Use What is AI?

Machine Learning

Al Risks

stem Risk

Crises

Bibliography

Global Financial Systems Chapter 20 Artificial Intelligence

Jon Danielsson London School of Economics © 2024

To accompany Global Financial Systems: Stability and Risk www.globalfinancialsystems.org/ Published by Pearson 2013

Version 11.0, August 2024

Global Financial Systems C 2024 Jon Danielsson, page 1 of 76

What is AI? Machine Learning

Use

AI Risks

DEADSO

tem Risk 000000000 Crises 0000000 olicy

Bibliography

Book and slides

Global Financial Systems Stability and Risk Jon Danielsson

 Updated versions of the slides can be downloaded from the book web page www.globalfinancialsystems.org

Private sector AI use

- Many financial institutions are rapidly adopting AI and have large AI teams
 - Some are very conservative, with legacy technology

AL Risks

Use

What is AI?

Machine Learning

- Others have modern technology stacks and staff attuned to new technology
- Even if many say publicly and privately that they do not intend to use much AI
- Very large cost savings in a highly competitive market
- Risk management, sales, product design, pipeline management, credit allocation, compliance, AML, fraud, KYC, ...

Machine Learning AI Risks

System Risk

Crises 0000000 **Policy**

Bibliography

Authority use of AI

Central banks, various regulators, ...

- The authorities are much slower in engaging with AI than the private sector
- Most appear to ignore AI entirely

Use

0.

- The risk is that they will be faced with fait accompli
- Al is entering by stealth into the authorities
- The authorities will have no choice but to keep up with AI if they wish to be relevant
- We return to this theme later

Use

What is AI?

•0000000000

Machine Learning

AI Risks

System Risk

Crises

licy

Bibliography

What is artificial intelligence (AI)?

Global Financial Systems © 2024 Jon Danielsson, page 5 of 76

Common definitions

AL Risks

Use

What is AI?

0000000000

Machine Learning

- "The science and engineering of making intelligent machines." McCarthy (1956)
- "The science of making machines do things that would require intelligence if done by men." Marvin Minsky (1968) (no relation to Hyman)
- "The study of agents that receive percepts from the environment and perform actions." Russell and Norvig (2020)
- "Al is the study of how to make computers do things which, at the moment, people do better." Rich (1983)
- "Artificial Intelligence is the field of study and technology development that focuses on creating systems capable of performing tasks that typically require human intelligence." ChatGPT (2023)

Our notion of AI

• I see it as a *rational maximising agent*

Machine Learning

Use

What is AI?

0000000000

• Following Russell and Norvig's (2020) taxonomy

AL Risks

- Because that resonates with how economists think
- A computer algorithm performing tasks usually done by humans
- Which differs from machine learning and traditional statistics
- Since it not only provides quantitative analysis
- But also gives recommendations and makes decisions
- While generative AI takes most of the oxygen, other types can be more important in finance

Recommend Russel (2019) for a general overview of AI



Strengths

AI can

- Interpolate in highly dimensional spaces
- Identify patterns in large data sets
- Use game theory to become better than humans at dealing with humans
- Be reliable, fast, quick and cheap
- Scale easily
- Reduce human errors
- Provide new insights and solutions
- Enhance security

Weaknesses

• Needs relevant data in its training set

Machine Learning

Use

What is AI?

0000000000

• Usually bad at *extrapolating* — *hallucinates* (later slide)

AL Risks

- May need precise instruction from seemingly unrelated domains
 - Economic theory, history, ethics, politics, psychology, ...
- Humans have markets, firms, and organisations, artefacts that allow them to pool their intelligence to make better decisions. It is not clear that interacting Als will be able to do so
- We don't know how to incentivise it to align its behaviour with our objectives
- We don't know how AI strategies
- Explainability still in its infancy (later slide)
- For risks, see later Section



Types of AI

- 1. Reactive machines basic AI that reacts to specific inputs with predefined responses. Rigid and inflexible, designed for specific tasks
- 2. Rule-based systems AI that operates based on predefined rules or logic. Flexible within the domain of rules but not adaptable
- 3. Limited memory AI (discussed later)
- 4. Narrow (weak) AI (discussed later)
- 5. Self-aware Al
- 6. Artificial general intelligence (AGI)

