AI and Financial Stability

Jón Daníelsson
London School of Economics

modelsandrisk.org/AI

10 April 2025

International Conference on Generative Artificial Intelligence
Hong Kong Monetary Authority

Bibliography

- Joint work with Andreas Uthemann, Bank of Canada! `authe.github.io`
- My AI work `modelsandrisk.org/AI`

Some recent AI cases

a. An AI was instructed to fully comply with all securities laws and maximise profits. When given private information, it proceeded to do insider trading and lie about it to its human overseers

b. Bank of Canada assessed impact of 25% US tariffs — GDP $\downarrow 6\%$
Deepseek took 12 seconds to make its own impact model, finding GDP $\downarrow 4\% \in [1\%, 8\%]$

c. Google Research's AI "co-scientist" cracks superbug problem in two days Imperial College London scientists had worked on for years

- Prof Penadés said the tool had, in fact, done more than successfully replicating his research
- "It's not just that the top hypothesis they provide was the right one," he said.
- "It's that they provide another four, and all of them made sense"
- "And for one of them, we never thought about it, and we're now working on that"

At the onset of every crisis

- Some shock happens
- Banks decide to run or stay — stabilise or destabilise
- If shock is not too serious, optimal to absorb and even trade against shock
- If avoiding bankruptcy demands a swift, decisive action, such as selling into a falling market, do exactly that

	Run	Stay
Crisis	✓Right decision	✗Wrong decision
No Crisis	✗Wrong decision	✓Right decision

Important to anticipate what the majority will do and do that first

AI generally improves the financial system

- Makes financial intermediation cheaper, more robust and efficient
- Could increase competition (caveat discussed later)
- Makes regulators and central bank financial stability experts more productive
- Allows regulators to directly benchmark regulations and interventions
- Helps in optimising policy responses in crisis
- Allows the financial authorities to allocate resources better

But there is a tradeoff...

AI can also undermine stability

A. Malicious and damaging use of AI (by AI and humans)

B. Wrong way risk — Model risk

C. Synchronised behaviour, monoculture, procyclicality

D. Speeding up and intensifying crises

A.1 Malicious use of AI by humans

- We have always exploited technology for malicious purposes
  - Find legal or regulatory loopholes
  - Crime
  - Terrorism
  - Nation-state attacks
- These people will not follow ethical guidelines or regulations or AI acts
- The job of malicious agents becomes easier the better AI becomes

A.2 The defender's dilemma

The problem has always been

- Attackers need one vulnerability, while defenders must monitor everywhere
- Monitoring gaps in the space between regulation silos
- Innovation outpaces regulatory frameworks
- Defense requires vastly more resources than attack

It gets worse with AI because the asymmetry in computing power grows with system complexity and AI ability

A.3 Model risk

AI doing what it thinks it is supposed to do

- Insider trading example at the start
- Model risk especially arises when
  1. AI faces a complex problem
  2. Needs to find multidimensional solutions
  3. Is given simple rules
- We cannot evaluate LLMs like we evaluate other models by checking their code and making sure that for a given input, their outputs are correct/safe.
- AI admiral sinking its own ships

Telling AI to keep the system safe (public sector) and avoid bankruptcy (private sector) is too high-level to be useful

A.4 Human experts in charge

- Principal-agent problems — AI does not care about punishment or bonuses
- Regular ways to incentivise — carrots and sticks — don't work with AI
- So have a human expert in charge
  1. Human in the loop (human making decision)
  2. Human on the loop (human supervising system)
- But then we have the speed issue discussed below

B.1 Trust and how we come to over-rely on AI

- AI builds up trust by being good at simple tasks that play to its strength
- We may end up with the AI version of the Peter principle
- Usually not credible when someone says
  - "We would never use AI for X (in or on the loop)."
- So long as it delivers significant cost and efficiency savings
- Competitive pressures drive AI adoption (see slide on market structure below)

Human out of the loop

B.2 Wrong way risk

- Want highest explainability/least hallucination for most vital problems
- Extreme financial system outcomes unique
- AI knows little about the most important causal relationships
- The reaction functions (public and private) sectors are mostly unknown
- Such a problem is the opposite of what AI is good for
- AI knowledge is negatively correlated with the importance of decisions

When AI is needed the most, it knows the least — wrong way risk

C.1 The problem of common knowledge and synchronised behaviour

- Synchronised behaviour
  1. The more similar our understanding of the world is
  2. It benefits us to coordinate
- Creates problems, such as
  1. Speculative attacks
  2. Liquidity crises
  3. Selling spirals

The more similar the neural network is, (next slide) the taking of risk becomes increasingly correlated (procyclical)

C.2 Neural networks

- Different institutions with heterogeneous objectives will use neural networks to inform decision-making
  1. Mathematical design
  2. Training and optimising — data
- Are the neural networks more similar or more heterogeneous than the human centred structure that came before it?
- Likely that the answer is more similar
  1. For broad categories of important data single data vendor may have an effective monopoly
  2. Small number of open and closed source engines

C.3 Seeking alpha and neural networks

- Many firms will want their own unique neural network
- But not many can afford to
  1. Acquire the necessary human capital to design their own networks
  2. Have the compute to optimise them
  3. Obtain the necessary unique data
- Low hanging fruits disappear, so heterogeneity becomes increasingly valuable, giving advantage to institutions best able to master it

C.4 Market structures in competitive markets

- The largest banks — GSIBs — find it easier to fund their own neural networks
- And easier to impose AI on its staff
- Middle tier banks, with traditional staff, legacy systems and technical dept may have to use commodity engines
- The smallest neo institutions with nimble technology stacks and staff may find creative uses for commodity engines
- GSIBs $\uparrow$, middle tier $\downarrow$, neo banks $\uparrow$ $\quad \Rightarrow$ Market concentration

AI can further entrench GSIBs and hence increase systemic risk

D.1 Debt, liquidity and crises

- A systemic financial crisis is characterised by the disappearance of liquidity
- HFT, flash crashes, market function, ..., are important but not systemic
- Liquidity and safe assets are particularly valuable in crisis
  1. Deposits
     - 1.1 The case of SVB
     - 1.2 Deposit aggregators
  2. Liquidity supplied to the markets and real economy
     - 2.1 ETFs rapidly withdraw liquidity
     - 2.2 Liquidity providers, banks and others, stop providing liquidity
     - 2.3 Banks prefer central bank reserves
     - 2.4 Cancel standing orders, refuse loans, ...

