Can we Have Pro-Worker AI?

Choosing a path of machines in service of minds

Daron Acemoglu
David Autor
Simon Johnson

Revised: September 19, 2023
Summary

- Over the past 40 years, the diffusion of digital technologies significantly increased income inequality.
- Generative Artificial Intelligence (AI) will surely impact inequality, but the nature of that effect depends on exactly how this technology is developed and applied. Nothing about the path of this (or any) technology is inevitable.
- The private sector is pursuing a path for generative AI that emphasizes automation and the displacement of labor, along with intrusive workplace surveillance.
- Simply displacing workers is never good for the labor market, even when the displaced are highly paid. Displaced formerly high-paid workers are forced to compete for jobs with lower-wage workers, leading to a downward cascade in wage levels.
- A better path is available, along which generative AI would be complementary to most humans—augmenting their capabilities—including people without a four-year college degree.
- Choosing the human-complementary path is feasible but will require changes in the direction of technological innovation, as well as in corporate norms and behavior.
- The goal should be to deploy generative AI to create and support new occupational tasks and new capabilities for workers. If AI tools can enable teachers, nurse practitioners, nurses, medical technicians, electricians, plumbers, and other modern craft workers to do more expert work, this can reduce inequality, raise productivity, and boost pay by leveling workers up.
- Public policy has a central role in encouraging this positive path of technology to complement all workers, elevating the achievable level of skill and expertise for everyone.
- At this time, the five most important federal policies should be:

  1. Equalize tax rates on employing workers and on owning equipment/algorithms to level the playing field between people and machines.
  2. Update Occupational Safety and Health Administration rules to create safeguards (i.e., limitations) on the surveillance of workers. Finding ways to elevate worker voice on the direction of development could also be helpful.
  3. Increase funding for human-complementary technology research, recognizing that this is not currently a private sector priority.
  4. Create an AI center of expertise within the government, to help share knowledge among regulators and other officials.
  5. Use that federal expertise to advise on whether purported human-complementary technology is appropriate to adopt in publicly provided education and healthcare programs, including at the state and local level.
Introduction

The world is about to experience transformative and disruptive advances in generative Artificial Intelligence. A major set of concerns centers around the labor market and economic inequality implications of these advances. Will AI eliminate jobs in net? Will it further inflame the decades-long phenomenon of rising economic inequality? Will it boost labor earnings or instead make machines more valuable and workers more expendable?

The consensus in the economic literature is that previous waves of digital technologies—including personal computers, numerically controlled machinery, robotics, and office automation—have increased inequality. This is both because some of these technologies, such as personal computers, have been highly complementary to more-educated workers (Autor, Katz and Krueger, 1998; Autor, Levy and Murnane, 2003; Goldin and Katz, 2008), and also because many of these tools have been used for automating work, with an unequal impact on different types of workers (Autor, Levy and Murnane, 2003; Acemoglu and Restrepo, 2022a, 2022b). While digital technologies have undoubtedly created new goods/services and boosted productivity in some activities (e.g., Brynjolfsson and McAfee, 2015), there is also evidence that productivity gains from these technologies have sometimes fallen well below expectations (e.g., Acemoglu, Autor, Dorn, Hanson, and Price, 2016).

Generative AI will have a substantial impact on the future of work and the trajectory of inequality. The nature of that impact is not an inevitable consequence of the technology itself but instead depends on how society develops and shapes AI.

- The currently predominant direction for AI emphasizes automation, displacement of skilled labor, and diminished worker voice due to stepped-up monitoring and surveillance.
- An alternative, “human-complementary” path could contribute more to productivity growth and could help reduce economic inequality.

In the next section, we outline what the automation path looks like and what its implications would be for work, inequality, and productivity. We then describe the alternative human-complementary path, drawing on both general principles and specific examples. We also explain why, despite its advantages, the human-complementary approach is not likely to prevail based on current investments and
corporate attitudes. We suggest policies that could help steer AI development and implementation in the more constructive direction.