Our focus is on 3 and 4



- Machines that can interact with humans in an intuitive and human-like manner
- Understand and anticipate human actions based on an understanding of their mental states

6. Artificial general intelligence (AGI) (hypothetical)

- Al capable of understanding, learning, and applying knowledge across a wide range of tasks at a human level
- Extremely complex
- Highly adaptable and generalisable
- AI makes new and better AI hardware and software
 - Recall Google's AlphaZero
- AGI is an ongoing research goal, with significant advancements anticipated in the coming decades

Is it only a question of data and compute?

System Risk

• Narrow AI (weak AI)

Machine Learning

Use

What is AI?

• Designed for a specific task or narrow set of tasks

AI Risks

- Operates based on pre-programmed rules and learned patterns.
- Does not update its parameters with new data after initial training
- Example: A speech recognition system that doesn't adapt to new accents
- Limited memory AI
 - Can update and adjust parameters based on new data
 - Has a memory that allows for ongoing learning and adaptation
 - More dynamic than narrow AI, often using reinforcement learning or continual learning
 - Example: A self-driving car that improves its performance with more driving experience



Generative Al

- Models that generate new content similar to the data they were trained on
- Applications:
 - Text generation: Chatbots, writing articles and course assignments
 - Image generation: Creating art, design prototypes, deep fakes
 - Music generation: Composing new pieces
 - Data augmentation: Enhancing training datasets for machine learning models, stress testing
 - GPT, DeepArt, DALL-E, Jukedeck, MuseNet, ...

Use What is AI? Machine Learning AI Risks System Risk Crises Policy E

Non-generative AI

- Analyse data, make predictions and classify information without generating new content
 - Predictive analytics, forecasting
 - Natural language processing (NLP): Sentiment analysis, language translation, entity recognition
 - Autonomous systems: Self-driving cars, robotics
 - Image classifiers
 - Speech-to-text systems
 - Recommendation systems
 - Fraud detection systems

Use What is AI?

Machine Learning

AI Risks

System Risk

Crises

Policy

Bibliography

Machine learning (ML)

Global Financial Systems C 2024 Jon Danielsson, page 16 of 76

What is Machine learning (ML)?

- Algorithms (models) enabling computers to learn from and make predictions based on data
- Train a model on a dataset so it can do analysis and predictions
- Regressions and other traditional statistical methods are ML
- Useful sources Murphy (2023) and Prince (2023)

Supervised learning

• Model is trained on a labelled dataset

AL Risks

Machine Learning

0000000000

• $\{P_{AAPL,i}\}_{i=1}^t$

Use

- Each training sample is paired with an output label that the model aims to predict
 - P_{AAPL,t+1}
- Model learns to map inputs to outputs by minimising the error between the predicted and the true values
 - $\left| \hat{P}_{\mathsf{AAPL},t+1} P_{\mathsf{AAPL},t+1} \right|$
- Primarily used for classification (predicting categorical labels) and regression (predicting continuous values)
- Common algorithms include linear regression, GARCH, logistic regression, support vector machines, decision trees and k-nearest neighbours

Unsupervised learning

- Model is trained on a dataset without labelled outputs to infer the natural structure of the data
 - Entire text of the *Financial Risk Forecasting* book

Use

- Model tries to learn patterns and structures from the input data without supervision or labelled responses, looking for relationships, groupings and structures in the data
- Used for clustering (grouping data points into clusters) and association (finding rules that describe large portions of the data)

Reinforcement learning

- Involves
 - States the current situation
 - Actions choices available
 - Rewards feedback from the environment
 - Policy strategy used to determine actions
- Interact with the environment, take action and receive feedback
- Learn to associate actions with feedback and adjust its strategy to maximise cumulative rewards over time
- Used, e.g. in robotics, autonomous driving, game playing (e.g. AlphaGo), resource management, recommendation systems, algorithmic trading, fraud detection, credit scoring and risk management

