeing things that were 25-standard deviation moves, several days in a row".[10] Indeed, some of the Basel III reforms are aimed at reducing the role of such models and enhancing their robustness.[11]

Nothing suggests at this stage that AI/ML models will fare any better. If anything, their automaticity, speed and ubiquity could potentially aggravate future financial stress episodes by exacerbating amplification mechanisms such as fire sales, runs and deleveraging. The number of parameters in notable AI systems is well north of 10 billion, with some now exceeding 1 trillion. The risk of model misspecification is therefore incredibly high.[12]

There is also an indirect structural risk to banks from AI/ML models. Recent studies have suggested that AI/ML has the potential to reshape labour markets.[13] Under some scenarios, income inequality and unemployment may increase for some labour segments or industries. This could result in a negative second-round effect – potentially in a short period of time – for banks' exposures to such households or corporates. Banks must therefore be vigilant to such potential scenarios as part of their risk management.

We also know that existing large language models face important cognitive limits (eg Perez-Cruz and Shin (2024)). Today's generation of models cannot understand the rules of games like tic tac toe or basketball, even if they can beat humans in games such as chess or Go. How can we sensibly rely on them to understand the complex game of finance, which is constantly evolving? Put differently, while AI/ML models may be good at finding a needle in an existing haystack, they are unlikely to perform as well when the haystack itself evolves over time.

This is why human judgment must remain a core part of banking, with banks' own risk management and governance arrangements at the heart of it.[14] The "A" in AI will always remain artificial, and cannot replace the critical role of humans, be it through the relationship side of banking or the ability to challenge and override model-based outputs. Banking without banks, bankers or supervisors would be a brave new world.

In fact, this observation was not lost on Kranzberg. Another of his laws of technology states that "although technology might be a prime element in many public issues, nontechnical factors take precedence in technology-policy decisions".[15] We cannot lose sight of these bigger picture dynamics when discussing the role of AI/ML in banking.

---

[10] Thal Larsen (2007).

[11] BCBS (2017).

[12] Giattino et al (2023).

[1313] For example, Cazzaniga et al (2024).

[14] Hernández de Cos (2022a, 2022b).

[15] Kranzberg (1986).

Conclusion

Digital innovation will further fuel cross-border and cross-sectoral financial interconnections. Safeguarding global financial stability will therefore require ongoing collaboration to ensure that we achieve an appropriate baseline regulatory and supervisory approach to overseeing the use of AI/ML in banking and beyond. This collaboration is likely to extend across a wide range of authorities – going beyond just central banks and bank supervisors – given the ongoing growth of non-bank players and the blurring of regulatory boundaries. Some have even called for a "technoprudential" approach to AI governance, building on the macroprudential framework to overseeing system-wide financial risks.[16]

When it comes to banking, it is critical that banks anticipate and oversee the risks and challenges posed by AI/ML – both at the micro and the macro level – and incorporate them in their day-to-day risk management and governance arrangements. And it is critical that human judgment plays a central role as part of this framework.

Central banks and supervisory authorities must also anticipate and mitigate the potential risks and vulnerabilities to the banking system. For its part, the Committee has already issued supervisory newsletters on AI/ML and the use of third- and fourth-party service providers.[17] In addition, it will soon be publishing a more comprehensive report on the digitalisation of finance and the implications for regulation and supervision.

Let me end by going back to Kranzberg's final law of technology: "technology is a very human activity".[18] Despite attempts to paint a robotic future for banking, the fact is that humans – and in particular human judgment – can and must continue to play an integral role.

[16] Bremmer and Suleyman (2023).

[17] BCBS (2022a, 2022b).

[18] Kranzberg (1986). This law continues "…and so is the history of technology".

References

Basel Committee on Banking Supervision (BCBS) (2017): *High-level summary of Basel III reforms*, December.

——— (2022a): "Newsletter on artificial intelligence and machine learning", March.

——— (2022b): "Newsletter on third- and fourth-party risk management and concentration risk", March.

BIS Innovation Hub (2022): *Project Ellipse: an integrated regulatory data and analytics platform*, March.

——— (2023): *Project Aurora: the power of data, technology and collaboration to combat money laundering across institutions and borders*, May.

——— (2024): *Project Gaia: enabling climate risk analysis*, March.

Bremmer, I and M Suleyman (2023): "The AI power paradox", *Foreign Affairs*, August.

Cazzaniga, M, F Jaumotte, L Li, G Melina, A Panton, C Pizzinelli, E Rockall and M Tavares (2024): "Gen-AI: artifical intelligence and the future of work", *IMF Staff Discussion Note,* SDN/2024/001, January.

Daft Punk (2001): "Harder, better, faster, stronger", *Discovery*, October.

Danielsson, J and H S Shin (2002): "Endogenous risk", in P Field (ed), *Modern Risk Management*, Risk Books.

Danielsson, J, H S Shin and J-P Zigrand (2012): "Procyclical leverage and endogenous risk", working paper, October.

Danielsson, J and A Uthemann (2024): "On the use of artificial intelligence in financial regulations and the impact on financial stability", February.

Giattino, C, E Mathieu, V Samborska and M Roser (2023): "Data page: parameters in notable artificial intelligence systems", *Our World in Data*.

Hernández de Cos, P (2022a): "Trust, digitalisation and banking: from my word is my bond to my code is my bond?", speech at the Eurofi 2022 Financial Forum, Prague, 9 September.

——— (2022b): "International cooperation in a world of digitalisation", speech at the 22nd International Conference of Banking Supervisors (virtual), 29 November.

——— (2023): "Back to the (macroprudential) future: Reflections and questions on macroprudential policy", speech at the HKMA-BIS joint conference "Future-proof supervision for an innovative banking world", Hong Kong SAR, 24 March.

Kranzberg, M (1986): “Technology and history: Kranzberg’s Laws”, *Technology and Culture*, vol 27, no 3, July.

Perez-Cruz, F and H S Shin (2024): “Testing the cognitive limits of large language models”, *BIS Bulletin*, no 83, January.

Silva, L (2024): “How artificial general intelligence will drive an inclusive financial sector in Latin America”, *World Economic Forum Annual Meeting*, January.

Thal Larsen, P (2007): “Goldman pays the price of being big”, *Financial Times*, 13 August.

The Economist (2023): “AI is setting off a great scramble for data”, 13 August.

Villalobos, P, J Sevilla, L Heim, T Besiroglu, M Hobbhahn and A Ho (2022): “Will we run out of data? An analysis of the limits of scaling datasets in machine learning”, *arXiv preprint*, 2306.155456.