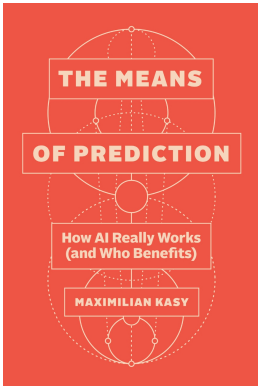
The Means of Prediction
How AI Really Works (And Who Benefits)

Maximilian Kasy

Department of Economics, University of Oxford

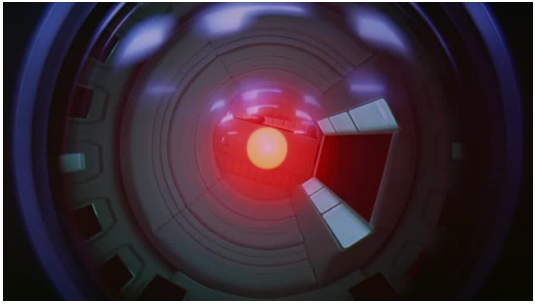
Fall 2025



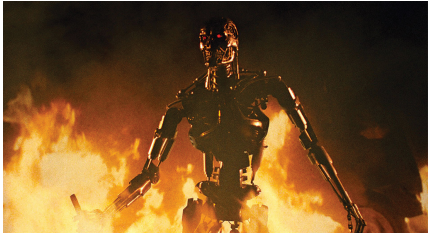
Are you scared of AI?

A popular dystopian story:

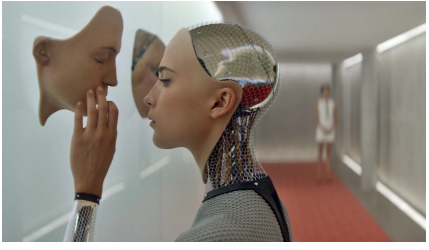
- AI will attain superhuman capabilities,
- will start to self-improve exponentially,
- and will threaten human existence in the name of self-preservation.



2001: A Space Odyssey



Terminator



Ex Machina

- Such stories touch on our deepest fears:
  - Losing our livelihoods, autonomy, lives, and loved ones,
  - to inscrutable and inevitable forces.
- But they don't enable good decisions:
  - Make it seem like AI and its use are fate.
  - Obscure conflicts over who controls AI.
- Intentional obfuscation by tech players?

A more accurate story

1. AI is automated decision-making using *optimization*.
2. Key issue: Who gets to pick the *objectives* that AI optimizes?
   (Not: Did the AI fail to optimize?)
3. Power flows from control of AI *inputs*:
   data, compute, expertise, energy.
4. We need *democratic control* of AI objectives
   by those affected by AI decisions.

Some examples

Beyond the headline-grabbing large language models:

- Algorithmic management of gig-workers.
- Automatic screening of job candidates.
- Filtering and selection of social media feeds, search engine results.
- Ad targeting.
- Predictive policing and incarceration.
- Automated choice of bombing/assassination targets (e.g. Gaza).

How AI works

The political economy of AI

Regulating algorithms

How AI works

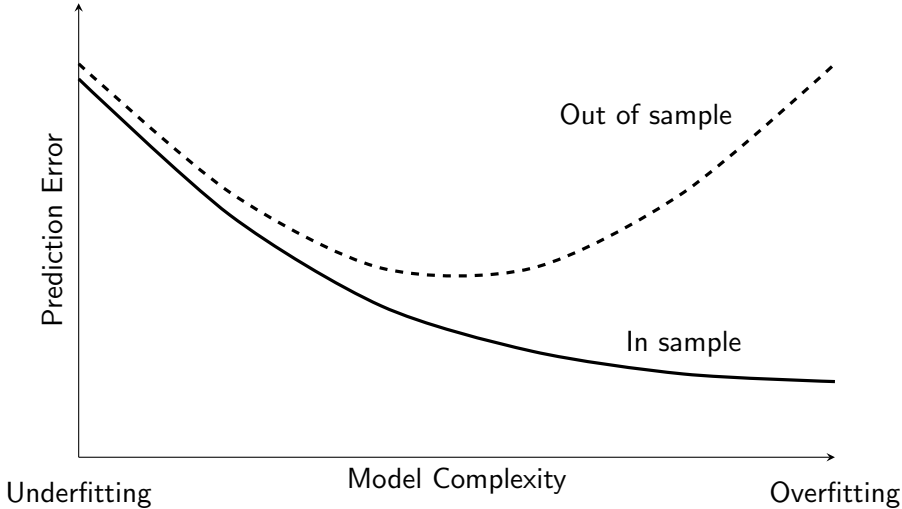
The book explains the foundations of machine learning and AI *without math*:

