


Defining our future with generative AI

Siddharth Suri

 Check for updates

We can design, build and use AI systems with intentionality, to make them an equalizing force within society, or we can use AI without intentionality, in which case AI could become a force that exacerbates inequality, or both. Society has the power to decide which.

The recent and upcoming advances in artificial intelligence (AI) represent a phase transition in the ability of such systems to solve problems previously thought to be intractable. Given this huge technological leap forward, now is when we, as a global society, must define the trajectory of our future. As companies continue to innovate AI systems and integrate them into current products, it is our responsibility to ask ourselves: what is the future that we want to build? As a society, we must take a stance and define the relationship that we want between people and AI systems. We are still in the early stages of the AI revolution, so it is easier to set our trajectory on a conscientious path now than it would be to correct our course later. We can either use AI with intentionality to make it an equalizing force within society, or we can use AI without intentionality, and it can become a force that exacerbates inequality, or both. Society has the power to decide which of these outcomes we drive towards.

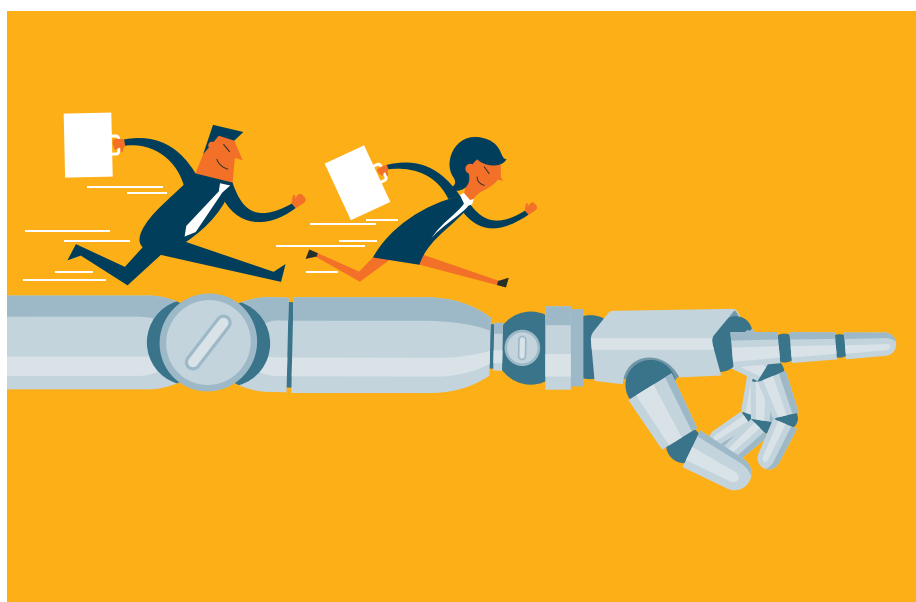
A potentially equalizing force

Within generative AI, large language models (LLMs) are responsible for much recent progress. A recent study¹ shows that most of the work

being done on LLMs is knowledge work, which is any work involving handling or using information. There is also a string of recent experimental papers showing that AI improved the productivity of participants in various knowledge work contexts, such as programming², writing³, consulting⁴ and customer support⁵. Importantly, these papers show that there is an economic benefit to using generative AI. Notably, the boost in productivity was not uniformly distributed among the participants. In fact, AI helped those who needed it the most, more. That is, generative AI helped novice and lower-skill workers more than it helped the more experienced and skilled workers. To the extent that productivity is correlated with earnings, generative AI allows the less experienced or skilled workers to close the gap. If these results generalize, the results show that within the field of knowledge work, AI is acting as an equalizing force.

Inequality in accessing the technology

Whenever there is a technological innovation on the scale of generative AI, it is important to first understand who has access to that innovation, because only those exposed to AI can use it to accrue productivity and economic benefits. A recent Pew survey shows that Americans that know about ChatGPT are more likely to have higher household incomes and more formal education. For example, 79% of adults with postgraduate degrees have heard of ChatGPT, whereas only 41% of those with high school education or less have heard of it. Similarly, 76% of those in the upper-income tier (family income over US\$131,500) have heard of ChatGPT versus only 44% of those in the lower-income tier (family income less than US\$43,800). Finally, men are more likely to have heard of ChatGPT than women (67% versus 49%). According to another Pew survey, of those who have heard of ChatGPT, 33% of those



with a postgraduate education have used it versus only 15% of those with a high school education or less, and men are more likely to have used it than women (29% versus 19%).