**The Automation Path**

Automation—the substitution of machines and, more recently, algorithms for tasks previously performed by humans—has been a constant since at least the beginning of the Industrial Revolution. Machines were developed to perform tasks that have a high degree of predictability and are carried out in stable environments. Examples include spinning and weaving in the eighteenth-century textile industry, harvesting in nineteenth-century agriculture, and many office and clerical tasks in the twentieth century, such as telephone switchboard operation and routine bookkeeping. Mass production, which vastly reduced the cost of everyday products, depends fundamentally—though not exclusively—on assembly lines made possible by automation.

Not all automation is highly productive, however. When machines are deployed to perform tasks in which they are not particularly effective, lackluster productivity gains follow. Most people are familiar with the frustration of seeking customer service from an airline, credit card provider, or computer manufacturer, only to be diverted through mazes of unhelpful computer menus. Firms may find such automation to be cost-effective, but it is not a meaningful productivity advance.

Whether it has large or small effects on productivity, automation tends to have major distributional consequences. The reason is that automation displaces workers who were specialized in the tasks that automation reallocates to machines and algorithms. The automation of blue-collar and office jobs using digital technologies has been an important driver of the rise in inequality since 1980 (Acemoglu and Restrepo, 2022a).

It is inevitable that AI systems will be used for some automation, both for technical and business strategy reasons. On the technical front, a major barrier to automation of many service and production tasks has been that they require flexibility, judgment, and common sense—things that are notably absent from pre-AI forms of automation. Artificial intelligence, especially generative AI, can potentially master such tasks (Susskind, 2021). A broad swath of computer security tasks that used to be performed by skilled human operators can now be performed by AI bots. Similarly, generative AI systems can write advertising copy, parse legal documents, transcribe physicians’ medical notes, and perform language translation. It is unclear how much this type of automation will contribute to aggregate productivity growth while these
technologies are immature, but they could contribute to sizable productivity gains as costs fall and reliability improves.

Businesses may choose machines over workers for reasons other than productivity. Automation appeals to managers who are seeking greater consistency and less opposition from organized or unorganized labor (Acemoglu and Johnson, 2023).

All too often, businesses prefer to focus on automation rather than creating new job tasks and enabling workers to build new skills. Automation is always an easy case for a manager to make because it appears to save costs. Investing to make workers more productive or more useful may be a harder sell, since it’s seen as messy, uncertain, and expensive. Some managers simply prefer to “hire machines” rather than hiring workers because machines don’t complain about pay or working conditions, and they certainly don’t join unions. But a country is not a business. We have a shared interest in ensuring that adults are productively employed. This promotes economic resilience, social cohesion, and a strong tax base. Policymakers care more about the quality and quantity of jobs than do employers, and policy should support institutions, incentives, and investments with this in mind.

Beyond economic incentives, the dominant intellectual paradigm in today’s digital tech sector—among both business leaders and academic researchers—favors the automation path. A major focus of AI research is to attain human parity in a vast range of cognitive tasks and, more generally, to achieve “artificial general intelligence” that fully mimics and then surpasses capabilities of the human mind. This intellectual focus encourages automation rather than the development of human-complementary technologies (Acemoglu and Johnson, 2023).

There is a widely shared optimism that human-replacing automation will produce so much productivity gain that all types of workers will benefit. It is true that if automation is sufficiently productive (and thus reduces costs by a significant amount), this can generate demand for other goods and services and, as a result, workers may benefit as well.

However, while productivity gains are obviously welcome, there are two problems with this line of thought. First, the benefits may be highly unequally distributed across different skill groups. For example, AI-based productivity gains might increase the demand for so-called “prompt engineers,” but this would not help workers displaced from accounting or financial services jobs—assuming they do not have a comparative advantage in prompt engineering (Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2022a). Second, automation tends to reduce the labor share of national income, so even if workers benefit, most of the gains flow to entrepreneurs and capital owners.
(Acemoglu and Restrepo, 2018). There are limits to how much shared prosperity can be generated exclusively by automation.

Another common prediction is that, because generative AI may automate managerial or knowledge tasks typically performed by professional workers, it could have equalizing effects. For instance, if accountants and financial analysts lose their jobs, this might reduce inequality between retail workers and highly paid financial sector workers. This logic is faulty. Studies of previous waves of digital automation show that workers directly displaced by new technologies not only experience lower pay growth but also start competing with other groups of lower-paid workers, whose pay then declines (e.g., Acemoglu and Restrepo, 2022a). Simply put, displacing workers will never be good for workers or for the labor market. Instead, AI can reduce inequality if it enables lower-ranked workers to perform more valuable work—but not if it merely knocks rungs out of the existing job ladder.

