

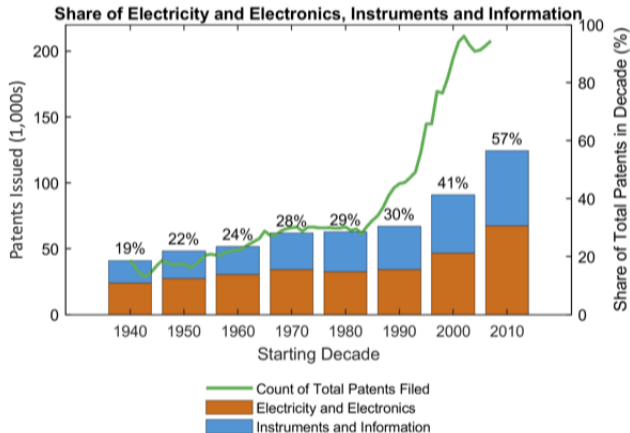
End of Innovation or the Distorted Future of AI?

Daron Acemoglu

AoM, August 2023

End of Innovation?

- ▶ Hardly.
- ▶ In fact, we are going through a period of very rapid innovation by many metrics—number of new widgets and apps; and more objectively, number of patents, especially from IT.



- ▶ And now (generative) AI.

Hopes of AI

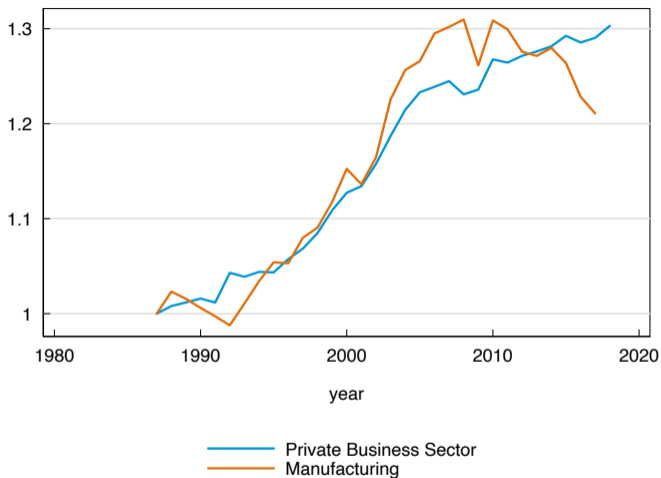
- ▶ Major advances in certain aspects of AI—especially those relying on machine learning tools applied to unstructured, huge data sets, and narrow, well-defined tasks, including quite complex ones such as chess and Go.
- ▶ Recent research showing that AI-related publications have shifted from the confines of computer science journals to a much broader range of application domains, indicating the onset of the more applied stage of the technology's advances.
- ▶ A lot of optimism. *The Economist* magazine recently argued that fears of job losses from AI are exaggerated and

“by lowering costs of production, [AI-based] automation can create more demand for goods and services, boosting jobs that are hard to automate. The economy may need fewer checkout attendance at supermarkets, but more massage therapists.”

- ▶ McKinsey-Davos statement for 2022 also dismisses fears of automation from AI:

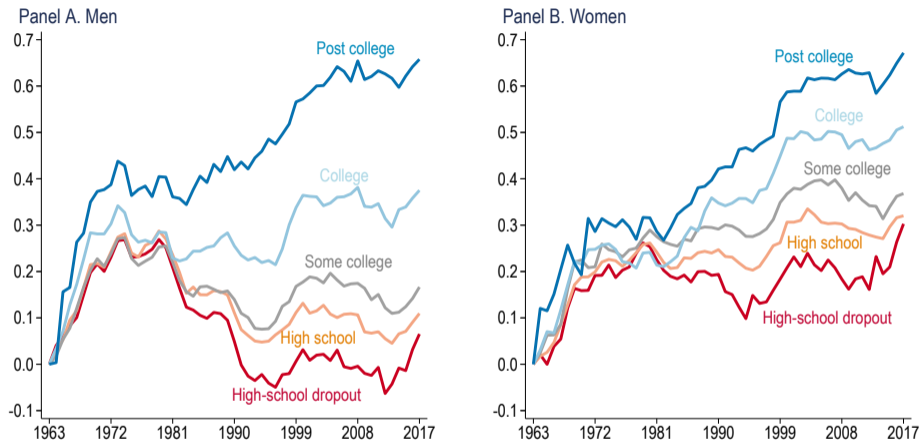
“with Fourth Industrial Revolution technologies driving productivity and growth across manufacturing and production at brownfield and greenfield sites. These technologies are creating more and different jobs that are transforming manufacturing and helping to build fulfilling, rewarding, and sustainable careers.”