There is already substantial inequality across education, income and gender in the USA⁶. For example, on average, women in the USA still only earn **82 cents for every dollar** earned by a man. Worldwide, economic inequality is such an important problem that it is one of the seventeen **United Nations Development Goals**. Looking at the labor force, given that generative AI could provide an economic benefit and there is disparity in who is aware of it and who is using it, there is a concern that AI could exacerbate existing economic inequality. To counteract this effect, those building AI systems should intentionally target those who are not as likely to be using AI, helping to reduce some of these inequalities by understanding their use cases and tailoring AI systems to deliver value to these populations. Whether or not through community-based participatory design, engineers working on generative AI should collaborate with members of the communities that they are not currently serving to understand how AI might provide value to those communities.

Impact in the workforce

AI can also profoundly affect tasks, work and jobs. At any moment, there is a set of problems that machines can solve. Any problem outside that set needs to be solved, at least in part, by humans. When there is a technological innovation, like generative AI, the set of problems that machines can solve grows, encompassing some problems that humans used to solve. Thus, some of the humans who used to do the newly encompassed tasks get displaced and a new frontier of problems to be solved emerges. We called this idea “the paradox of automation’s last mile”⁷, but the idea is not new. For instance, economists assume that a job is comprised of tasks. When there is a technological innovation, some of those tasks can become automated. When the tasks that are automated are core to a job, that job will be dramatically reshaped. When the tasks that are automated are peripheral to a job, the impact will be smaller^{8,9}. Nonetheless, jobs do get reshaped, and the dynamic of the human–AI relationship shifts. This paradox is one of the fundamental components of our relationship with AI.

The first thing to understand about the paradox of automation’s last mile is that, overall, we are not going to run out of work any time soon. In fact, this dynamic is responsible for creating most of our current jobs. According to **David Autor**, “the majority of contemporary jobs are not remnants of historical occupations that have so far escaped automation. Instead, they are new job specialties that are inextricably linked to specific technological innovation.” The second thing to understand is that the people whose jobs are affected by automation need not be the people for whom new work is created. For example, if automation affects workers’ jobs in one city but creates jobs in another, those with more resources can more easily move, upskill or reskill, thereby adapting to the new economy, whereas those with fewer resources are less able to do so⁶. As the dynamic between humans and AI shifts, it’s essential to remember that people’s occupations are often intertwined with their own personal identities and self-worth⁶. How we treat the people whose jobs are affected by AI will be a key determining factor in what kind of future we end up with. In the past, automation has been the cause of roughly half of the loss in wages of many middle-class jobs in the USA⁹ and we have the opportunity to avoid that outcome this time around.

There are at least two ways we could define a more positive relationship with AI going forward. First, an example of this paradox, and an

illustration of how to protect those whose jobs are affected by AI, can be found in Hollywood. The task of generating movie and TV scripts used to be a problem that only humans could solve, but now LLMs can generate text and dialog for fiction writing. In response, the Writers Guild of America (WGA) struck an **agreement** with the Alliance of Motion Picture and Television Producers to ensure that “a writer can choose to use AI when performing writing services... but the company can’t require the writer to use AI software (e.g., ChatGPT) when performing writing services”. This is just one of many protections for writers in the agreement, along with protections that prohibit companies from using the output of screen writers to train AI systems or crediting AI for material. The agreement intentionally specifies the relationship that the writers want with AI systems. These protections empower writers to use AI, as opposed to being replaced by AI. Such a labor agreement is important because, although it protects all writers in the guild, those with fewer resources are often affected disproportionately by automation. More generally, the WGA agreement could provide a blueprint for future labor groups defining their relationship with AI. Having these agreements in place, in a variety of industries, could pre-emptively help those who would otherwise be displaced by AI by specifying productive relationships between workers in that industry and AI. One lesson learned from this example is that it takes a variety of institutions (such as technology companies, writers, studios, unions and so on) to come together to guide society towards a productive path.