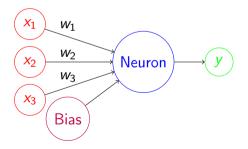


Neural networks (NN)

- Inspired by the human brain
- Interconnected layers of nodes (neurons) that process data in a hierarchical manner (next two slides)
- Structure
 - The input layer receives the initial data
 - The hidden layer performs computations and extracts features through weighted connections
 - Output layer produces the final prediction or classification
- Learning process
 - Neural networks learn by adjusting the weights of connections based on the error of their predictions
 - Backpropagation, a model's errors are propagated backwards through the layers to update and optimise the weights

Receives inputs, applies weights, adds bias, and passes through activation function to produce output

AL Risks



Machine Learning

000000000000

Use

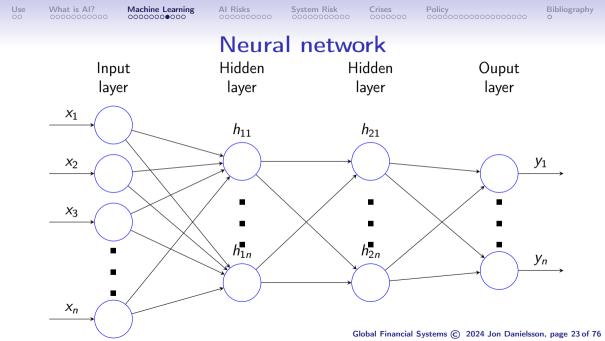
What is AI?

$$y = \sigma\left(\sum_{i=1}^n w_i x_i + b\right)$$

where σ is the activation function, w_i the weights, x_i the inputs and b the bias (constant)

Most common activation function, rectified linear unit (ReLU)

 $\max(0,x)$ Global Financial Systems © 2024 Jon Danielsson, page 22 of 76



Deep learning

• Neural network with many layers

Machine Learning

Use

What is AI?

- Each layer transforms the input data into more abstract representations
- Many hidden layers allow the model to learn hierarchical representations of the data
- Automatically extract relevant features from raw data

AL Risks

- Apply non-linear transformations to the data, enabling the network to model complex relationships and patterns
- Require large amounts of data and significant computational resources for training
- Overfitting controversial, as increasing the size and compute for the same data might or might not have worse out-of-sample performance

Active learning

- Model can query a user or information source to obtain labels for new data points
- Select the most informative data points to label

AL Risks

Machine Learning

Use

- Iteratively selects data points, queries for their labels and retrains itself
- By focusing on the most informative examples reduces the amount of labelled data needed, making it useful when labelling data is expensive or time-consuming
- Often involves human experts who provide the labels for the queried instances

Machine Learning 0000000000● AI Risks

stem Risk 0

0000000

Bibliography

Transformer

Vaswani et al. (2023)

- Model weighs the importance of different words in a sequence, regardless of their position
- Injects information (positional encoding) about the position of each word in the sequence, as neural networks can lack an inherent sense of order
- Can handle very large datasets and model complex patterns, making it highly effective for natural language processing tasks
- Applications:

Use

What is AI?

- Language translation
- Text generation
- Sentiment analysis
- Named entity recognition
- GPT (Generative Pre-trained Transformer) is a type of transformer

Global Financial Systems © 2024 Jon Danielsson, page 26 of 76

Use What is AI?

Machine Learning

Al Risks

System Risk

Crises

Policy

Bibliography

AI Risks and Issues

Global Financial Systems © 2024 Jon Danielsson, page 27 of 76

Al Risks

System Risk

Crises Pol 0000000 00

Bibliography

Main AI societal risks

Weidinger et al. (2022), Bengio et al. (2023), Shevlane et al. (2023), Bostrom (2014), Arthur (1994) and Varian (2018)

- 1. Trust, embedding and over-reliance
- 2. Misalignment with human values
- 3. Malicious use
- 4. Market power and concentration
- 5. Bias and discrimination
- 6. AI feedback loops and self-propagation risks
- 7. Misinformation and disinformation
- 8. Autonomy and loss of human control
- **9.** Erosion of privacy
- 10. Environmental impact
- 11. Ethical dilemmas in autonomous decision-making