The Human Complementary Path

New technologies need not merely replace workers in existing tasks. They may also complement workers by enabling them to work more efficiently, perform higher-quality work, or accomplish new tasks that were previously infeasible (Acemoglu and Restrepo, 2018; Autor, Chin, Salomons, and Seegmiller, 2022; Acemoglu and Johnson, 2023). For example, even as mechanization gradually pushed more than half of the U.S. labor force out of agriculture, a range of new blue-collar and clerical tasks in factories and newly emerging service industries generated significant demand for skilled labor. The expansion of employment in industry and services between (roughly) the years 1870 and 1970 led to work that was not only better paid but also less dangerous and less physically exhausting, and increasingly rewarded the formal literacy and numeracy skills created by the expansion of universal public high school education.

This virtuous combination—automation of traditional work alongside creation of new tasks—proceeded in relative balance for much of the twentieth century. But sometime after approximately 1970, this balance was lost. While automation has maintained its pace or even accelerated over the ensuing five decades, the offsetting force of new task creation has slowed, particularly for workers without four-year college degrees (Acemoglu and Restrepo, 2019; Autor, Chin, Salomons, and Seegmiller, 2022). Non-college workers have been displaced from factories and offices by computerization and, for blue-collar workers, also by import competition (Autor, Dorn, and Hanson, 2013), but no new equivalently well-paid opportunities have emerged to attract these
workers. As a result, non-college educated workers are increasingly found in low-paid services such as cleaning, security, food service, recreation, and entertainment. These jobs are socially valuable, but they require little specialized education, training, or expertise, and hence pay poorly.

The critical question we face in the new era of generative AI is whether this technology will primarily accelerate the existing trend of automation without the offsetting force of good job creation—particularly for non-college workers—or whether it will instead enable the introduction of new labor-complementary tasks for workers with diverse skillsets and a wide range of educational backgrounds.

There is a case for qualified optimism: generative AI offers an opportunity to complement worker skill and expertise.

Because so many of the routine tasks that workers previously performed have already been automated, a large fraction of current jobs require non-routine problem-solving and decision-making tasks. Empowering workers to perform these tasks more effectively, and to accomplish even more sophisticated decision-making tasks, will require providing workers with better information and decision-support tools. Generative AI is particularly well-suited to this type of information provision. An irony of today’s information era is that people are overwhelmed by information but often lack the time and expertise to parse this information effectively. Generative AI is particularly well-suited to address this problem. With proper development, AI tools can help surface relevant information at the right time to enable better decision-making.

Additionally, and closely related, human productivity is often hampered by lack of specific knowledge or expertise, which could be readily supplemented by AI. For example, an electrician might be unable to diagnose a rare problem that she could readily address if given relevant tools or appropriate training. Or a highly trained immigrant to the U.S. may be inhibited from fully using her abilities because of limited English language skills. Generative AI tools can assist in such cases by boosting human expertise, supporting workers in unfamiliar situations, providing on-the-spot training, and improving all forms of information translation. Overall, AI holds great potential for training and retraining expert workers, such as educators, medical personnel, software developers, and other workers with modern “crafts” (such as electricians and plumbers).

Finally, while generative AI may take over more of the operational tasks in certain occupations, such as accounting, financial analysis, or computer programming, if developed in the right manner, it could create new demands for human expertise and
judgment in overseeing these processes, communicating with customers, and enabling more sophisticated services that leverage these tools.

Several recent studies provide “proof-of-concept” examples that demonstrate how generative AI can supplement expertise rather than displace experts. Peng, Kalliamvakou, Cihon, and Demirer (2023) demonstrate that GitHub Copilot, a generative AI-based programming aid, can significantly increase programmer productivity. In a controlled experiment, the treatment group that was given access to this generative AI tool completed the required programming task about 56% faster than the control group without access to Copilot.