But Obviously Not All is Well: US Total Factor Productivity



- ▶ Anemic productivity growth during the age of spellbinding innovations.
- ▶ Some claim this is mismeasurement, but no evidence that this problem explains the disappointing productivity performance or is worse than before.

But Obviously Not All is Well: US Wages



Source: Autor (AEA 2019)

- ▶ Huge increase in inequality, and significant declines in real wages for low-education groups.

What is Going on?

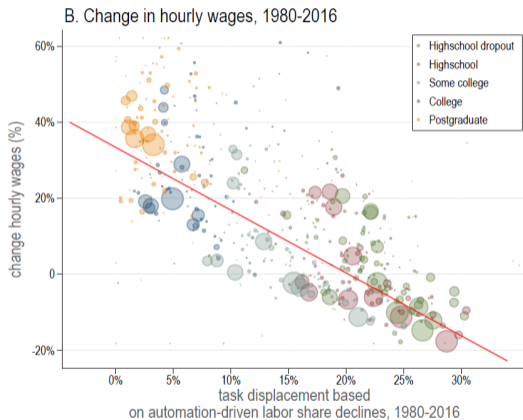
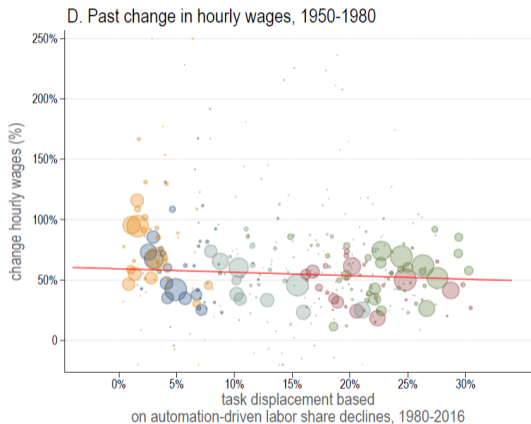
- ▶ My proposed explanation:
 - ▶ There are tremendous innovation opportunities based on digital technologies, advanced algorithms and new information gathering and sharing instruments.
 - ▶ But they are being squandered, because of an excessive focus on **automation**.
 - ▶ Automation—the substitution of machines for tasks previously performed by humans—can generate major efficiency gains, but is not the only source of productivity growth.
 - ▶ It is even more important to increase the (marginal) productivity of workers—**treat workers as a key resource, not a cost to be cut**.
 - ▶ Particularly important is introduction of **new tasks** for workers with diverse skills.
 - ▶ Excessive automation is boosting inequality and failing to realize the productivity benefits that are in principle possible.
 - ▶ This is a trend that started long before AI, but is accelerating with AI.
- ▶ Let us first understand the potentials and pitfalls of AI.

What Can AI Do?

- ▶ Automating tasks: very much in line with the vision much of the AI-CS community, going back to Alan Turing—e.g., reaching human parity.
- ▶ Task automation can be useful, but only if it increases productivity by a lot (whereas a lot of automation in the past have been "so-so") and also crucially, if it is coupled with new things—**new tasks** and better human decision-making.
- ▶ This is what some of the early pioneers like Norbert Wiener, Douglas Engelbart and JCR Licklider understood.
- ▶ Licklider advocated "human-machine symbiosis" as a way of complementing human capabilities, while Engelbart suggested "augmenting human intellect".
- ▶ More broadly, we should be focusing on **machine usefulness**—rather than machine intelligence. (Calculators and Wikipedia are not "intelligent" but very useful).
- ▶ Significant progress in this direction over the last six decades, some of it leading to huge breakthroughs in computer technology, including the mouse, graphic user interface, hyperlinks and the World Wide Web, etc.
- ▶ Also significant new platforms enabled by AI, such as Airbnb and ride-sharing apps.
- ▶ However, much of recent AI activity is focused on automation.

Pre-AI Automation and Inequality

- ▶ But digital technologies have prioritized automation over the last 45 years.

