Can We Have Pro-Human AI?

Daron Acemoglu
MIT

February 2024

SHAPING THE FUTURE OF WORK

AI Acceleration

AI research is picking up, and models are becoming increasingly powerful

Training computation used to train notable AI systems (in petaFLOPS, logarithmic scale)

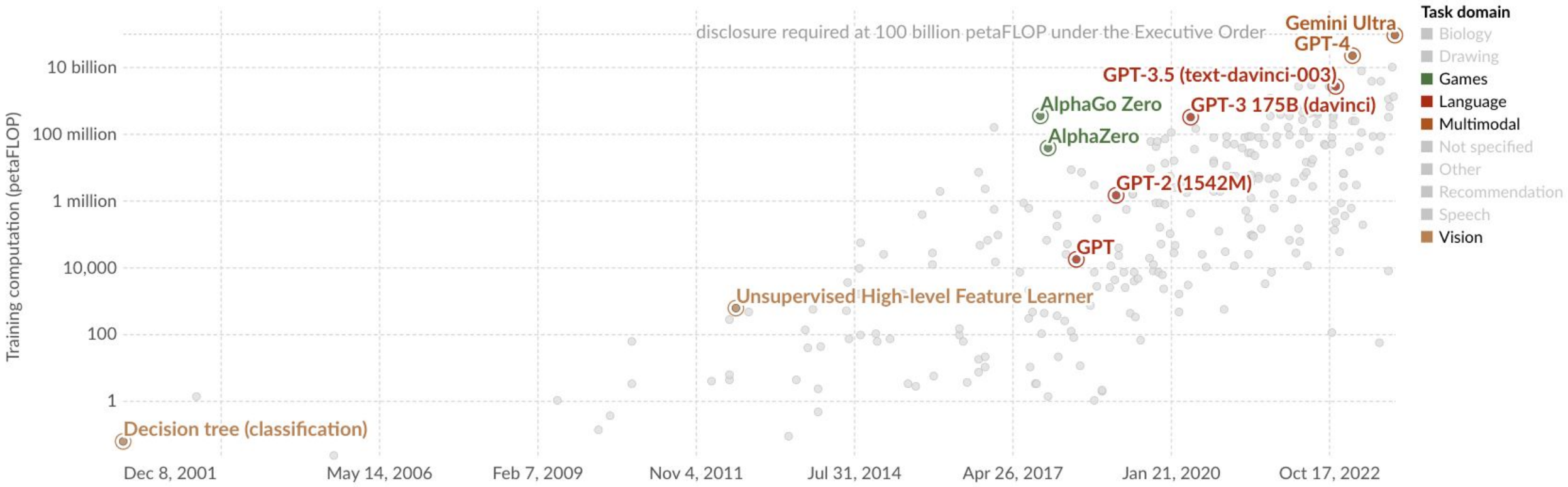


***Exhibit source**: Giattino, Charlie, Edouard Mathieu, Veronika Samborska and Max Roser. 2023. "Artificial Intelligence." Our World in Data.*
***Calculations based on**: Indiana University, University Information Technology Services. 2023. "Understand measures of supercomputer performance and storage system capacity."; NanoReview.net. 2023 "Apple M3 Max vs M2 Max."*

One petaFLOPS is one-thousand trillion (10^{15}) "floating-point operations per second". The most powerful MacBook Pro can perform 0.0164 petaFLOPS, about 1.42 quadrillion (10^18) operations per day, whereas the training of GPT-4 involved >10 billion petaFLOPS.

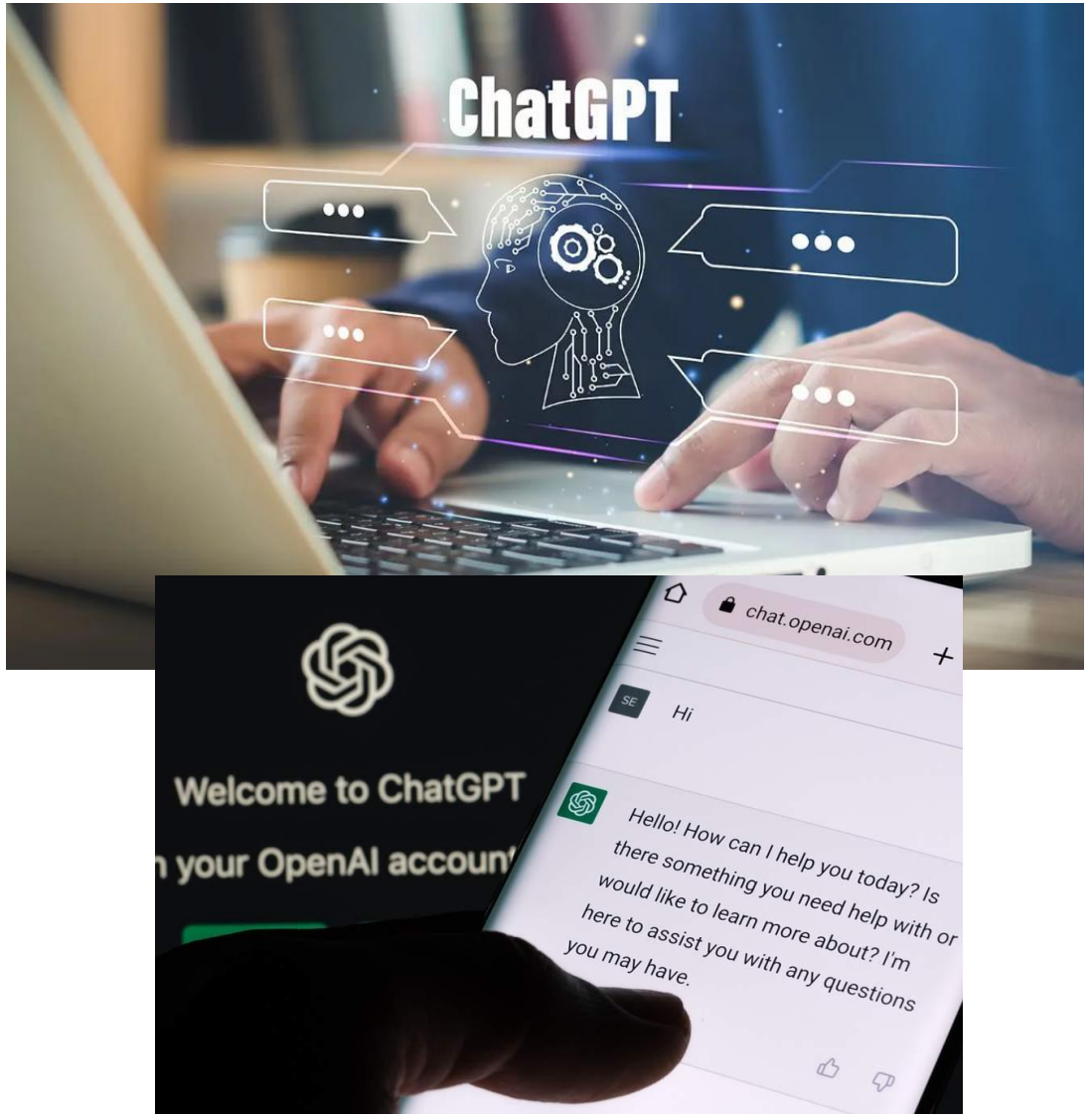
Generative AI

Generative AI and large language models (LLMs):

The promise is intelligent, flexible and easily usable tools that are complementary to human decision-making, technical work and creative tasks.

Can Generative AI and LLMs make this aspiration a reality?

Yes, there is promise, but also major roadblocks on the way, related to excessive automation, loss of informational diversity, human-AI misalignment, and monopolized control of information.