System Risk

- Al builds up trust by being good at simple tasks that play to its strength
- As it gets promoted to increasingly sophisticated tasks

AL Ricks

000000000

- We may end up with the AI version of the Peter principle
- Al presents advice in a way that is hard to reject
 - Ends up a shadow decision-maker

Machine Learning

Use

- And becomes essential no matter what senior decision-makers wish
- Usually not credible when someone says
 - "We would never use AI for X"
 - "We will always have a human in the loop"
- So long as it delivers significant cost and efficiency savings

2. Misalignment — Objectives

System Risk

• How do we ensure AI does what it is supposed to do?

AI Risks

000000000

- One of the hardest problems for AI in complex social settings
- Engine needs to be given a precise objective function and cost-benefit calculations
- Impossible to do for problems that are effectively infinitely complex
- Note analogy with the legal system and incomplete contracts
- Object misspecification leads to suboptimal decisions, even bias (3.)
- And disastrous outcomes

Machine Learning

Use

- EURISKO and the canon of human knowledge and ethics
- Meting objective may depend on intuition to find analogous problems from seemingly unrelated domains that can provide creative solutions
 - Al is not good at intuition

1 and 2. Trust and misalignment — Explainability

• We demand human decision-makers explain their decisions

AI Risks

000000000

• Before hiring them, we can ask them to explain how they would react in particular cases

System Risk

• Explainable AI

Use

What is AI?

- Transparent and understandable AI decision-making processes
- Essential in many, even most, financial applications
 - Builds trust
 - Accountability
 - Compliance
 - Ethics
 - Model improvement

Machine Learning

- Most current AI models have no explainability
- But we are seeing new techniques emerge that help

2. Misalignment — Hallucination

- Al may generate outputs not based on reality
- When asked to extrapolate to cases not in its training dataset
- Because
 - Overgeneralisation, broad assumptions
 - Lack of context, insufficient understanding
- In, for example
 - Text, images
 - Implausible financial stress scenarios that can lead to wrong decisions
 - Beware of GenAl stress testing
- Difficult to detect except for experts
- Undermines trust

2 and 5. Misalignment, bias and discrimination

- Algorithmic bias
 - Erroneous internal representation that biases decision-making away from the desired outcome
- Due to objective misspecification, inadequate training data and/or spurious statistical patterns that change once the AI engine attempts to exploit them
 - Incorrect refusal of products
 - Racial bias
 - Inappropriate language in communication
- Can have serious consequences for the institution employing it
- Key risk holding back the use of generative AI in financial applications
- Significant progress being made

AL Ricks

3. Malicious Use of Al

- Use of AI technologies for harmful purposes
 - Find legal or regulatory loopholes
 - Crime

Use

What is AI?

- Terrorism
- Nation-state attacks

Machine Learning

- Non-financial examples:
 - Al-driven cyberattacks
 - Autonomous weapons systems
 - Manipulation of public opinion or elections
 - Deepfakes for deception or fraud
- Risks to security and stability

Bibliography

Al Risks

• The technology behind AI

Machine Learning

- Human capital
- Compute
- Data

Use

- Currently has increasing returns to scale
- Where all three reinforce the other
- A handful of IT firms end up with the people, data and compute
- Oligopolies and concentration
- Monopolies in specialised areas

Al engines train each other towards undesirable

outcomes

• Training data for other engines fed by

Machine Learning

• Visible outcomes of other Als' advice/decisions

AL Ricks

- Generated data from other AI
- Errors in one AI system propagate to others
- Feeds into

Use

- Feedback loops that reinforce undesirable behaviours
- Can lead
 - Bias amplification
 - Exploitation of weaknesses/loopholes
 - Coordination on undesirable equilibria

Use What is AI?

