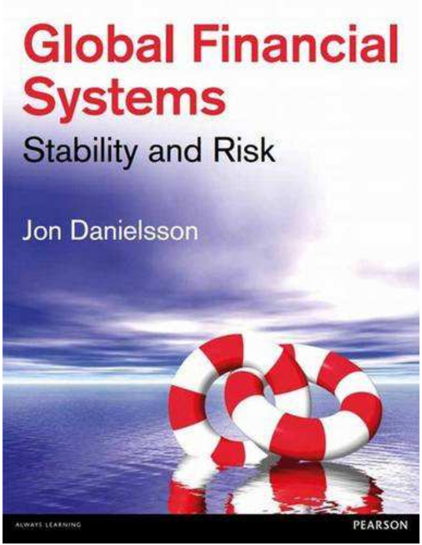
Global Financial Systems Chapter 20 Artificial Intelligence

Jon Danielsson London School of Economics

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

Version 11.1, December 2024

Book and slides



- 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
  - 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, ...

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
- The risk is that they will be faced with *fait accompli*
- AI 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

What is artificial intelligence (AI)?

Common definitions

- "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)
- "AI 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*
  - Following Russell and Norvig's (2020) taxonomy
- 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
- May need precise instruction from unrelated domains
- Humans have markets, firms, and organisations, artefacts that allow them to pool their intelligence to make better decisions. It is not clear that interacting AIs 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
- Matches patterns, does not understand (Mirzadeh et al. 2024)

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
4. Narrow (weak) AI
5. Self-aware AI
6. Artificial general intelligence (AGI)

Our focus is on 3 and 4

5. Self-aware AI — Theory of mind (hypothetical)

- 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)

- AI 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?

Generative AI

- 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, ...

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

Machine learning (ML)

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
  - Apple prices: $\{P_{\text{AAPL},i}\}_{i=1}^{t}$
- Each training sample is paired with an output label that the model aims to predict prices tomorrow
  - $P_{\text{AAPL},t+1}$
- Model learns to map inputs to outputs by minimising the error between the predicted and the true values
  - $\left|\hat{P}_{\text{AAPL},t+1} - P_{\text{AAPL},t+1}\right|$
- 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 *Financial Risk Forecasting*
  - Entire text of *Global Financial Systems*
- 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

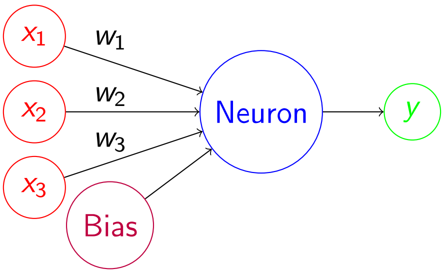
- Involves
  - States — the current situation
  - Actions — choices available
  - Rewards — feedback from the environment
  - Policy — strategy used to determine actions
- Interact with the environment, take action and receive feedback
- Learn to associate actions with feedback and adjust its strategy 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

Example of a neuron, or node

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



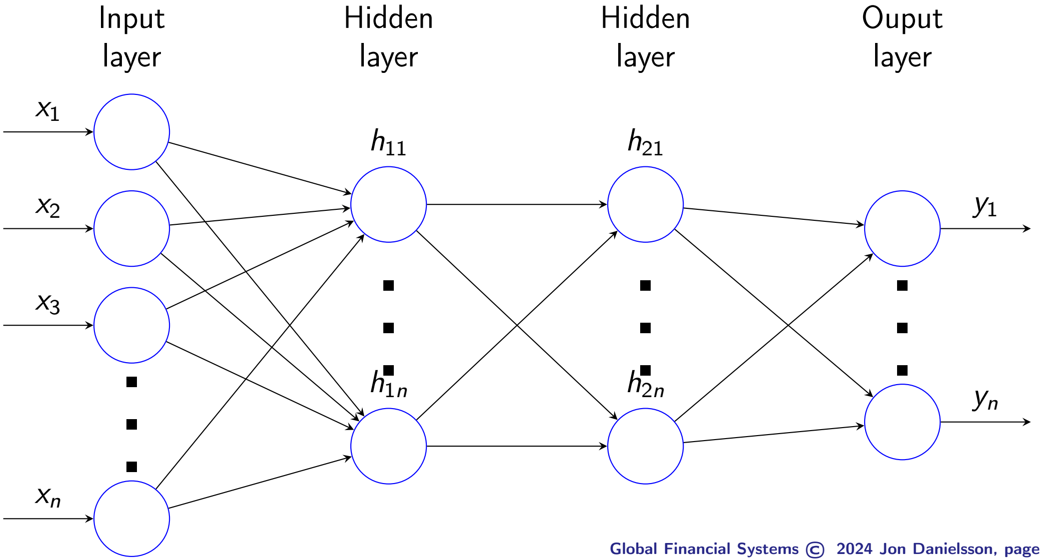
$$y = \sigma\left(\sum_{i=1}^{n} w_i x_i + b\right)$$

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)

$$\max(0, x)$$

Neural network



Deep learning

- Neural network with many layers
- 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
- 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
- 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

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

AI Risks and Issues

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

1. Trust, embedding and over-reliance

- 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
- Presents advice in a way that is hard to reject
- 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