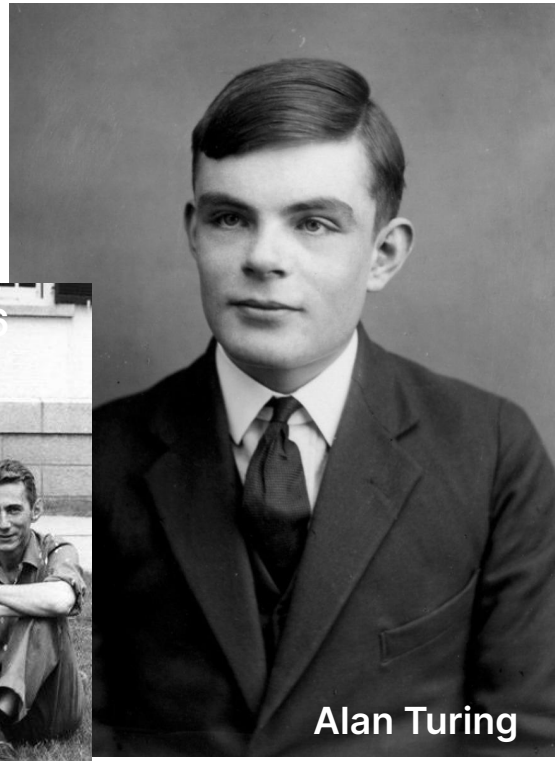


AI Beginnings: The Battle of Two Visions

Two fundamentally different visions of AI

1. Machines designed to be *smarter and more powerful* than (most) humans.

- The first vision – machine intelligence – refers to Alan Turing's conceptualization of how the mind works and how computers could imitate.



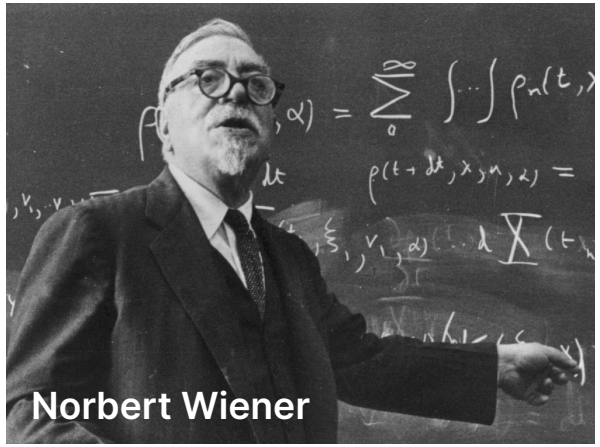
Source: *American Academy of Achievement*

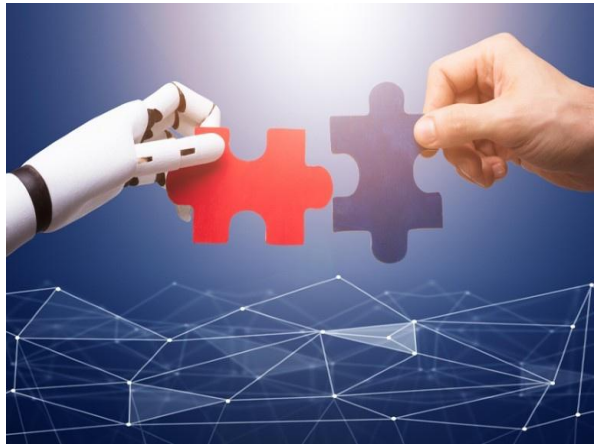
AI Beginnings: The Battle of Two Visions

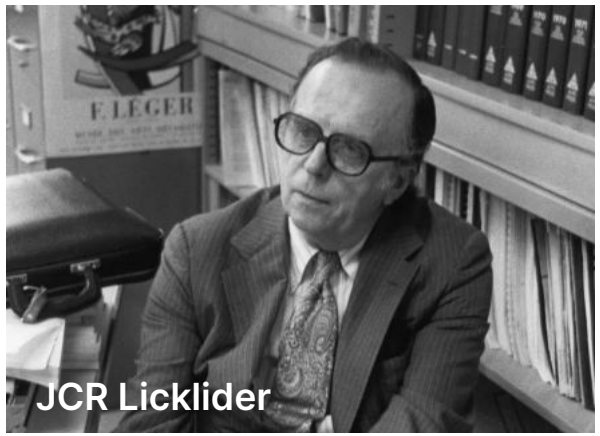
Two fundamentally different visions of AI

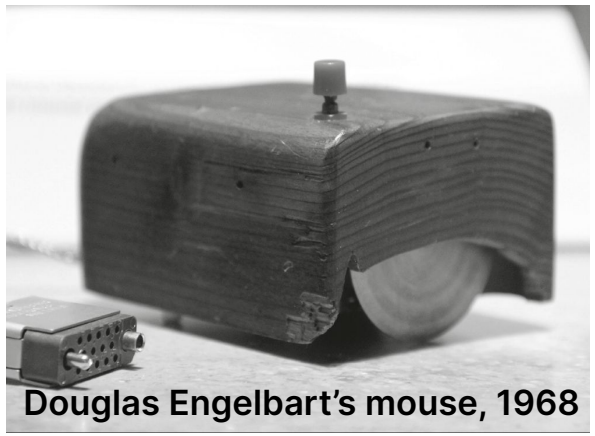
1. Machines designed to be *smarter and more powerful* than (most) humans.
2. Machines to *complement* human abilities.

- The first vision – machine intelligence – refers to Alan Turing's conceptualization of how the mind works and how computers could imitate.
- The second vision – let's call it machine usefulness or "pro-human AI" – starts with Norbert Wiener.
- Articulated and put into practice by computer scientists, such as JCR Licklider and Douglas Engelbart, "human-machine symbiosis"


Norbert Wiener




JCR Licklider


Douglas Engelbart's mouse, 1968

***Source**: Ben Shneiderman*

Pro-Human Generative AI

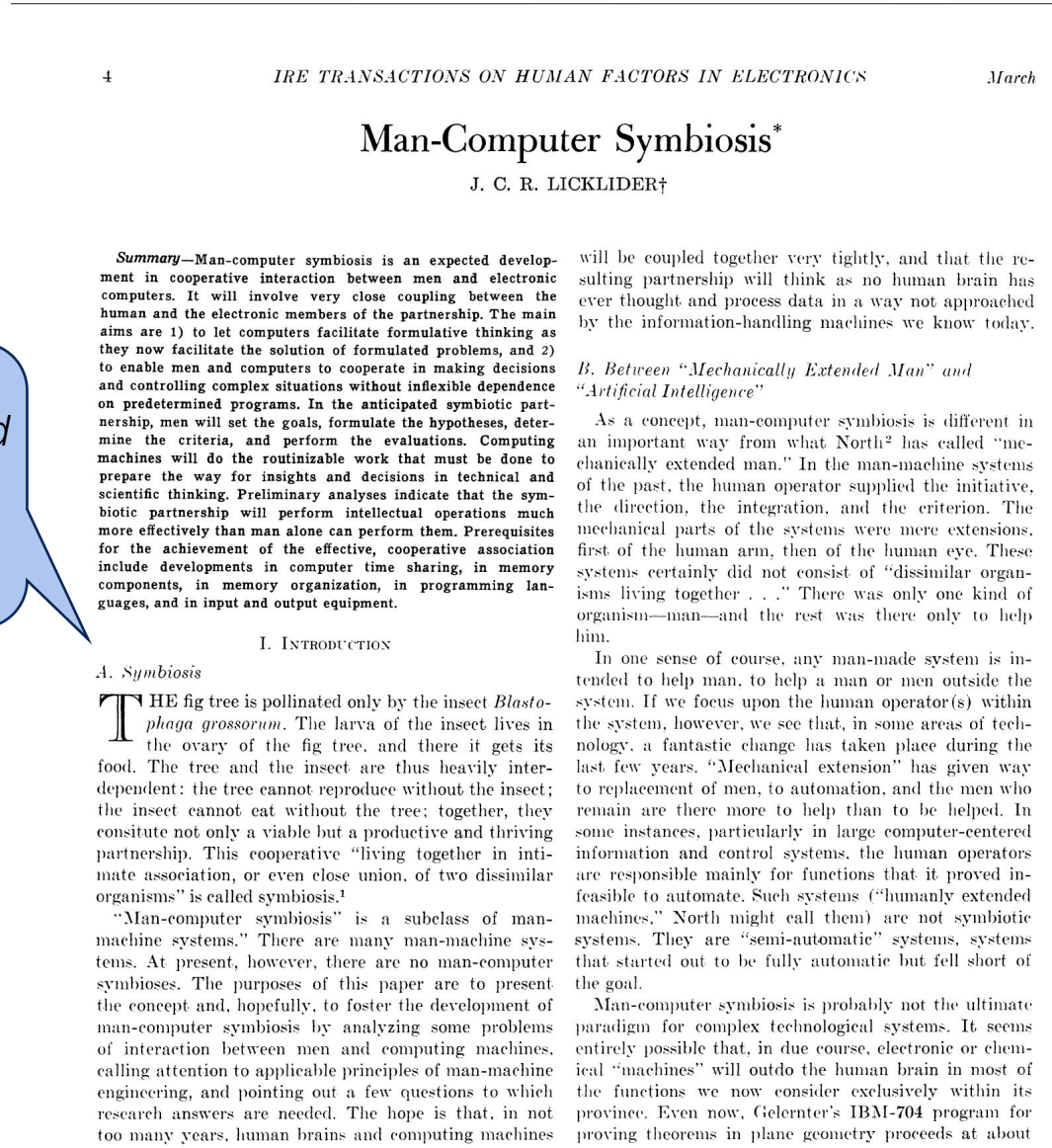
Generative AI could provide the tools for humans to get better in knowledge work

