When AI becomes a central banker and supervisor: opportunities and pitfalls

Jón Daníelsson
London School of Economics
modelsandrisk.org

22 April 2024

AI in Central Banking
SUERF, BIS, ECB , Bank of Finland workshop

Bibliography

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

- My AI work modelsandrisk.org/appendix/AI

  a. "How AI can undermine financial stability"

     cepr.org/voxeu/columns/how-ai-can-undermine-financial-stability

  b. "On the use of artificial intelligence in financial regulations and the impact on financial stability"

     papers.ssrn.com/sol3/papers.cfm?abstract_id=4604628

  c. "Artificial intelligence and financial stability"

     cepr.org/voxeu/columns/artificial-intelligence-and-financial-stability

  d. "When artificial intelligence becomes a central banker"

     cepr.org/voxeu/columns/when-artificial-intelligence-becomes-central-banker

Our GPT — Illusion of Control

- chat.openai.com/g/g-dqceh7EH5-illusion-of-control
- Trained on our AI, risk and regulation work

What is AI?

- We see it as a rational maximising agent
- A computer algorithm performing tasks usually done by humans
- Making decisions and providing recommendations
- Which differs from machine learning and traditional statistics
- It not only provides quantitative analysis
- But also gives recommendations and makes decisions

AI strengths and weaknesses

- Great at identifying patterns in large data sets
- Can be very useful in interpolating in highly dimensional spaces
- Fast, quick and cheap — reliable in certain tasks
- Important that relevant data is in its training set
- Bad at extrapolating as lacks causal models, explicitly needs precise instruction from unrelated domains — econ theory, history, ethics, politics, psychology — Hallucination
- Unlike for humans, we have markets, firms, organisations, not clear how AI interacts with other AI and humans
- How does AI strategise and how to incentivise it to align its behaviour with our objectives?

Private sector use

- Banks are rapidly adopting AI and have large AI teams
- Even if many say publicly and privately that they are not intending to use much AI
- Very large cost savings in a highly competitive market
- Risk management, credit allocation, compliance, AML, fraud, KYC, ...
- The authorities will have no choice but to keep up if they wish to be relevant

What can AI do for the authorities?

- Design rules and enforce compliance with them
- Advise on and implement stress/crisis interventions
- Search for best outcomes given objectives and understanding
- Will likely enter by stealth into the authorities
- Presents advice in a way that is hard to reject — becomes a shadow decision-maker
- And becomes essential no matter what senior decision-makers wish

Five conceptual challenges to AI use. 1-3

1. **Data limitations**
   - System generates petabytes daily
   - Often badly measured and confined to silos
   - Crises are rare (1 in 43 years)
2. **Unknown unknowns**
   - Common crisis fundamentals
   - Every crisis is unique in detail
   - Crises are *unknown-unknowns* or uncertain
3. **System responses**
   - The system *changes in response* to regulations — Goodhart's law and the Lucas critique
   - Most reaction functions are *hidden* until we encounter stress

Five conceptual challenges to AI use. 4-5

4. Objectives

- Micropru rulebook known and immutable, not in macropru
- Mutability increases along with longer time scales and severity
- Most important macroprudential objectives not known except at the highest levels of abstraction

5. Incentives

- Regulations align private incentives with society
- The one-sided PA problem (institution – regulator) becomes two-sided (institution – regulator – AI)

1. Malicious use of AI

- Highly resourced profit-maximising agents not concerned about social consequences
- Bypassing controls/changing the system in a way benefiting them while difficult for others to detect
- Deliberately creating market stress
- Directly manipulating AI engines or using them to find loopholes
- Socially undesirable, even against the interests of the institution operating AI
- Most common are those careful to stay on the right side of the law
- Illegal activities. Rogue traders and criminals, terrorists and nation-states

2. Misinformed use and overreliance on AI

- When it does routine tasks well, trust builds up until
- Algorithms extrapolate to areas where data is scarce and objectives unclear
- AI presents confident recommendations about outcomes it knows little/nothing about — AI hallucination
- AI should have to provide an assessment of the statistical accuracy of its recommendations
- Authorities need to overcome their reluctance to provide statistical accuracy

3. AI misalignment and evasion of control

- No guarantee AI will do what it is instructed to do
- Impossible to pre-specify all the objectives
- Very good at manipulating markets — collusion, insider trading
- Can destabilise the system even when only doing what it is supposed to do
- When the objective of financial institutions is survival, AI amplifies existing destabilising behaviour — flights to safety, fire sales and investor runs
- AI will find it easy to evade oversight
- The authorities have to contend with both needing AI and it aiding the forces of instability
- We suspect the second factor dominates
- The more we use AI, the more difficult the computational problem for the authorities becomes

4.Risk monoculture and oligopolies

- AI business model is increasing returns to scale
- Three scarce resources: compute, human capital and data
- Harmonises beliefs and action
- Amplified procyclicality
- When authorities also depend on the same AI engine, as they will, they may not be able to identify the resulting fragilities until it is too late
- Oligopolistic nature of the AI analytic business increases systemic financial risk.
- It is a concern that neither the competition nor the financial authorities appear to have fully appreciated the potential for increased systemic risk due to oligopolistic AI technology in the recent wave of data vendor mergers

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
Micropru 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