- How do we ensure AI does what it is supposed to do? settings
- Engine needs to be given a precise objective function and cost-benefit calculations
- Impossible to do for problems that are effectively infinitely complex
- Suboptimal decisions, even bias
- AI is not good at intuition

1 and 2. Trust and misalignment — Explainability

- We demand human decision-makers explain their decisions
- Explainable AI
- Most current AI models have no explainability
- But we are seeing new techniques emerge that help

2. Misalignment — Hallucination

- AI may generate outputs not based on reality
- When asked to extrapolate to cases not in its training dataset
- Difficult to detect except for experts
- Undermines trust

2 and 5. Misalignment, bias and discrimination

- Algorithmic bias
  - 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

3. Malicious Use of AI

- Use of AI technologies for harmful purposes
  - Find legal or regulatory loopholes
  - Crime
  - Terrorism
  - Nation-state attacks

4. Market power and concentration

- The technology behind AI
  - Human capital
  - Compute
  - Data
- Currently has increasing returns to scale
- Oligopolies and concentration
- Monopolies in specialised areas

AI engines train each other towards undesirable outcomes

- Training data for other engines fed by
  - Visible outcomes of other AIs' advice/decisions
  - Generated data from other AI
- Feedback loops that reinforce undesirable behaviours
  - Bias amplification
  - Exploitation of weaknesses/loopholes
  - Coordination on undesirable equilibria

AI Economic/Financial System Risks and Issues

Based on
Danielsson and Uthemann (2024)

1. Data

- AI 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
    - No data on liquidity impact of large trades

AI will not learn much about tail events — crises

2. Unknown-unknowns

- All crises have the same code and common fundamentals
    - Liquidity, leverage, one day in 1,000 problem, opacity
- While every crisis is unique in detail
- 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)

AI will not learn much about tail events (crises) since little data on them

3. Endogenous system responses

- Neural networks are a reduced-form model ("super VAR") needing policy experiments in their training data (Sims 1980)
- The reaction functions of market participants and the authorities is *hidden* until we encounter stress
  - Lucas (1976) and Goodhart (1974)

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

3. Endogenous system responses — Wrong way risk

- 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

Wrong way risk

4. Objectives — Principal-agent

- Regulations align private incentives with society
- 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
- Scheurer, Balesni, and Hobbhahn (2023) insider trading

We don't know how to incentivise AI

4. Objectives — What to tell it to do?

- How to ensure AI follows the objectives desired by its human operators?
- AI 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
- Mutability increases along with longer time scales and severity
- 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

- AI speeds up analysis and decisions
- Particularly problematic in crises (next Section)

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

5. Strategic complementarities

- The best course of action for one player increases incentives for others to do the same
- Private-sector AI might coordinate on undesirable outcomes
  - Run equilibria
  - Market manipulation

AI engines train each other towards undesirable outcomes

6. Malicious use

- AI helps those intent on exploiting systemic vulnerabilities
  - LTCM, Volkswagen, Hunt brothers, GameStop + Reddit, ...
- Pushing into regions of the state space with the strongest complementarities
  - Collusive strategies
  - Chasing the strongest possible bubbles and crashes
  - Complex cross-market strategies that humans have so far failed to identify
- Rogue traders, criminals, terrorists, nation states

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

7. Risk monoculture

Increasing returns to scale of AI technology

- Risk monoculture drives booms and busts
- Humans are more heterogeneous than AI and hence are more stabilising
- 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

Kill switches

- Now suspend trading
- Not as viable for AI
- 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
    - AI systems are designed to learn and adapt to their environment.
- Difficult to design a kill switch that can be effectively deployed without additional harm

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

Problem

- 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

AI to crises

- AI excels at extracting complex patterns from data and, hence, can react faster and with more sophisticated strategies than humans
- AI 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
- AI is a stabilising force

If it wants to run

- 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
- Extreme volatility
- AI will significantly speed up and strengthen responses, making crises particularly quick and vicious
- AI acts as a crisis amplifier
- *Days or weeks to minutes or hours*

Relevance

- To remain relevant
  - Monitor and regulate private sector AI
  - Address how private AI affect the effectiveness of regulations
  - Face the system risk it poses
  - 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

- Today: PDF reports, database dumps and conversations
- Does not allow gauging of
  - Potential regulations
  - Interventions
  - Identify stress resulting from 1 day out of 1,000 problem
  - Particularly from institution interactions
- AI 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 AI-to-AI links

C. Triggered central bank standing facilities

D. Public-private partnerships

A. Authority AI system — Setup

- Set up their own AI engines
  - To design regulations
  - Evaluate the effectiveness of interventions
  - Gain expertise in AI
- 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. AI-to-AI links — Setup

- AI-to-AI 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 AI-to-AI 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, AI 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

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.

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.

Mirzadeh, Iman, Keivan Alizadeh, Hooman Shahrokhi, Oncel Tuzel, Samy Bengio, and Mehrdad Farajtabar. 2024. *GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models.* arXiv: 2410.05229 [cs.LG]. https://arxiv.org/abs/2410.05229.

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.

Bibliography III

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].

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.