This is JCR Licklider's vision from 60 years ago:

> *The hope is that, in not too many years, human brains and computing machines will be coupled together very tightly, and that the resulting partnership will think as no human brain has ever thought and process data in a way not approached by the information-handling machines we know today.*

It requires generative AI tools to be useful to humans in better decision-making, problem identification, and information retrieval, filtering, and curation.

Example: how generative AI can help electricians and how AI can help blue-collar work.



4 IRE TRANSACTIONS ON HUMAN FACTORS IN ELECTRONICS March

Man-Computer Symbiosis*

J. C. R. LICKLIDER†

Summary—**Man-computer symbiosis is an expected development in cooperative interaction between men and electronic computers. It will involve very close coupling between the human and the electronic members of the partnership. The main aims are 1) to let computers facilitate formulative thinking as they now facilitate the solution of formulated problems, and 2) to enable men and computers to cooperate in making decisions and controlling complex situations without inflexible dependence on predetermined programs. In the anticipated symbiotic partnership, men will set the goals, formulate the hypotheses, determine the criteria, and perform the evaluations. Computing machines will do the routinizable work that must be done to prepare the way for insights and decisions in technical and scientific thinking. Preliminary analyses indicate that the symbiotic partnership will perform intellectual operations much more effectively than man alone can perform them. Prerequisites for the achievement of the effective, cooperative association include developments in computer time sharing, in memory components, in memory organization, in programming languages, and in input and output equipment.**

I. Introduction

A. Symbiosis

THE fig tree is pollinated only by the insect *Blastophaga grossorum*. The larva of the insect lives in the ovary of the fig tree, and there it gets its food. The tree and the insect are thus heavily interdependent: the tree cannot reproduce without the insect; the insect cannot eat without the tree; together, they consitute not only a viable but a productive and thriving partnership. This cooperative "living together in intimate association, or even close union, of two dissimilar organisms" is called symbiosis.[1]

"Man-computer symbiosis" is a subclass of man-machine systems." There are many man-machine systems. At present, however, there are no man-computer symbioses. The purposes of this paper are to present the concept and, hopefully, to foster the development of man-computer symbiosis by analyzing some problems of interaction between men and computing machines, calling attention to applicable principles of man-machine engineering, and pointing out a few questions to which research answers are needed. The hope is that, in not too many years, human brains and computing machines will be coupled together very tightly, and that the resulting partnership will think as no human brain has ever thought and process data in a way not approached by the information-handling machines we know today.

B. Between "Mechanically Extended Man" and "Artificial Intelligence"

As a concept, man-computer symbiosis is different in an important way from what North[2] has called "mechanically extended man." In the man-machine systems of the past, the human operator supplied the initiative, the direction, the integration, and the criterion. The mechanical parts of the systems were mere extensions, first of the human arm, then of the human eye. These systems certainly did not consist of "dissimilar organisms living together . . ." There was only one kind of organism—man—and the rest was there only to help him.

In one sense of course, any man-made system is intended to help man, to help a man or men outside the system. If we focus upon the human operator(s) within the system, however, we see that, in some areas of technology, a fantastic change has taken place during the last few years. "Mechanical extension" has given way to replacement of men, to automation, and the men who remain are there more to help than to be helped. In some instances, particularly in large computer-centered information and control systems, the human operators are responsible mainly for functions that it proved infeasible to automate. Such systems ("humanly extended machines," North might call them) are not symbiotic systems. They are "semi-automatic" systems, systems that started out to be fully automatic but fell short of the goal.

Man-computer symbiosis is probably not the ultimate paradigm for complex technological systems. It seems entirely possible that, in due course, electronic or chemical "machines" will outdo the human brain in most of the functions we now consider exclusively within its province. Even now, Gelernter's IBM-704 program for proving theorems in plane geometry proceeds at about

Gen-AI and Human Decision-Making: Proof of Concept

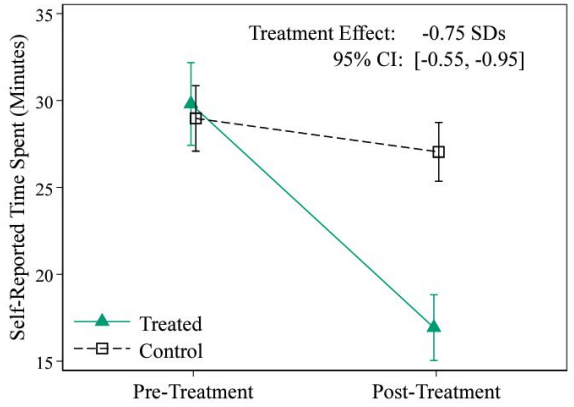
Programming: Peng et al. (2023) show software engineers with GitHub Copilot can be twice as fast.

Writing tasks: Noy and Zhang (2023) show that lower-productivity workers, given access to ChatGPT, improve performance in writing tasks.

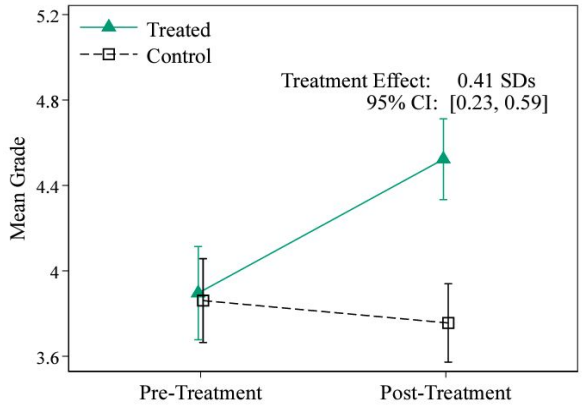
Customer service: Brynjolfsson et al. (2023) show improvement in worker productivity when ChatGPT is used in a human complementary manner—to provide information to operators.

Common element: the use of better information (from LLMs) as input for human decision-making to increase the effectiveness of human skills and expertise.