- AI is automated decision-making, maximizing some reward.
- Machine learning is AI using statistics.
  - Supervised learning: Prediction
    - Overfitting versus underfitting, tuning.
    - Deep learning, transformers.
  - Online learning: Choosing actions over time.
    - Exploration versus exploitation.
    - Planning.

Variance/bias tradeoff

- Prediction errors are due to either
  - estimation errors (variance), or
  - approximation errors (bias).
- More data $\rightarrow$ variance goes down.
- More model complexity (and thus compute) $\rightarrow$ bias goes down, variance goes up.

Tuning of supervised learning algorithms



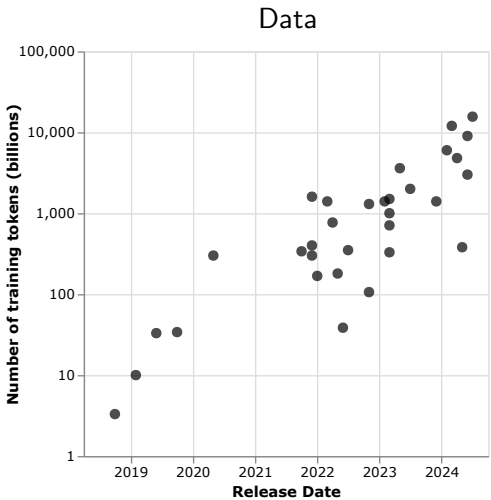
Scaling laws and the production function of AI

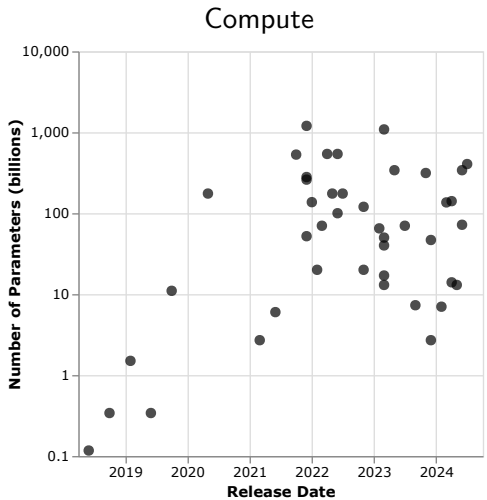
- Empirical counterpart: Scaling laws of LLMs. Write
  - L for the prediction loss (e.g. negative log likelihood),
  - N for model size,
  - D for data size.
- (Hoffmann et al., 2022): For $\alpha = .34$ and $\beta = .28$.,

$$L(N, D) = \frac{A}{N^{\alpha}} + \frac{B}{D^{\beta}} + L_0,$$

- Key motivation for the *bet on scale* of the AI industry.

Scaling of large language models





How AI works

The political economy of AI

Regulating algorithms

The means of prediction

- These foundations clarify what resources are needed for AI:
  - Data
  - Compute
  - Expertise
  - Energy
- Implications:
  - Potential for future improvements (domain-dependent).
  - Control of AI by controlling its inputs.
  - Contests over property rights, externalities.

Agents of change

Who can align AI objectives with social welfare?

- Interests, values, and strategic leverage.
- AI companies? Constrained by profit maximization.
- AI discourse should address others:
  - Workers (click-, gig-, tech-), consumers,
  - media and public opinion, state and law.
- Ultimate goal:
  - Democratic control of AI objectives
  - by those impacted by AI decisions.

Ideological obfuscation

- Ideology: Represents
  - Interests of a particular group as those of society at large.
  - Contingent choices as objective necessity.
  - Social relationships as technical ones.
- Popular AI stories that prevent change:
  1. *Man versus machine*: Obfuscates conflicts within society.
  2. *Intelligence explosion*: Not human choices but autonomous process.
  3. *Only experts understand AI*: Prevents democratic control of tech companies.
  4. *If we don't do it, China will*: Political inevitability.