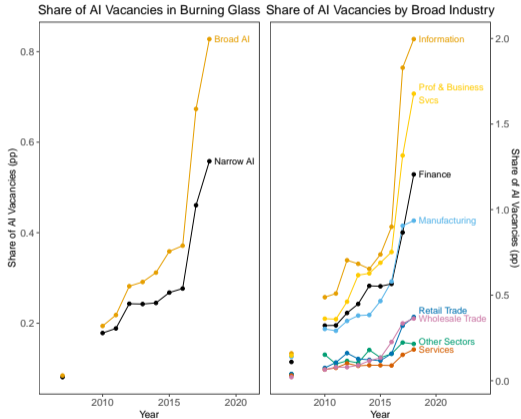


Source: Acemoglu and Restrepo (Emet 2020)

- ▶ Across 500 demographic groups (distinguished by education, gender, age, ethnicity and foreign/domestic), **task displacement** from 1980 onwards explains between 50-70% of

AI is (Almost) Here

- ▶ Measure (pre-LLM) AI from its footprints in job postings.
- ▶ Huge increase in AI since 2015, across most sectors of the economy.

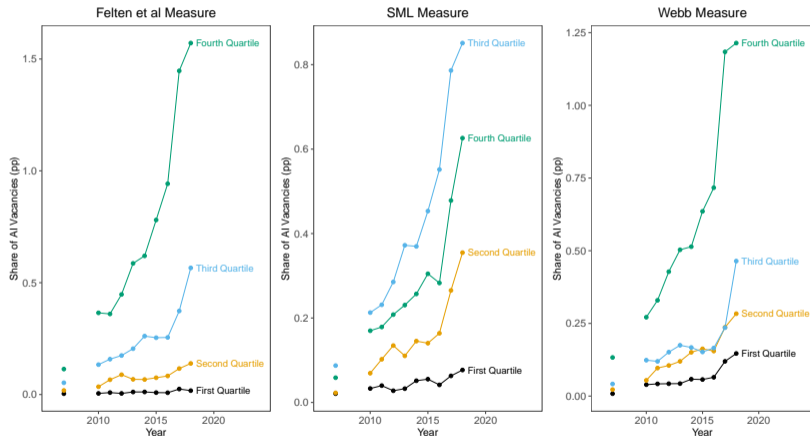


Source: Acemoglu, Autor, Hazell and Restrepo (J. of Labor Econ. 2022)

- ▶ Narrow AI vacancies up from 0.1% to 0.6%.

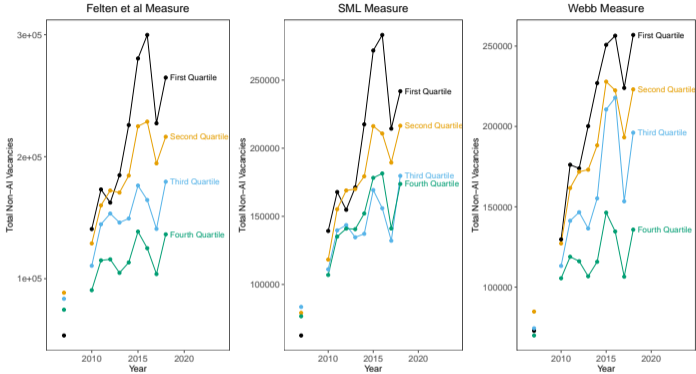
Establishment Share of AI Vacancies by Quartile of AI Exposure

- ▶ AI surge driven by establishments with more AI-replaceable tasks.



Source: Acemoglu, Autor, Hazell and Restrepo (J. of Labor Econ. 2022)

AI Negatively Associated with Establishment Hiring

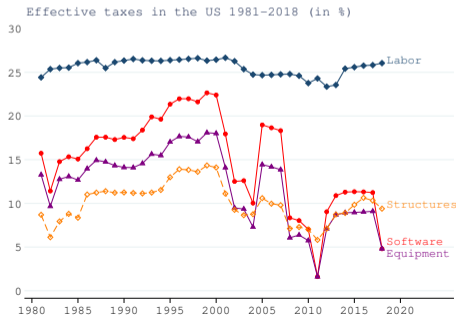


Source: Acemoglu, Autor, Hazell and Restrepo (J. of Labor Econ. 2022)

- ▶ This pattern is robust and quantitatively large: AI adoption, at the moment, going hand-in-hand with reduce hiring.

Could This be Excessive Automation?

- ▶ Excessive automation may also be fueled by additional factors:
 1. Business models and growing size of Big Tech (which is largely inconsistent with machine usefulness).
 2. Excessive focus on cost-cutting (and recently on monitoring).
 3. Massive data collection.
 4. Changing nature of government support for research (working much more to support established corporate priorities).
 5. The US tax code favors capital and powerfully encourages excessive automation.