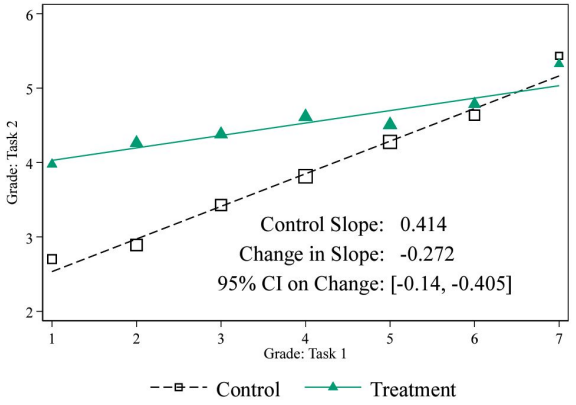
ChatGPT may reduce time to complete writing tasks…



…improve grades…

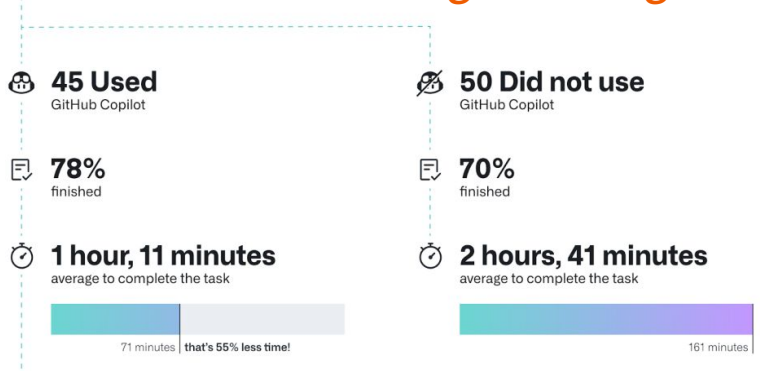


…and lessen grade inequality.



***Source**: Noy, Shakked and Whitney Zhang. 2023. "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence." Science. 381(6654): 187–192.*

Use of AI "Pair Programming"



Roadblocks to the Pro-Human Vision of AI

Four related, but distinct, roadblocks can be identified

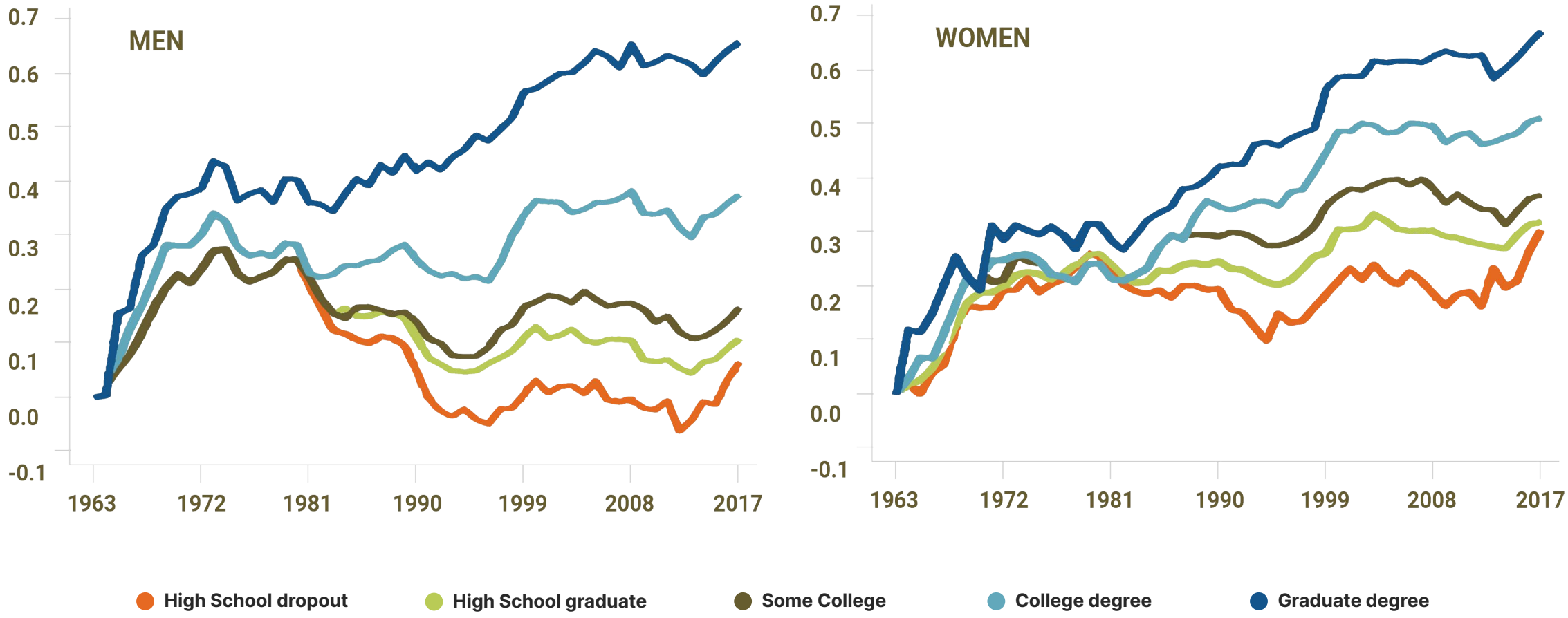
1. **Excessive automation** rather than input into human decision-making.
2. **Loss of informational diversity**—greater conformity, less diversity, less "information production", and externalities from new information.
3. **Misalignment between human cognition and AI algorithms**—incorrect perception or processing of AI inputs, could lead to bad feedback loops between machine and human.
4. **Monopolized control of information**—information from generative AI to manipulate rather than help people.

All four of these roadblocks have parallels with (but also differences from) previous digital technologies and waves of AI, from which we can learn.

***Source**: Generated using Midjourney*

Roadblock I: Inequality in the Age of Digital Technologies

The change in real (log) weekly earnings
Working age adults, ages 18–64, since 1963



Breakdown of shared prosperity during the era of digital technologies

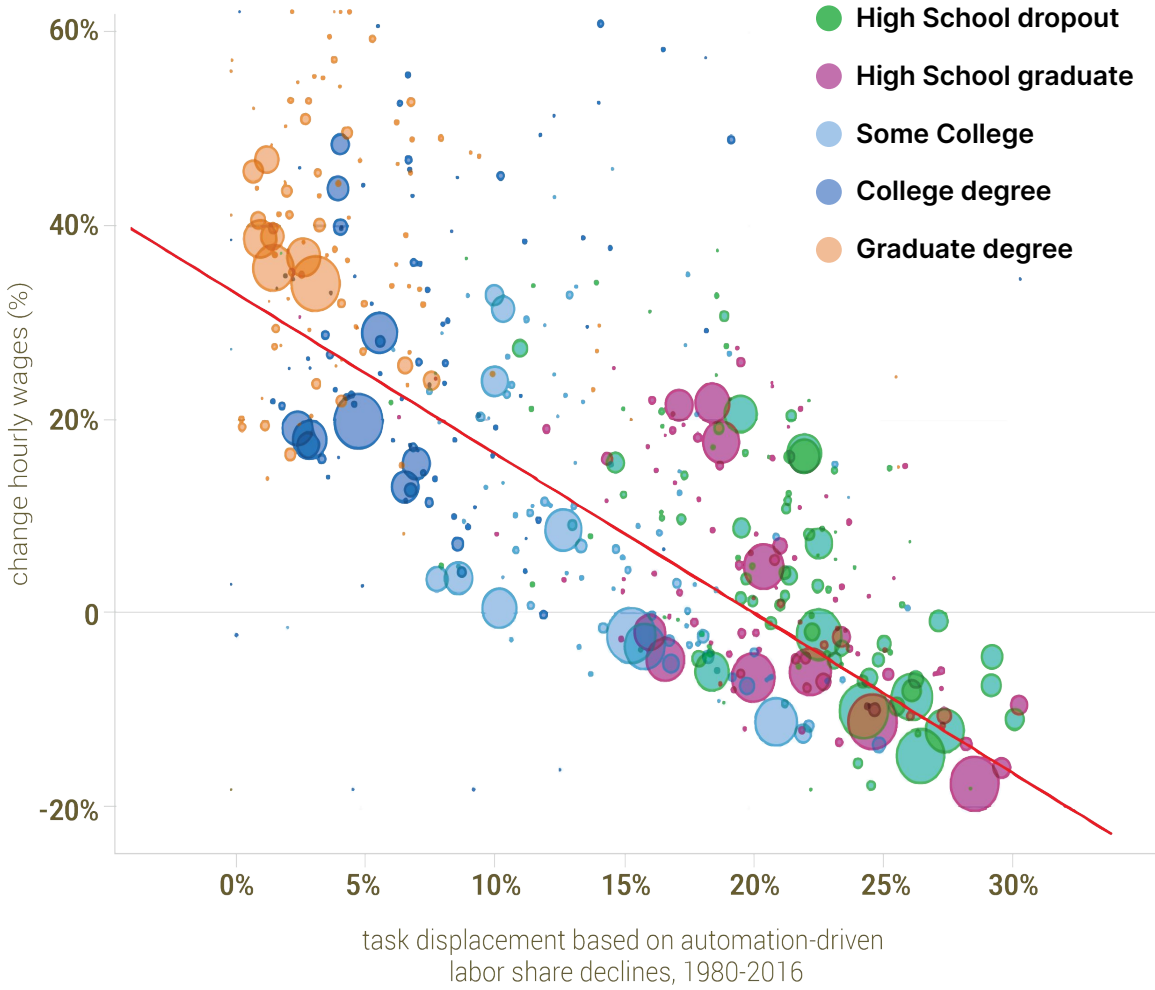
Source: *Autor, David. 2019. "Work of the Past, Work of the Future." AEA Papers and Proceedings, 109: 1–32.*