How AI works

The political economy of AI

Regulating algorithms

Regulating algorithms

Ramifications of this perspective for various policy domains:

1. Value alignment and the limits of AI
2. Privacy and data ownership
3. Workplace automation and the labor market
4. Fairness and algorithmic discrimination
5. Explainability of algorithms and algorithmic decisions

Value alignment and the limits of AI

- Value alignment and AI safety:
  - Maximizing a slightly mis-specified objective can have bad consequences.
  - This formalizes the *man versus machine* stories.
- Analogous to multi-tasking (e.g. *teaching to the test*).
  - Reward design for AI $\approx$ incentive design for contracts.
- More important: Democratic control to align
  - the objectives of those controlling AI,
  - with the objectives of society at large.

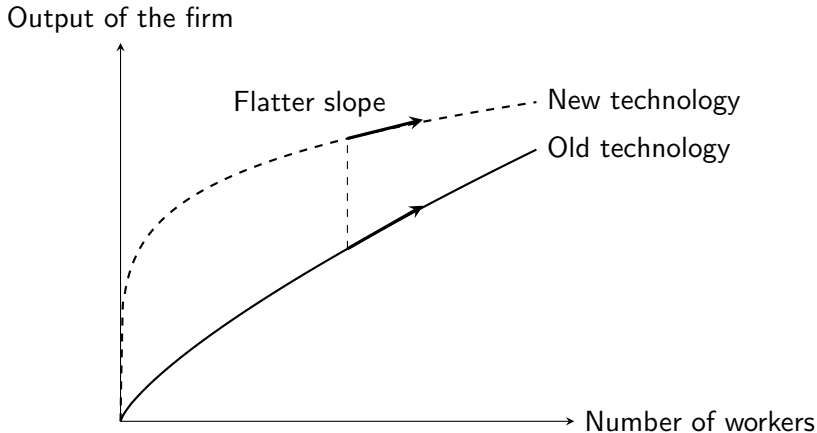
Privacy and data ownership

- Differential privacy:
  - (Almost) no observable difference whether your data are in a dataset.
- Individual property rights (e.g. GDPR):
  - Control over whom to share data with.
- But: Learning is all about the externalities.
  - Learning patterns, not individual observations.
  - Individual privacy / property rights cannot prevent harms from AI.
  - Need collective democratic governance of data.

Workplace automation and the labor market

- Micro-theory:
  - New technologies unambiguously increase average output, given inputs.
  - But the effect on marginal output is ambiguous.
  - It depends on technological choices.
- Automation and growth without shared prosperity:
  - Increased average output, decreased marginal output for workers.
- Who controls the development and deployment of new technology?
  - Who controls the objectives of workplace AI?
  - Co-determination and workplace democracy matter!

Growth without shared prosperity



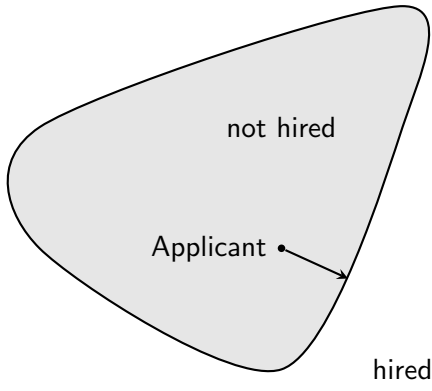
Fairness and algorithmic discrimination

- Most definitions of algorithmic fairness:
  - Treating people of the same "merit" (productivity, risk, etc.)
  - independently of their group membership.
- Algorithmic version of Becker's definition of taste-based discrimination:
  - Can decision be justified based on monetary objectives alone?
- Both:
  - Are supposed to reflect interests of disadvantaged groups,
  - but instead measure deviations from profit maximization.
- Alternative to this optimization-error perspective:
  - What is the causal impact of algorithms on inequality
  - between and within groups.

Explainability

1. Explaining decision functions:
    - Simple approximations to complicated functions.
    - Motivated by incompletely specified objectives.
2. Explaining decisions:
    - How would inputs need to change, to change decision?
    - Motivated by individual recourse.
3. Explaining decision problems:
    - What is the objective, action space, data used?
    - Motivated by collective democratic control.

Counterfactual explanations of decisions



Thank you!

Book available for ordering here:

https://press.uchicago.edu/ucp/books/book/chicago/M/bo255887145.html