Noy and Zhang (2023) performed a related online randomized controlled trial, focused on writing tasks. Among the set of white-collar workers recruited for the study, half were randomly given access to (and encouraged to use) ChatGPT for writing tasks. Noy and Zhang (2023) found significant improvements in the speed and quality of writing output. Most importantly, the biggest improvements were concentrated among the least-capable writers. Although generative AI did not make the least-skilled writers as effective as the most-skilled writers, it made all writers faster and substantially reduced the quality gap between the two groups.

Finally, Brynjolfsson, Li, and Raymond (2023) evaluated the use of generative AI tools used for providing background information to customer service agents. They also estimated a significant improvement (about 14%) in productivity, and like Noy and Zhang's study, these gains were the most pronounced for novice workers. Using these AI tools, novice workers were able to reach a level of proficiency within three months that previously took workers ten months to attain.

In all three cases, generative AI tools automate and augment human work simultaneously. The automation stems from time-savings: AI writes the first draft of computer code, advertising copy, and customer support responses. Augmentation happens because workers are called upon to apply expertise and judgment to intermediate between the AI's suggestions and the final product—whether it is software, text, or customer support.

**Promising Applications**

Looking forward, we see at least three major sectors where human-complementary AI could be transformative, both for productivity and for shared prosperity.
Education

Generative AI tools can enable major advancements in education, together with new productivity-enhancing roles for educators. Classroom instruction is hindered by the fact that the teacher must choose one pace at which an entire class proceeds, even if it is too fast for some students and too slow for others. Individualized education programs and personalized teaching tools can be effective in enabling less prepared students to excel, but these tools are labor-intensive and hence expensive. AI-enabled tools have the potential to vastly improve tutoring and self-instruction. Khanmigo, an app built on ChatGPT-4, is an infinitely patient, highly adaptable tutor that can break complex problems into their constituent parts, walk students step-by-step through solving them, and provide hints and examples along the way without directly answering questions.\textsuperscript{3} Research shows that Large Language Models (LLMs) can anticipate which parts of a problem humans will find difficult and suggest simplifications to improve understanding (see Bubeck et al., 2023).

These technologies can support educators as well as students. Teachers could focus more of their time on instruction and less on remediation. They could develop richer lesson plans that harness new tools, such as visualization, simulation, and even real-time “interaction” with fictitious or historical figures. The right AI investments could enable significant gains in student learning, especially among currently underperforming pupils. Such reorganization would also create a range of new tasks for educators—and would almost certainly necessitate further productive training for existing teachers and hiring of additional AI-savvy teachers.

It is not clear, however, that the private sector (or any public school district) has the right incentives to develop such AI tools. Will AI be developed to reduce gaps between more-capable and less-capable learners—as ChatGPT does for professional writers—or will it instead be deployed to reduce teacher headcounts and substitute for personal attention? This is not a question of what the technology can do but how we collectively decide to develop and deploy it. Applications of AI that appear to reduce labor costs may be more appealing to many cash-strapped school districts in the United States. Thus, our concern is that AI will be used to automate teaching, testing, and grading, rather than for new personalized education-targeting tasks.

Healthcare

Given that one in five U.S. dollars is spent on healthcare, any technology that improves efficiency, lowers costs, or broadens access to the healthcare system has potentially enormous benefits. Generative AI tools can improve healthcare delivery
and accessibility, enabling productivity gains and generating valuable new worker tasks. For example, generative AI tools could support expanding scope of practice boundaries, enabling medical professionals at all levels to accomplish a broader range of tasks. Just as nurse practitioners have proved effective at diagnosing, prescribing, and treating (Asubonteng, McCleary, and Munchus, 1995; Li, Westbrook, Callen, and Georgiou, 2012)—tasks formerly done exclusively by doctors—the decision-support capacities of AI can enable a larger set of trained medical professionals to accomplish expert tasks without exclusively relying on the most elite medical professionals. Using AI, qualified nurse practitioners, nurses, and health technicians could diagnose routine health problems, recommend courses of treatment, and more efficiently route patients to further care options.