Roadblock I: Automation's Effects, Inequality and Stagnation

Changes in Real Wages due to Automation of Job Tasks

Companies may be tempted to use generative AI just for automation

- This will not unleash the full potential of generative AI in complementing human-decision-making. And will likely create more inequality (Acemoglu & Restrepo, 2018).
- **Learning from the past**: Past digital technologies used for automation, with adverse consequences for distribution and wages (Acemoglu & Restrepo, 2022):
  1. Limited productivity benefits because of "so-so automation"
  2. Inequality: groups more affected by automation suffered wage declines



Roadblock I: Automation in History

Same tensions in history

- Automation and excessive control over workers deepens inequality and does not raise productivity by much.
- **No automatic correction mechanism** to bend the arc towards shared prosperity.
- First phase of Industrial Revolution: emblematic of stagnant (even declining) wages, focus on automation, much worse working conditions.
- The second phase was better for workers, but not automatic in arrival, but was the result of:
  - Fundamental political reform;
  - Labor organization;
  - Redirection of technology.

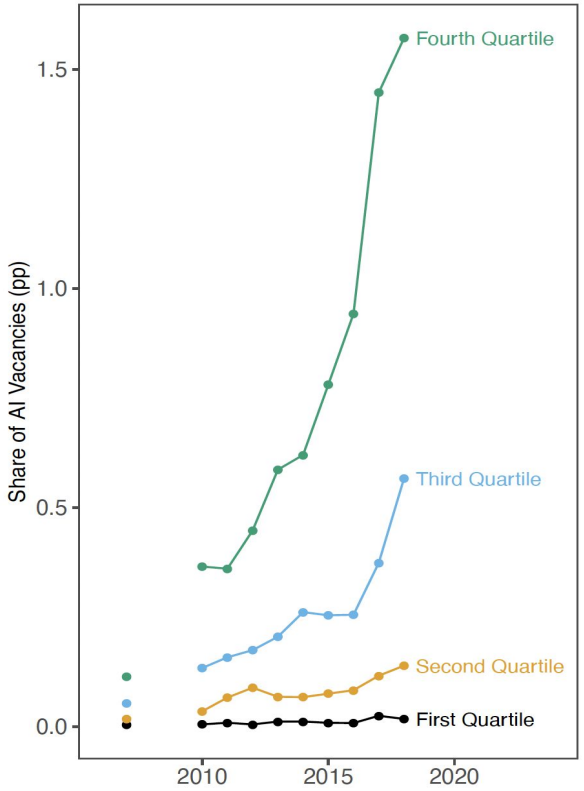

Power Loom, Lancastershire, 1835

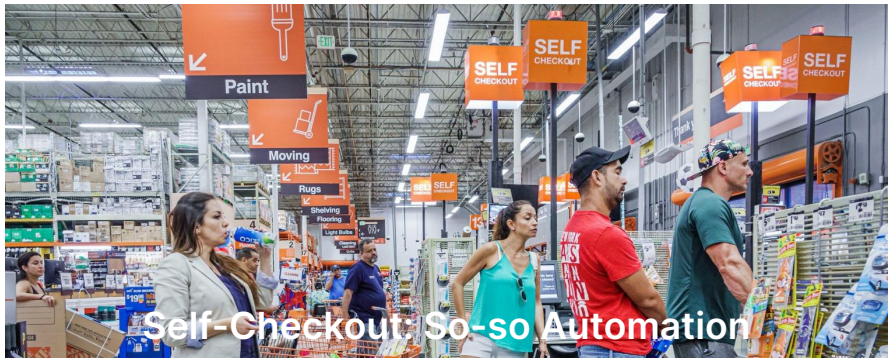
Roadblock I: AI and Automation

AI Adoption by Quartile of AI Task Exposure

There is already evidence that AI has been used for automation (rather than complementing humans)

- Establishments investing most in AI are those that used to perform tasks that were replaceable by basic AI (Acemoglu et al., 2022).
- And these same establishments slowed down their hiring after AI adoption.
- Also, concerns of so-so automation.
- LLMs seem to be going the same way—simple writing and analytical tasks are being automated in companies such as Buzzfeed and Bloomberg.
- The control of data from creative output, at the center of the Writers Guild strike, is going to be crucial for AI's use toward automation.