The World of Large Language Models

- ▶ LLMs have arrived, almost suddenly, upon humanity, and are here to stay.
- ▶ Tremendous potential.
- ▶ But a major problem (and barrier to potential): the vision and architecture of LLMs continuing from the earlier automation and huge data collection focus.
- ▶ This translates into four related architectural problems, which have to be overcome if these models are going to go in the machine usefulness direction.
 1. **Illegibility**: It is impossible to understand what LLMs are doing and why they are making the recommendations that they are making. This problem is more general than “explainability”. It makes advancing human cognition and better decision-making all but impossible. It also encourages bad feedback loops between humans and machines.
 2. **Excessive authoritativeness**: As a consequence, LLMs become excessively authoritative—providing ill-fitting recommendations with great certitude.
 3. **Excessive centralization**: Huge data collection and illegibility also leads to excessive centralization of power in the hands of whoever controls base models. E.g.: what is allowed speech? What is correct information? Who decides those?
 4. **Excessive automation**: If humans cannot be complemented, then automation becomes more likely.

Need for New Architecture

- ▶ If this diagnosis is right, we can get much more from LLMs, but we need a course correction—both in the architecture of LLMs and in the apps building on them.
- ▶ Let us consider three examples where generative AI can be hugely useful in a human-complementary way:
 1. Education.
 2. Healthcare delivery.
 3. Electricians.
- ▶ What do we need from LLMs to make new tasks and human complementarity possible?
 - ▶ incorporate reliability scores (together with better filtering and verification);
 - ▶ allow reasoning exchanges, sensitivity analysis and broader interrogation by humans;
 - ▶ structure that facilitates human-complementary applications;
 - ▶ more selective use of (higher-quality) data;
 - ▶ internal guardrails, perhaps with two LLMs consistently checking each other (to prevent venturing into excessive authoritativeness);
 - ▶ internal structure to facilitate regulation.
- ▶ Some of these may be almost impossible with the current architecture.
- ▶ But more broadly, also a new vision of what we should be using LLMs for (together with new ethical precepts in the business and academic AI communities).

Implications for the Developing World

- ▶ This has implications not just for industrialized nations.
- ▶ LLMs will spread to the emerging world (and have already started doing so).
- ▶ Even more importantly, they will change the international division of labor: deindustrialization in the South from automation in the North.
- ▶ But automation technologies are **inappropriate technologies** for the developing world and new AI technologies may go in the same direction—they economize on the factors that are abundant in the developing world: labor, especially semi-skilled labor.
- ▶ They will increase inequality between the North and the South, as well as within the emerging world (Acemoglu and Zilibotti, 2001).
- ▶ Future of work in the developed world is thus intimately linked to future of growth in the developing world.

Implications for the Future of Work and Life

- ▶ There are really two faces of AI.
- ▶ **Good AI**—high-productivity automation technology needs to be accompanied with **new tasks** and broader human empowerment—can contribute to productivity and labor demand.
- ▶ **Bad AI**—leading to excessive centralization, massive data collection, **so-so automation**, and inequality, and perhaps even failing to deliver in terms of productivity gains.
- ▶ The problem is particularly severe when automation is not counterbalanced by new tasks, but also exacerbated by current business models and data collection modes.
- ▶ If the future is one of ceaseless algorithmic automation and nothing else, then the future of work will not be bright. There would be lower and lower labor share across industries and in national income. And there would be no guarantee of sufficient job growth.
- ▶ Improving labor market institutions, by itself, cannot be the solution — if we push wages up, this will cause more automation, unless technology becomes more “human-friendly”.
- ▶ But good AI, especially enabling the creation of new tasks for workers and empowering workers and citizens, can be powerful engine of growth, prosperity and human flourishing.

Conclusion

- ▶ We are not at the end of innovation.
- ▶ On the contrary, we may be at the brink of a new age of very different and potentially transformative innovations related to more powerful chips, more advanced algorithms and more capable ways of dealing with vast amounts of information.
- ▶ But hopes that these technological trends will **automatically** bring shared prosperity are not just wrong, they are deeply misleading.
- ▶ The **direction of technology** matters and matters greatly.
- ▶ The real problem facing us today is that we have a highly distorted direction of technology, led by gargantuan companies, distorted incentives and dangerous technological visions.
- ▶ The good news is that redirecting technology is possible.
- ▶ The bad news is that doing so in the current episode appears very challenging.