Keynote speech

Artificial intelligence and systemic risk

Jón Daníelsson
London School of Economics
modelsandrisk.org

16 November 2023

Seventh annual conference of the European Systemic Risk Board
Financial stability challenges ahead: Emerging risks and regulation
www.esrb.europa.eu/news/schedule/2023/html/20231116_7th_annual_conference.en.html

Bibliography

- Joint work with Andreas Uthemann, Bank of Canada
  authe.github.io

- My AI work
  modelsandrisk.org/appendix/AI

  a. "On the use of artificial intelligence in financial regulations and the impact on financial stability"
     papers.ssrn.com/sol3/papers.cfm?abstract_id=4604628
  b. "Artificial intelligence and financial stability"
     cepr.org/voxeu/columns/artificial-intelligence-and-financial-stability
  c. "When artificial intelligence becomes a central banker"
     cepr.org/voxeu/columns/when-artificial-intelligence-becomes-central-banker

What artificial intelligence (AI) am I talking about?

- There are many AIs — Data driven ML with reinforcement learning to achieve objectives
- Not the "singularity"
- Computer algorithm that makes decisions that humans would normally do
- Searches for best outcome given its objectives and understanding of the world
  1. Advising human decision makers
  2. Making independent decisions
- Uses data (like prices, rulebook and human decisions) to learn
- AI needs objectives more than humans
- *Compute* costs in the many billions — Increasing returns to scale business

Summary conclusion

- Private sector and microprudential AI use *generally* positive
  - Ample data, mostly immutable rules and low cost of mistakes
  - Faster and more accurate decisions, with much less staff than now
  - Supervisors, risk managers, and central bankers are training their AI successors
- AI can *undermine macroprudential objectives*
  - Collusion, stress amplifying, booms and busts, criminality/terrorism and nation state attacks
- It will be essential for *crisis resolution* which is also where it poses the largest danger
- And may present its advice in a way that does not allow rejection — *decision-maker-in-effect*
- Leads to difficult human capital issues

Criteria for evaluating AI use in the financial authorities

1. Does the AI engine have *enough data*?
2. Are the rules *immutable* (static)?
3. Can AI be given *clear objectives*?
4. Does the authority the AI works for *make decisions on its own*?
5. Can we *attribute* responsibility for misbehaviour and mistakes?
6. Are the consequences of mistakes *catastrophic*?

Four conceptual challenges to AI use

1. Data limitations

- System generates petabytes daily
- May be badly measured (solvable)
- Confined to silos (hard to solve)
- Crises are rare (1 in 43 years)

2. Crises are unique

- Common crisis fundamentals
  - Leverage, self-preservation and complexity/information asymmetry
- Every crisis is unique in detail
- Crises are *unknown-unknowns* or uncertain

Both frustrate macroprudential AI learning

3. Strategic response

- The system *changes in response* to regulations — Goodhart's law and the Lucas critique
- Problem for all data driven analysis — particularly AI
- Usually manageable in microprudential regulations
- The macroprudential designers and supervisors need to consider the private sector's strategic response
- But most reaction functions are *hidden* until we encounter stress
  1. Danielsson-Shin — risk is *exogenous* or *endogenous*
  2. AI focuses on exogenous risk while endogenous matters for macroprudential
  3. Less important for microprudential as it mostly can work with exogenous risk

4. Mutable (non-static) objectives

- Rulebook is known in microprudential regulations and mostly immutable (on operational time scales)
- But in macroprudential policy
  - Mutability increases along with longer time scales and severity
  - Most important macroprudential objectives not known except at the highest levels of abstraction
  - We do what it takes to resolve crises
    - Change/suspend the law in the name of the higher objective of crisis resolution
    - Significant reallocation of resources
    - The political leadership takes charge
    - Resolution critically depends on information and interests that only emerge endogenously and *intuitively*
- AI has a stronger need to know objectives than humans but will find learning hard — AI is not good at intuition

Five destabilisation channels

1. Booms and busts — Procyclicality

- High fixed costs of control systems — increasing return to scale business
- Very expensive to run in-house for both private and public sector
- Handful of AI vendors
  - Risk management as a service (RMaaS) — BlackRock's Aladdin
- AI better than humans at finding best practices and state-of-the-art models

All of these

- Lead to more homogeneity in beliefs and actions — see the world and react to it in the same way
- Amplifying the cycle — procyclicality — more booms and busts

2. Self-preservation and stress amplification

- Private sector maximises profits 999 days out of 1000 and survival on the last 1 day
- Self-preservation during crises is destabilising — amplifying stress
  1. Flights to safety — investor strikes, liquidity hoarding and credit crunch
  2. Bank runs
  3. Fire sales
- AI speed and accuracy advantages over humans work against the system

3. AI interacting with AI

- Private AI may find it can best meet its objectives by bypassing or manipulating rules and regulations
  - Attack competing AI
  - Collude to manipulate markets
  - Collaborate to attack the authorities' AI
- Easier for AI as such behaviour is both very complex and often illegal
- It is better at handling complexity and coordination
- And may be *unaware of the legal nuances* unless explicitly instructed
  - It can be hard in an infinitely complex system to tell it all the things it is not supposed to do
- AI cannot be held to account, and its operators have a layer of deniability

4. Patrol an infinity complex system

Mistakes, misbehaviour, criminality and terrorism

- As the financial system becomes more complex
- Those finding loopholes increasingly gain an advantage
- Criminals and terrorists only need to find *one weakness*
- While the authorities have to monitor the *entire system*
- The system is, in effect, infinitely complex
- May be a *NP-hard problem* — impossible for the central bank's AI to handle

5. AI vs. humans intent on damage

Nation state attacks on the financial system

- As advice and decisions become increasingly automatic
- And humans left out of the loop
- Hostile nation states gain an advantage
- Can use hacking or humans to manipulate AI in preparation for attacking
- Which can be very hard to identify
- Humans know they are not supposed to attack. Does AI?
- Attack vectors can be in place for a long time
- And nation states can solve the problem of double coincidence

Human capital implications

- AI adoptions lead to cycles in staff skill sets
  - Case of AI in fraudulent transactions
- Over time, fewer and more highly skilled staff
- Both junior and senior staff are increasingly expected to have both *AI* and *domain knowledge*
- But the human capital pool for such people is very shallow
- And in demand across the economy
- Supervision and regulation design outsourced?

Criteria for evaluating AI use in the financial authorities

1. Does the AI engine have *enough data*?
2. Are the rules *immutable* (static)?
3. Can AI be given *clear objectives*?
4. Does the authority the AI works for *make decisions on its own*?
5. Can we *attribute* responsibility for misbehaviour and mistakes?
6. Are the consequences of mistakes *catastrophic*?

Task	Data	Mutability	Objectives	Authority	Responsibility	Consequences
Fraud/Compliance Consumer protection	Ample	Very low	Clear	Single	Mostly clear	Small
microprudential risk mana Routine forecasting	Ample	Very low	Mostly clear	Single	Clear	Moderate
Criminality Terrorism	Limited	Very low	Mostly clear	Multiple	Moderate	Moderate
Nation state atttacks	Limited	Full	Complex	Multiple & international	Moderate	Very severe
Resolution of small bank failure	Limited	Partial	Clear	Mostly single	Mostly clear	Moderate
Resolution of large bank failure Severe market turmoil	Rare	Full	Complex	Multiple	Often unclear	Severe
Global systemic crises	Very rare or not available	Full	Complex & conflicting	Multiple & international	Unclear even ex-post	Very severe