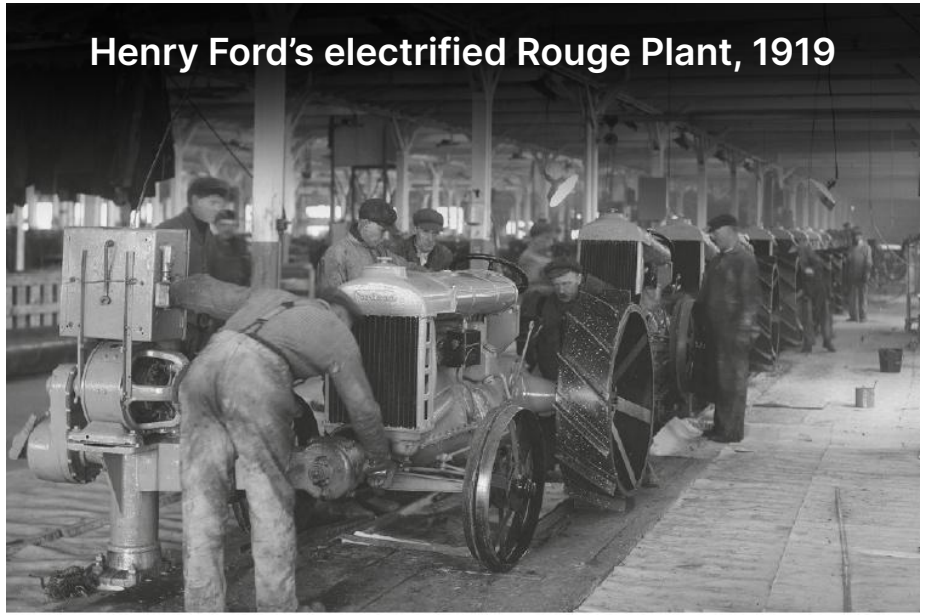


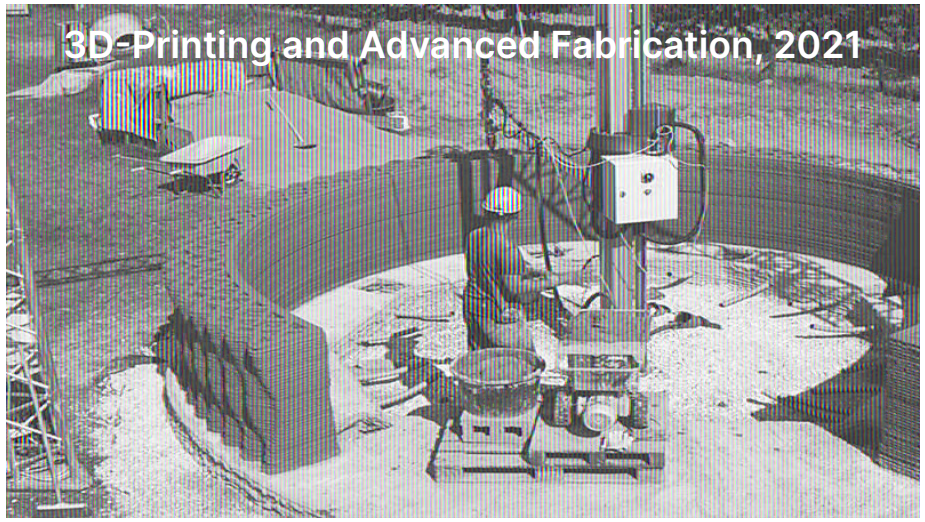
Self-Checkout: So-so Automation

Roadblock I: Better AI

The alternative path for AI is to create new human tasks

- Even using generative AI in existing human tasks to help workers is not enough.
- If this happens, it will likely devalue specific human skills (better AI-assisted writing would mean lower prices for writing skills and knowledge).
- This conundrum is solved with new tasks. These reinstate workers into the production process, increase worker contribution to productivity and boost earnings (Acemoglu and Restrepo, 2018).
- The promise of LLMs (and generative AI, more broadly) should be in this type of new-task creation.



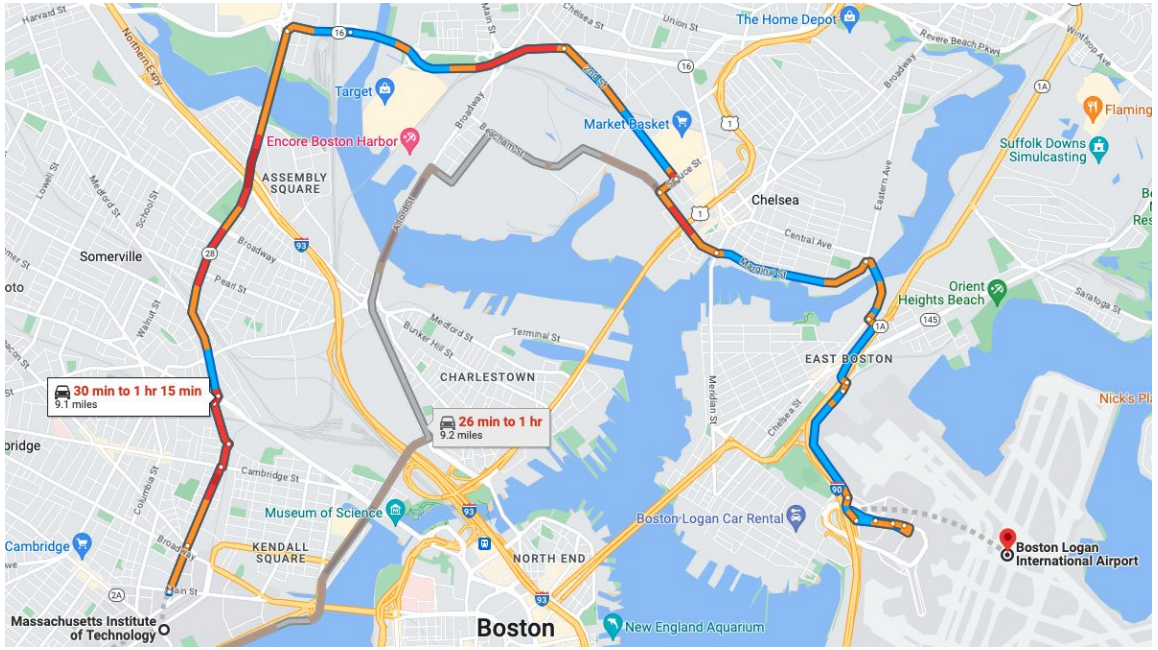


***Source**: 3D World's Advanced Saving Project, TECLA*

Roadblock II: Loss of Informational Diversity

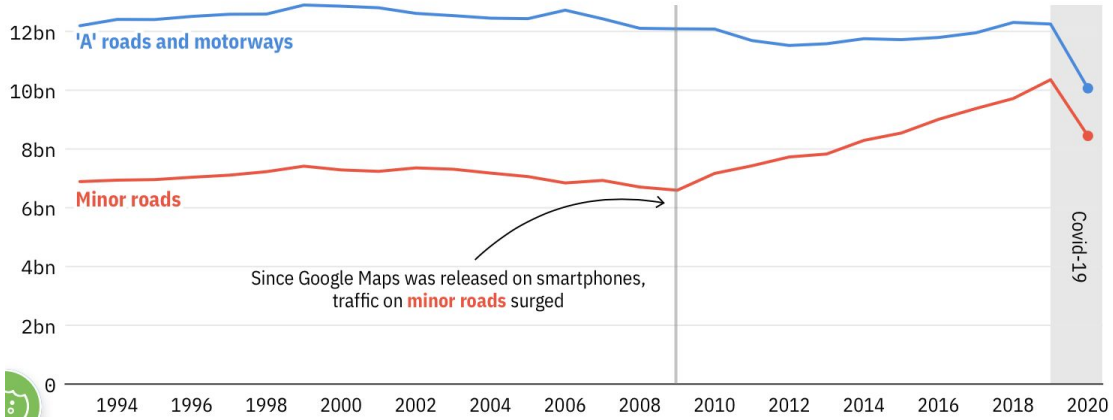
The promise of generative AI is in improving human decisions

- But can we do that if we lose human agency and diversity of human information?
- **Learning from the past**: GPS.
  - One of the first "intelligent" systems.
  - Its use for navigation hugely beneficial to individual drivers.
  - But systemic effects from better information may have been negative.
  - Why? Because traffic rerouted toward minor roads where congestion is more sensitive to traffic volume (theory: Acemoglu et al., 2018).



Traffic jams are moving to the backstreet

Vehicle miles in London traffic by road type



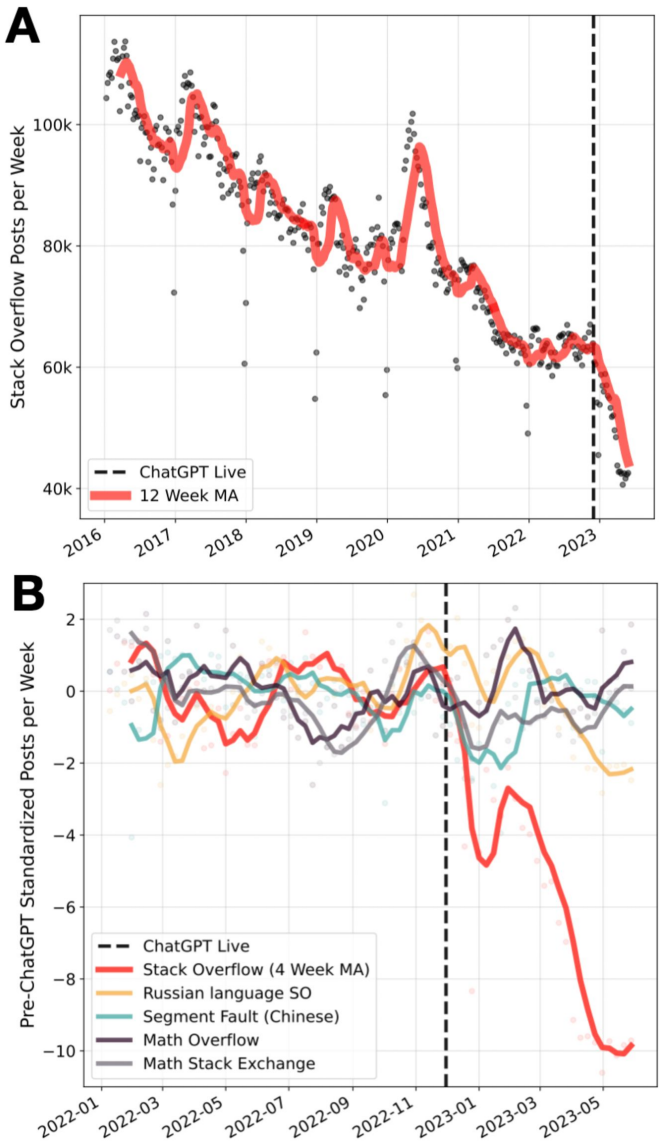
Roadblock II: AI Digging its Own Grave?