Machine Learning

AI Risks

System Risk

Crises

licy

Bibliography

AI Economic/Financial System Risks and Issues

Based on Danielsson and Uthemann (2024)

Global Financial Systems C 2024 Jon Danielsson, page 37 of 76



- Al depends on data
- Financial system generates terabytes of data daily
- It is often badly measured and confined to silos
- Little data on most important events as crises are rare (1 in 43 years)
- Observed data lives in the centre of the outcome distribution, not the tails

Al will not learn much about tail events — crises

2. Unknown-unknowns

System Risk

• All crises have the same code and common fundamentals

AL Risks

- Liquidity, leverage, one day in 1,000 problem, opacity
- While every crisis is unique in detail

Machine Learning

- Axiomatic, as else they (hopefully) would have been prevented
 - We don't know about crises successfully prevented
- Crises are unknown-unknowns or uncertain (Knight 1921)
- Little data to train on

Use

What is AI?

Al will not learn much about tail events — crises

3. Endogenous system responses — Setup

System Risk

- Neural networks are a reduced-form model ("super VAR") needing policy experiments in their training data (Sims 1980)
- Market participants change their behaviour when regulations/policy changes
 - Lucas (1976) and Goodhart (1974)

Machine Learning

Use

What is AI?

• The system changes in response to regulations and crises

AL Risks

• The reaction functions of market participants and the authorities is *hidden* until we encounter stress

Al finds it hard to learn about how the system behaves in crisis as there is little data to train on



- Neural networks don't learn much from past actions how future stress events play out as it knows little about causal relationships
- Such a problem is almost the opposite of what AI is good for
- When AI is needed most, it knows the least
- Al knowledge is negatively correlated with the importance of decisions

Wrong way risk

4. Objectives — Principal-agent

System Risk

• Regulations align private incentives with society

AL Risks

- Principal-agent problem between authority and institution (one-sided)
- Becomes two-sided (institution regulator AI)
- Regular ways to incentivise carrots and sticks don't work with AI
- Need pre-specification of detailed objectives
- Which is not sufficient

Machine Learning

Use

What is AI?

• Scheurer, Balesni, and Hobbhahn (2023) insider trading

We don't know how to incentivise AI

4. Objectives — What to tell it to do?

System Risk

- How to ensure AI follows the objectives desired by its human operators?
- Al needs clear objectives more than a human because of hallucination and lack of context
- Micro rulebook known and relatively immutable on operational time scales
- Macro rulebook is not

Machine Learning

Use

What is AI?

• Mutability increases along with longer time scales and severity

AL Risks

- Practical macro objectives not known
- The worst case is when the objective is unknown ex-ante and cannot be learned, as is the case with the worst financial crises

How to tell AI what to do when we don't know what we want?

4. Objectives — Speed and misalignment

- Al speeds up analysis and decisions
- Particularly problematic in crises
- Human regulators can no longer coordinate rescue efforts and "twist arms"
- See next Section on AI crises
- Al might cause a crisis before its human owner can answer the Fed Chair's call

How to tell AI what not to do when it acts so fast?

System Risk

- The best course of action for one player increases incentives for others to do the same
 - Multiple equilibria

Machine Learning

Use

What is AI?

• Rapid transition from one equilibrium to another

AL Risks

- Private-sector AI might coordinate on run equilibria
 - Their human owners would prefer to avoid
 - They recognise the damage of a systemic financial crisis to its owner

Al engines train each other towards undesirable outcomes

6. Malicious use

System Risk

• Al helps those intent on exploiting systemic vulnerabilities

AL Risks

- By coordinating and manipulating market conditions in a way that benefits them while destabilising the entire market
 - LTCM, Wolksvagen, Hunt brothers, GameStop + Reddit, ...
- Good at pushing into regions of the state space with the strongest complementarities
 - Converge on collusive strategies

Machine Learning

Use

What is AI?

- Jointly chasing the strongest possible bubbles and crashes
- Find complex cross-market strategies that humans have so far failed to identify
- Rogue traders, criminals, terrorists, nation states

Al helps all the highly resourced agents not concerned about social consequences

Bibliography

7. Risk monoculture

Increasing returns to scale of AI technology

- Risk monoculture drives booms and busts
- Few technology and information companies oligopolistic, and even monopolistic, market
- Financial data vendors have concentrated considerably over the past few years
- Analytics from the same vendor procyclicality

It is a concern that neither the competition nor the financial authorities have appreciated the potential for increased systemic risk due to oligopolistic AI technology in the recent wave of data vendor mergers Use What is AI?