D.2 Be the first to act in crises

	Run	Stay
Crisis	✓Right decision	✗Wrong decision
No Crisis	✗Wrong decision	✓Right decision

AI speeds up and intensifies crisis

D.3 AI train other AI for good and bad

- Training data for AI engines is fed by what other AI do
- Private AI can make "fake news" to manipulate markets
- AI output fed into other AI learning (manipulate, garbage,...)
- AI optimise to influence each other
- And simultaneously cooperate and compete

These AI-to-AI channels are likely hidden until it is too late

D.4 Speed and viciousness

- AI particularly good in rapidly processing new information and reacting quickly
- AI good at coordinating when mutually beneficial — global behaviour
- And undermining/attacking one not the case

D.5 AI can stabilise markets

- If AI thinks a shock is not serious, it is optimal not to panic sell — stay — buy risky assets
- Then AI is a force for stability
- By absorbing shocks

D.6 If AI concludes there will be a crisis — Speed and viciousness

1. AI good at identifying structural vulnerabilities/weaknesses (fundamental uncertainty)
2. And exploiting the weaknesses by preempting/coordinating (strategic uncertainty)

- Speed is of the essence
- The first to react gets the best prices
- The last to act faces bankruptcy
- Sell, calls in loans, run others as quickly as possible
- Makes the crisis worse in a vicious cycle

Days or weeks are reduced to minutes or hours

Implications for micro and macro

Micro

- Traditional regulation based on PDF files, database dumps, conversations, inspections
- When the private sector uses AI to comply, e.g. in reporting
- AI generates reported information (like pdf and data dumps)
- Easy to optimise against the authority
- Increasing asymmetry and undermining the effectiveness of regulations
- How are Basel methodologies like the LCR and its runoff assumptions affected by AI crisis speed?

Macro

- AI lowers volatility and fattens the tails
- Could the central bank systemic risk dashboards unambiguously conclude AI is stabilising as they predominantly focus on non-extreme risks?
- Or, more generally, are our existing systemic risk analytical frameworks likely to capture risks from AI?
- The traditional way of preventing stress and responding to crises may not work

Policy options

Who ‘takes a lead’ on AI

- The core function of central banks is monetary policy and financial stability
- And since AI can threaten financial stability
- The division taking a lead on AI should be financial stability
- Not data, IT or innovation

Put AI at the core of the financial stability function

Critical dilemma for authorities

- Either use open source models, local but less capable
- Or closed source models that are a black box and less secure
- Need for specialised financial AI with appropriate data protections
- Can the authority train its models with regulatory data? (see federated learning below)
- Hard to control staff that wants to use much better closed-source engines

Authorities may have to sacrifice
security if they want to effectively harness AI

Public-private partnerships

- Unlikely the authority will have the necessary AI experience or technology
- Set up strategic collaboration between financial authorities and AI providers
- Especially local and not international vendors
- Maybe replicate ‘market intelligence’ frameworks in the AI space

AI-to-AI links and benchmarking

- Authority AI directly communicates with private AI via API
- Can ask how it might react in particular cases
- Can examine iterative processes
- Investigate aggregate behaviour across the industry
- Identify industry-wide feedback
- Not based on much data sharing
- The technology for this exists today
- LLMs already useful in allowing single access to diverse APIs

Triggered facilities

- The bank AI might already react before the Governor has the chance to call the bank CEO
- Current liquidity facilities are mostly based on discretion, and committee meetings might be too slow
- Expand pre-committed (triggered) liquidity facilities that activate automatically
  - Reduces uncertainty during market stress
  - Prevents destructive AI-driven fire sales
  - Creates predictable stabilisation mechanisms
  - Counters the speed advantage of AI systems

Learning and confidential data

- Coordinate with other authorities on setting up neural networks
  - Share resources
  - Capture common vulnerabilities
  - Capture silo boundary vulnerabilities
- Use federated learning — train models across organisations by sharing weights but not data

7. Reporting/dashboarding on AI

- Track AI use on operational levels — training, engine source, data
- Coordination risks and emerging synchronisation
  - If the risk-taking operations across banks use similar engines $\Rightarrow$ procyclicality $\uparrow$ systemic risk $\uparrow$
  - If liquidity management across banks use similar engines $\Rightarrow$ potential for destructive synchronous behaviour $\uparrow \Rightarrow$ speed and viciousness of crises $\uparrow \Rightarrow$ systemic risk $\uparrow$
- Market structure
  - If only certain financial institutions develop their own engines when most cannot $\Rightarrow$ concentration $\uparrow \Rightarrow$ systemic risk $\uparrow$

Conclusion

- AI broadly positive for the users of the financial systemic
- Lower cost and better tailored services
- It will significantly help the macro and micro authorities
- Also raises new, poorly understood, systemic risks
- Likelihood of crisis is negatively correlated with the level of the authorities' understanding and use of AI

If the authorities don't effectively engage with AI
crises more likely