Posts on Stack Overflow and Related Sites

Generative AI intensifies these concerns

1. Greater conformity and informational "herding": If everyone uses LLMs for information, who produces new information?
    - New research on stack overflow from del Rio-Chanona et al. (2023) confirms these fears.
    - Similar issues from Wikipedia.
2. Bad information feedbacks, AI-AI interactions.
    - New research: significant generative AI model degradation from AI content

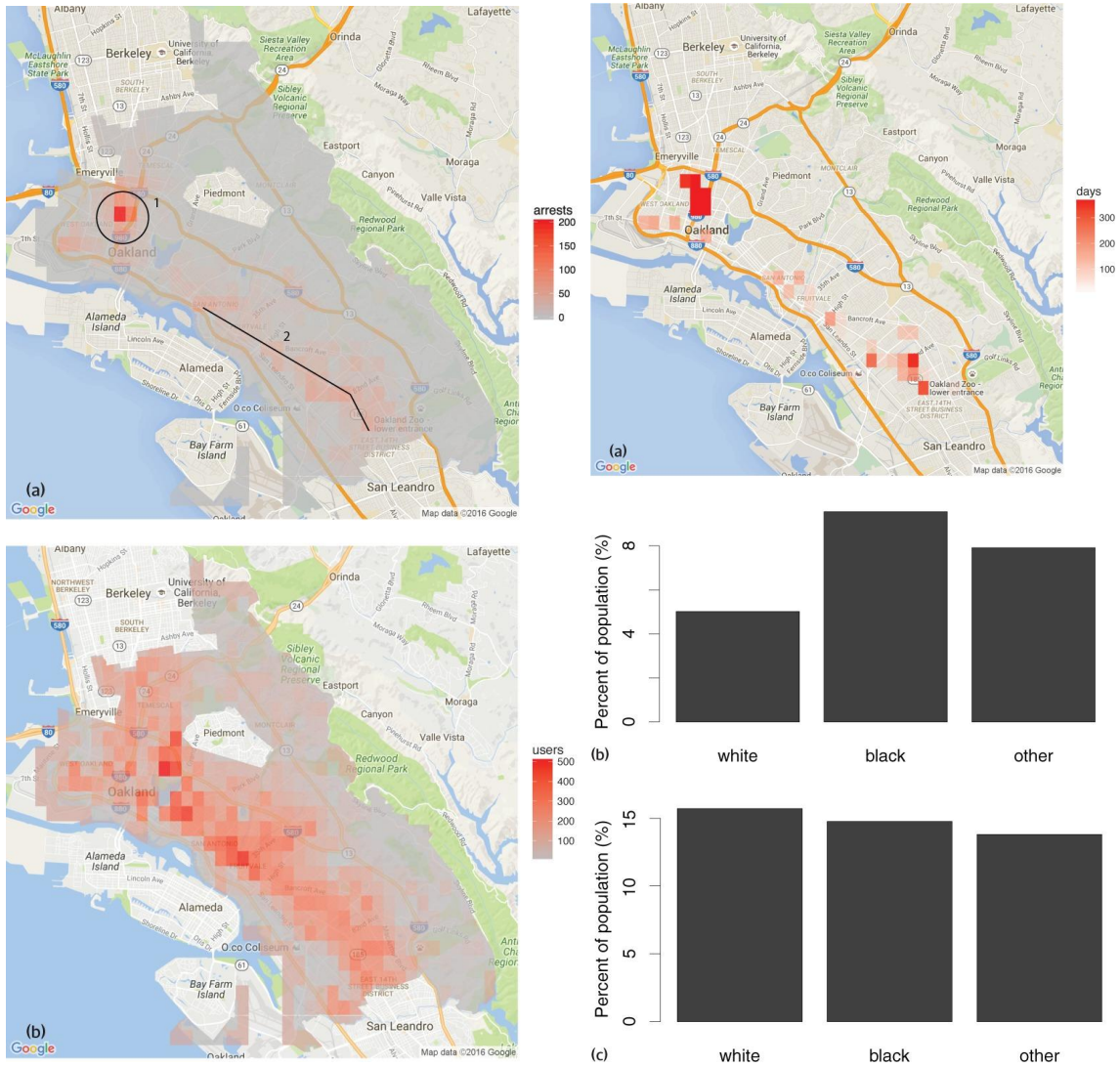
Bad human-AI feedback example:
1. Human Query: "Is policy X effective?"
2. LLM: "No"
3. Future Human Communication (e.g., on social media): "Policy X is not working"
4. LLM's new training data: "Policy X is not working"



Roadblock III: AI-Human Misalignment

Cognitive misalignment is a major problem

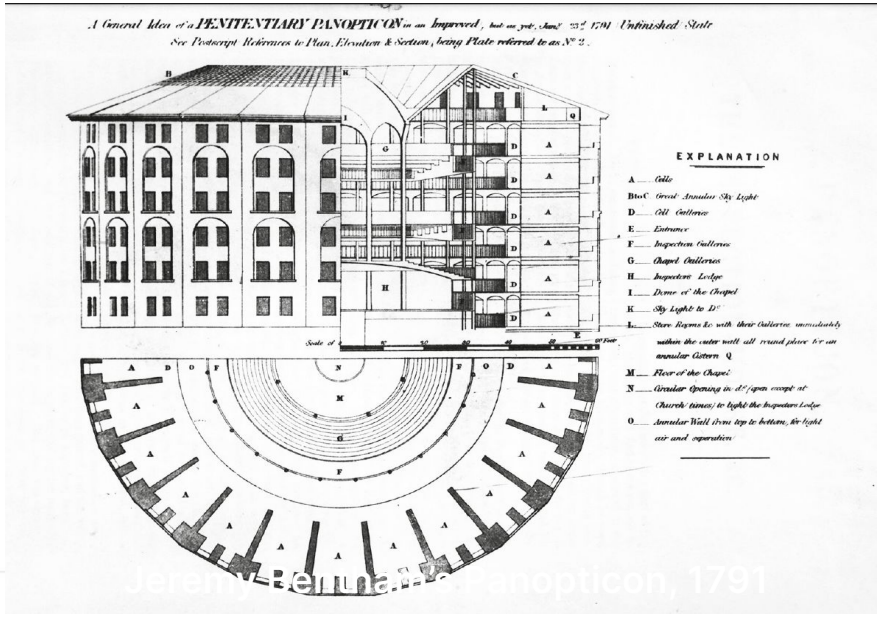
- Input into human decision-making is useful to the extent that humans use it correctly.
- Humans may misinterpret or mistrust algorithmic recommendations—or overreact to certain types of information and excessively change behavior.
- **Learning from the past**:
  - **Predictive policing**: Lum and Isaac (2016) find an overemphasis on where to expect crime.
  - **Misuse admits calibration of information**: Agarwal et al. (2023) show that physicians put more weight on information that AI recommends, and that they know already, but systematically underweight other information.



Roadblock IV: Control of Information

Who controls information?
Who benefits from information?

- **Learning from the past**: Historically, control of information has been a huge weapon.
- And very un-equalizing (surveillance in workplaces and the political arena).
- More recently, algorithmic recommendations for monetization of information—filter bubbles, viral content that is misleading or extremist.
- The key may be in how future business model of monetizing information will look.



Jeremy Bentham's Panopticon, 1791

Source: *Snopes*

Roadblock IV: Age of Manipulation?