Machine Learning

AI Risks

System Risk

Crises

olicy

Bibliography

AI Crises

Based on Danielsson and Uthemann (2024)

Global Financial Systems C 2024 Jon Danielsson, page 48 of 76

Technology and crises

- One of the earliest applications of the first transatlantic telegram cable in 1858 was the transmission of stock prices
- Nathan Rothschild supposedly used pigeons to get the first news of Napoleon's defeat at Waterloo in 1815 to manipulate the London stock market
- The one-day largest stock market crash in history, on 19 October 1987, was due to algorithmic trading, as was the stress event in June 2007 and several recent flash crashes

How financial institutions optimise

- Maximise profits given the acceptable risk
- Roy's (1952) criterion is useful
- Maximise profits subject to not going bankrupt
- That means financial institutions optimise for profits most of the time, perhaps 999 days out of 1,000
- However, on that one last day, when great upheaval hits the system and a crisis is on the horizon, survival, rather than profit, is what they care most about
- The "one day out of a thousand" problem

Use What is Al? Machine Learning Al Risks System Risk Crises Policy October Bibliography

- Options when faced with a shock 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



Al to crises

- Al excels at extracting complex patterns from data and, hence, can react faster and with more sophisticated strategies than humans
- Al engines can learn from their competitors
- What this means in practice is that AI engines in private firms and public organisations may end up optimising to influence one another
- Strategic complementarities



If AI decides to stay

- Will not sell or withdraw liquidity
- Buy
- Al is a stabilising force

Bibliography

If it wants to run

Crises

• Speed is of the essence

Use

What is AI?

- The first to react gets the best prices
- The last to act faces bankruptcy

Machine Learning

• Sell, calls in loans, run others as quickly as possible

AL Risks

- Makes the crisis worse in a vicious cycle
- Extreme volatility
- Al will significantly speed up and strengthen responses, making crises particularly quick and vicious
- Al acts as a crisis amplifier
- Days or weeks to minutes or hours

Use What is Al?

Machine Learning

Al Risks

System Risk

Crises

 Bibliography

Optimal policy response

Based on Danielsson and Uthemann (2024)

Global Financial Systems C 2024 Jon Danielsson, page 55 of 76



What can AI do for the authorities?

- Search for best decisions given objectives and understanding
- Design regulations and enforce compliance with them
- Advise on and implement stress/crisis interventions
- At a significantly lower cost than current arrangements
- But requires considerable changes in culture and resource allocation
- Very difficult for such conservative and bureaucratic entities



Policy

Bibliography

Relevance

- How the authorities respond will strongly impact the likelihood and severity of crises
- For them to remain relevant overseers of the financial system, they must meet the challenge of
 - Monitoring and regulating private sector AI
 - Address how private AI affect the effectiveness of regulations
 - Face the system risk it poses

Use

- Harnessing AI for their mission
- If the authorities do not credibly respond, the consequence will likely be more frequent and severe financial crises

Regulation and monitoring today to AI

- Done with
 - PDF reports, database dumps and conversations
- Does not easily allow the authorities to gauge how the private sector would react to
 - Potential regulations
 - Interventions
 - Identify stress resulting from 1 day out of 1,000 problem
 - Particularly from institution interactions
- Al can open up a new dimension in the authority-private sector interaction
 - Making regulations more robust and efficient
 - Allows better monitoring of misconduct and stress
 - Help in finding latent systemic risk
 - Design more efficient crisis interventions



Options

- A. Implement their own AI engines
- B. Set up Al-to-Al links
- C. Triggered central bank standing facilities
- D. Public-private partnerships

Use

Al Risks

System Risk

Crises

Policy

Bibliography

A. Authority AI system — Setup

- Set up their own AI engines
 - To design regulations
 - Evaluate the effectiveness of interventions
 - Gain expertise in Al
- Key characteristics
 - Speed
 - Objective specification