**Modern Craft Workers (MCWs)**

In line with the electrician example above, generative AI tools can be transformative for modern craft workers (MCWs) more broadly. The U.S. is currently undertaking a major infrastructure investment agenda, with growing employment in manufacturing, green energy production, and chip production, among other sectors. Skilled MCWs are in short supply due to an aging population and decades of over-investment in higher education at the expense of valuable vocational training. AI can be used to support training and enable MCWs to carry out a wider range of tasks that require specialized expertise. AI can provide relevant information, real-time instruction, and decision-making support in electrical work, plumbing, expert repair, design, and construction, among other activities. The current generation of AI technologies cannot replace the work of MCWs—whose tasks require dexterity, flexibility, and judgment that are far beyond the grasp of current robotics (even AI-enhanced robotics). But AI can enable these workers to do more with their skills by solving a broader and deeper range of applied problems in the field.

**What Can Be Done**

**Tax system:** The current U.S. tax code places a heavier burden on firms that hire labor than those that invest in algorithms to automate work (Acemoglu, Manera, and Restrepo, 2020). We should aim to create a more symmetric tax structure, where marginal taxes for hiring (and training) labor and for investing in equipment/software are equated. This will shift incentives toward human-complementary technological choices by reducing the bias of the tax code toward physical capital over human capital.
**Labor voice:** The direction of AI will have profound consequences for all workers. Creating an institutional framework in which workers also have a voice would be a constructive step—and there is an important role for civil society in pressing for this to happen, including through articulating needs at the local and state level. At a minimum, federal government policy should restrict deployment of untested (or insufficiently tested) AI for applications that could put workers at risk—for example, in high-stakes personnel decision-making tasks (including hiring and termination) or in workplace monitoring and surveillance. Health and safety rules need to be updated accordingly.

**Funding for more human-complementary research:** Given that the current path of research has a bias toward automation, additional support for the research and development of human-complementary AI technologies could have significant impact. It is hard to target human-complementary work in the abstract. It is feasible, however, to focus on specific sectors and activities where opportunities are already abundant. These include education, healthcare, and MCW training. Just as DARPA orchestrated investments and competitions to foster the development of self-driving cars and dexterous robotics, the federal government should foster competition and investment that pairs AI tools with human expertise, aiming to improve work in vital social sectors.

**AI expertise within the federal government:** AI will touch every area of government investment, regulation, and oversight, including (but not limited to): transportation, energy production, labor conditions, healthcare, education, environmental protection, public safety, and military capabilities. Developing a consultative AI division within the federal government that can support the many agencies, including regulators, tackling these challenges will support more timely and more effective decision-making at every level.

**Technology certification:** The federal government can encourage appropriate investments by advising on whether purported human-complementary technology is of sufficient quality to be adopted in publicly funded education and healthcare programs. For this advice to be meaningful, experts need to be engaged and independent—i.e., they should not be directly or indirectly working for the tech companies. It is hard to attract talent to government or universities when the private sector is paying top dollar for the relevant expertise. This further strengthens the case for building a high-prestige, cross-cutting federal AI service.
Conclusion

There is no guarantee that the transformative capabilities of generative AI will be used for the betterment of work or workers. The bias of the tax code, of the private sector generally, and of the technology sector specifically, leans towards automation over augmentation. But there are also potentially powerful AI-based tools that can be used to create new tasks, boosting expertise and productivity across a range of skills.

To redirect AI development onto the human-complementary path requires changes in the direction of technological innovation, as well as in corporate norms and behavior. This needs to be backed up by the right priorities at the federal level and a broader public understanding of the stakes and the available choices. We know this is a tall order. But this makes focusing on what is needed even more important.


1 The authors are co-directors of the MIT Shaping the Future of Work Initiative, which was established through a generous gift from the Hewlett Foundation. This research program and related results were also made possible with the support of the NOMIS Foundation. Acemoglu: Institute Professor, MIT; Autor: Ford Professor of Economics, MIT Department of Economics; Johnson: Kurtz Professor of Entrepreneurship, MIT Sloan School of Management. Relevant disclosures are available at shapingwork.mit.edu/power-and-progress, under “Policy Summary.”

2 https://github.com/features/copilot

3 https://www.khanacademy.org/khan-labs#khanmigo

4 Also known as “modern trades” or “tradespeople.”