Generative AI could intensify monopoly control of information and surveillance

- It can massively expand how information is presented in misleading ways.
  - Greater ability to manipulate via individually curated, emotionally charged material; deep fakes
  - Creating ads that are "more immersive and tailored" using generative AI.
- In the short run, this may amplify the amount of misinformation and disinformation.
- In the medium run, we may be in a world that Hannah Arendt foresaw:

  "If everybody lies to you, the consequence is not that you believe the lies, but rather that nobody believes anything any longer."

The world's biggest ad agency is going all in on AI with Nvidia's help

By Hanna Ziady, CNN

Published 8:47 AM EDT, Mon May 29, 2023

Sources: *Chris Ume (top); NVIDIA and WPP (bottom)*

How to Do Generative AI Better?

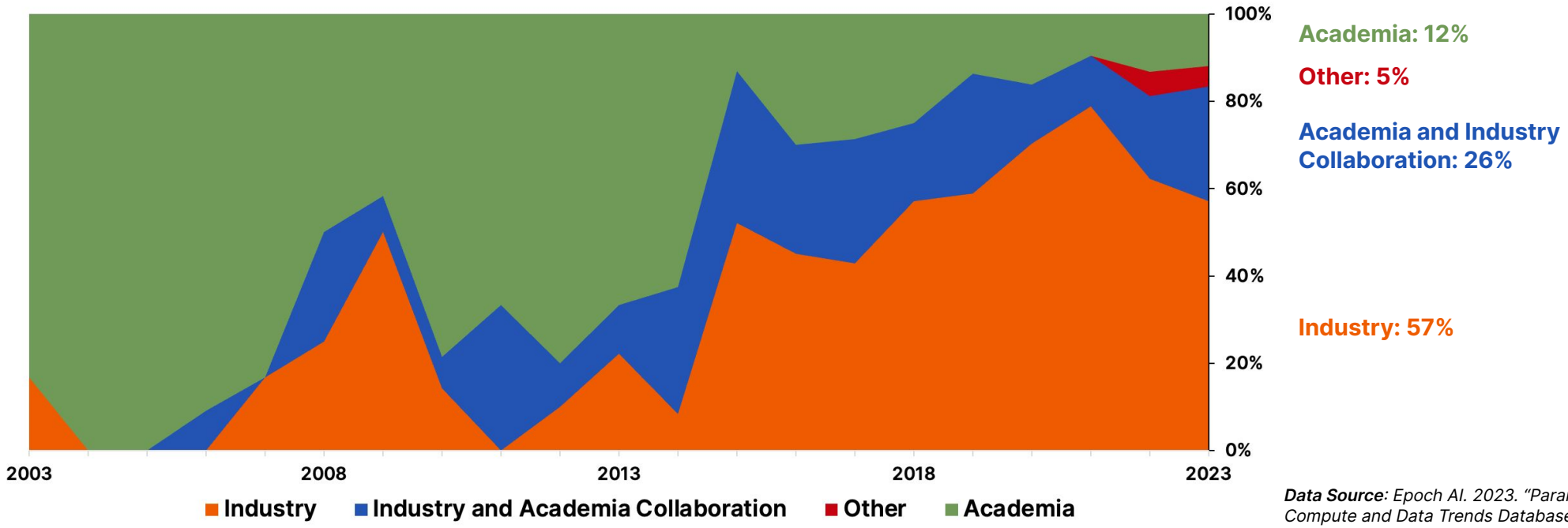
Redirection of AI: Make it more pro-worker and information-democratic

This will require new research and applications, with different focus. It will not happen by itself.

Good news: It is possible. **Bad news**: This is not where we are heading.

R&D of AI tools has become privatized and private incentives not always align with social ones.

Notable AI systems by researcher affiliation



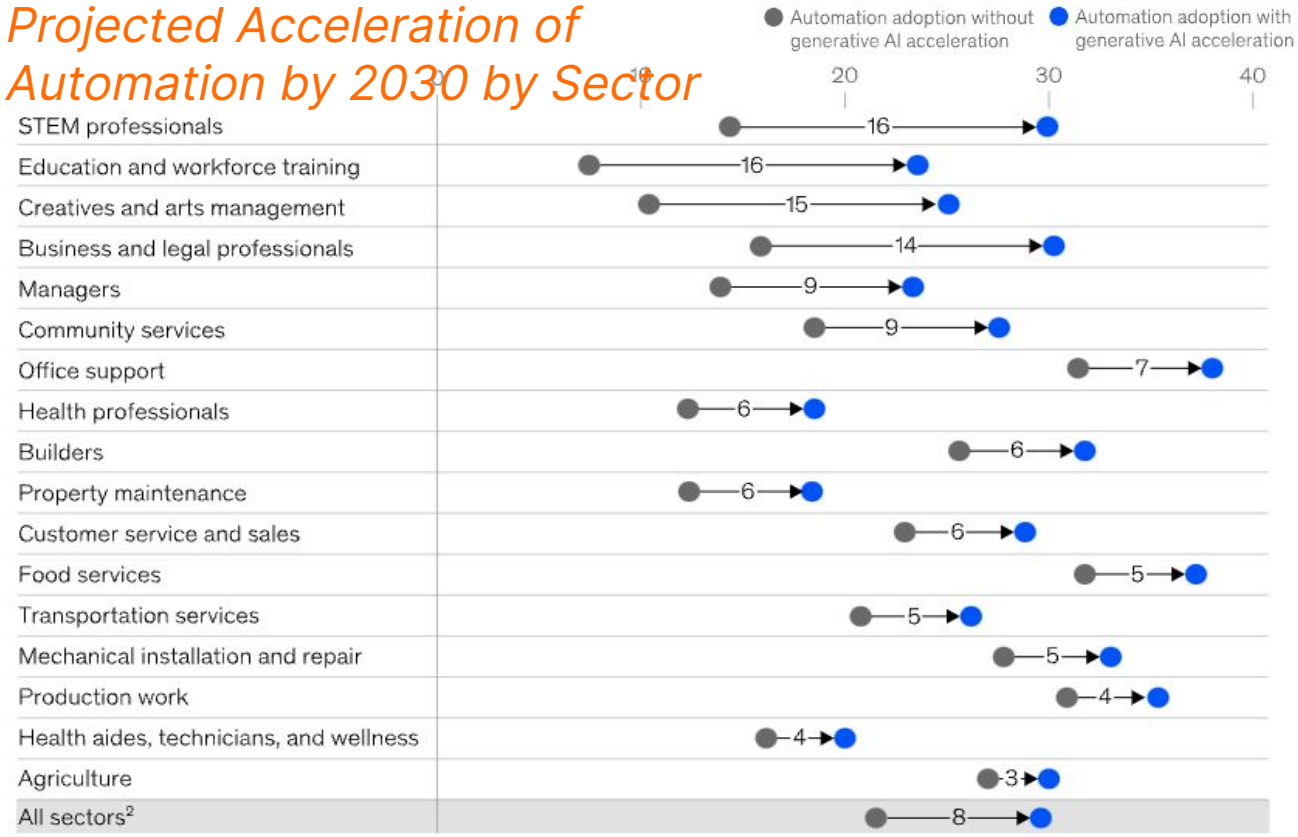
Data Source: *Epoch AI. 2023. "Parameter, Compute and Data Trends Database."*

How to Do Generative AI Better?

Policy and Social Changes

- **Reduce focus on automation**
- **Reform of tax policy** to remove capital-favoring asymmetries;
- **Data ownership and markets** to change business models and increase AI quality;
- **Digital ad taxes** to create room for new business models other than those that are exploitative of information;
- **Government subsidies to human-complementary AI** to counteract excessive focus on automation;
- **New norms and regulations** on the development and deployment of AI;
- Possibly also **new architecture of generative AI** for greater human complementarity

Projected Acceleration of Automation by 2030 by Sector



***Source**: McKinsey Global Institute. 2023. "Generative AI and the Future of Work in America."*