A. Authority AI system — Issues

- Such a system would not withstand the Lucas critique
- It would hence struggle with tail events
- However if
 - Well-designed
 - Informed both by
 - Confidential data
 - AI-to-AI benchmarking links discussed below,
- Would be of considerable benefit to both the micro and macro authorities



B. Al-to-Al links — Setup

- Al-to-Al links
- Authority AI directly talks to AI in private firms and other authorities
 - Legal framework
 - API
- Benefits
 - Regulating private-sector AI
 - Benchmarking
 - Test compliance
 - Perhaps how they decide on a loan application
 - Respond to a shock
 - And a crisis intervention
- Might not need data sharing



B AI-to-AI links — Gaming

- How would the system react
 - Lucas (1976) and Goodhart (1974)
- Depends on the effectiveness of the authority's setup
 - How it asks the private sector how it might respond to planned policy reactions
- To counteract gaming
 - Same policy intervention with a number of private AI
 - To ascertain whether a particular institution is attempting to game
 - Send a number of alternative queries
 - More difficult for the private AI to strategise



B Al-to-Al links — Hurdles

- No technological hurdles
- The data and sovereignty issues are more difficult
- But AI-to-AI API links focus on reactions and not data
- Benchmarking might not need access to the underlying technology
- Responses might be sufficient

C. Triggered central bank standing facilities

- Central banks prefer discretionary facilities *Constructive ambiguity*
- Too slow in AI crises (recall the last section)
- Facilities with predetermined rules for immediate triggered responses
- Can rule out bad equilibria
 - If private engines know CBs will intervene if prices drop by X
 - Will not coordinate on strategies that are only profitable if prices drop > X
 - Keeps moral hazard low
- If poorly designed, allow gaming and, hence, moral hazard

D. Public-private partnerships — Setup

- Authorities need AI that matches private-sector AI
- Hard to do in-house
 - Conservative, political and bureaucratic culture
 - Employment policies particularly difficult
 - High salaries of AI experts
 - Junior staff with high authority
- Outsourcing more likely
 - Private firms set up and run engines and AI-to-AI links
 - As already is increasingly common, e.g. in fraud, KYC, AML, credit risk, etc.

D. Public-private partnerships — Downsides

- Oligopolistic, even monopolistic, Al market structure
- Prevent the authorities from collecting information
 - Decision-making processes
- Private sector firms also prefer to keep technology proprietary and not disclose it, even to the authorities
- Lack of control
- Lack of knowledge of AI
 - But, does not prevent them from regulating derivatives without much derivative expertise

Use What is AI?

Machine Learning

Al Risks

System Risk

Crises

Policy

Bibliography

Applications

Global Financial Systems © 2024 Jon Danielsson, page 68 of 76

System Risk

• Design

What is AI?

Use

- Query private firms on hypothetical customer interactions
- Supervision
 - Query responses to routine interactions

AI Risks

Machine Learning

Policy

AL Risks

Crises

Policy

Bibliograp

Monitor stress

• Ask for stress scenarios

Use

What is AI?

• Simulate market stress scenarios

Machine Learning

- Aggregate responses and feedback across industry
- However, stress scenarios will likely continue to be backwards-looking, hence not improving the problem of the lack of tail observations

Interventions

- Simulations of alternative interventions via iterations
 - Iteratively query private sector AI and those of other agencies
 - Aggregate and feed responses back
- Find the best approach for
 - Liquidity injections
 - Forbearance
 - Bailouts

Use

- Identify the impact on market stability, credit availability, and overall economic health
- Allows for more informed decisions
- The mere presence of such an arrangement might act as a stabilising force in a crisis if perceived as competent and credible



- All four options need to be in place before the next stress event
- Compel the private sector to comply private sector will resist
 - *Might* see benefits in cooperating to prevent/resolve crises
- Stronger links only be activated in an emergency

 What is AI?
 Machine Learning
 AI Risks

 00000000000
 0000000000
 0000000000

System Risk

Crises

Policy

Bibliography

Kill switches

- Now suspend trading
- Not as viable for AI

Use

- System might not function effectively if a key AI engine is turned off
 - Significant market disruptions
- Difficult to predict AI responses to shutdown commands
 - might not account for the interconnected nature of modern financial markets where multiple AI systems interact
 - Al systems are designed to learn and adapt to their environment.
- Difficult to design a kill switch that can be effectively deployed without additional harm

Al Risks

System Risk

Crises 0000000

Bibliography

Bibliography I

Arthur, W. Brian. 1994. Increasing Returns and Path Dependence in the Economy. Ann Arbor, MI: University of Michigan Press. ISBN: 978-0472064960.

Bengio, Yoshua, Geoffrey Hinton, Andrew Yao, Dawn Song, Pieter Abbeel, Yuval Noah Harari, Ya-Qin Zhang, Lan Xue, Shai Shalev-Shwartz, Gillian Hadfield, et al. 2023. "Managing AI risks in an era of rapid progress." arXiv preprint arXiv:2310.17688.

Bostrom, Nick. 2014. Superintelligence: Paths, Dangers, Strategies. Oxford, UK: Oxford University Press. ISBN: 978-0199678112.

ChatGPT. 2023. ChatGPT's Definition of Artificial Intelligence. Https://www.openai.com/chatgpt.

Danielsson, Jon, and Andreas Uthemann. 2024. Artificial intelligence and financial crises.

Goodhart, Charles A. E. 1974. Public lecture at the Reserve bank of Australia.

Knight, Frank. 1921. Risk, Uncertainty and Profit. Houghton Mifflin.

Lucas, Robert E. 1976. "Econometric Policy Evaluation: A Critique." *Journal of Monetary Economics* 1.2 Supplementary:19–46.

Al Risks

System Risk

Crises 0000000 olicy

Bibliography

Bibliography II

McCarthy, John. 1956. "Dartmouth Conference: The Birth of Artificial Intelligence." In Proceedings of the Dartmouth Conference.

Minsky, Marvin. 1968. Semantic Information Processing. Cambridge, MA: MIT Press.

Murphy, Kevin P. 2023. *Probabilistic Machine Learning: Advanced Topics.* MIT Press. http://probml.github.io/book2.

Prince, Simon J.D. 2023. Understanding Deep Learning. MIT Press. http://udlbook.com.

Rich, Elaine. 1983. Artificial Intelligence. New York, NY: McGraw-Hill.

Roy, A. D. 1952. "Safety first and the holding of assets." Econometrica 20:431-449.

Russel, Stuart. 2019. Human compatible. Allen Lane.

Russell, Stuart, and Peter Norvig. 2020. Artificial Intelligence: A Modern Approach. 4th. Upper Saddle River, NJ: Pearson.

Scheurer, Jérémy, Mikita Balesni, and Marius Hobbhahn. 2023. *Technical Report: Large Language Models can Strategically Deceive their Users when Put Under Pressure*. arXiv: 2311.07590 [cs.CL].

Global Financial Systems © 2024 Jon Danielsson, page 75 of 76

Use

What is AI?

System Risk

Risk Crise

ises Polic

Bibliography

Bibliography III

Shevlane, Toby, Sebastian Farquhar, Ben Garfinkel, Mary Phuong, Jess Whittlestone, Jade Leung, Daniel Kokotajlo, Nahema Marchal, Markus Anderljung, Noam Kolt, et al. 2023. "Model evaluation for extreme risks." arXiv preprint arXiv:2305.15324.

Sims, Christopher A. 1980. "Macroeconomics and Reality." Econometrica 48, no. 1 (January): 1-48.

- Varian, Hal R. 2018. Artificial Intelligence, Economics, and Industrial Organization. NBER Working Paper 24839. National Bureau of Economic Research. https://doi.org/10.3386/w24839. http://www.nber.org/papers/w24839.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2023. Attention Is All You Need. arXiv: 1706.03762 [cs.CL].
- Weidinger, Laura, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, et al. 2022. "Taxonomy of risks posed by language models." In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 214–229.