Second, in deciding what positive human–AI relationships look like, we must also consider the types of role that each party could take on. Hofman and colleagues specify that AI can act in three different roles: (1) it could be a coach that improves our capabilities, (2) it could act like sneakers in that they augment our skills and help us get things done, or (3) it could act like a steroid on which we can rely in the short term, but that will weaken us in the long term¹⁰. If we intentionally design AIs that act as coaches, generative AI can step in when the user does not have the necessary expertise or access to the necessary expertise, and it might even help us to do our jobs better. In fact, if we apply the ‘AI as a coach’ idea across jobs and sectors, Autor suggests that AI could help “restore the quality, stature and agency that has been lost to too many workers and jobs” by extending “the relevance, reach and value of human expertise for a larger set of workers”. Since previous types of automation disproportionately affected middle-class and working-class jobs, there is a chance for AI to help those affected by previous types of automation, potentially mitigating previous inequities.

The human labor behind generative AI

In this context, it is particularly important to consider that there is inequity among those who have the power or autonomy to determine their relationship with AI systems. For instance, many LLM users do not realize that the AI systems that they use were trained, in part, by humans. LLMs, especially the earlier versions, can output text that is incorrect, uninformative or, in some cases, toxic. To help to fine-tune these models, humans were given example prompts and asked to write outputs they would like to receive and also to rank different possible outputs of these models¹¹. In fact, humans have been providing training data to machine learning algorithms, doing what is termed ‘ghost work’, through online labor platforms like **Mechanical Turk**, **Upwork** and **Scale AI** for well over a decade⁷.

As AI systems expand in reach and popularity, this type of training will happen more often and take on greater importance. These workers are disproportionately at the lower end of the income distribution,

and often work in countries where a dollar buys more labor than it does in the USA. In addition, these workers occupy a position where technology companies are paying them (either directly or through a third-party platform) to provide training data for AI systems. Since these workers often depend on the money they earn for basic needs¹², the workers may not have the power to define their relationship with the AI systems they train, yet they provide a service for every member of society with internet access. Thus, AI companies should recognize and value their contributions and offer them fair and ethical work practices. One way forward would be for society to agree on the criteria for ethical treatment of those who train AI systems and certify AI systems with a 'fair trade' stamp.

For the workers who train AI systems, there are at least three ways forward. The first is to unionize or at least gather together, to have a collective voice about the important work they do and advocate for fair and equitable working conditions. If companies and workers can come to a mutual agreement, workers would get fair and equitable working conditions and companies could take comfort in knowing that they are treating the workers that train their AI systems ethically. The second is to change the laws surrounding fair and equitable working practices, so that workers who train AI systems can access the social safety net that is usually reserved for those with full-time jobs in the USA. Not only would this provide fairer work practices, but also improved quality of life. Finally, in online labor markets, there are more workers than there is work to be done¹³. Therefore, companies that offer work on a given platform could leverage the power that they have to compel the platform to ensure that the workers are treated fairly. Whatever the method we use to affect change, it is incumbent upon society to ensure that these workers receive fair and ethical treatment.

Inequality in developing the technology

It is important to consider not just who has access to these models and who provides data to train them, but also who can build these models to ensure that everyone has a voice in their design. The use of generative AI is happening globally, but it takes hundreds of millions of dollars' worth of resources, such as GPUs and electricity, to develop LLMs¹⁴. Furthermore, these models are trained with as much data as can be gathered from the internet, along with human-labelled data. The resources required to build and train these models may put them out of reach of most people and institutions, which, for example, would disproportionately exclude institutions in the Global South. Thus, many people and institutions are left out of the conversation as to how to define our relationship with the AI systems we build, despite the importance of their inclusion to achieve equitable progress.

A nudge in the right direction

We can only improve what we measure. All sectors of global society, including workers, NGOs, governments, and technology companies need infrastructure to measure whether society is on the right track with respect to AI. We should build a planetary dashboard to measure who is being affected by AI and how. Doing so would allow us to monitor our relationship with AI systems and nudge them in the right direction as needed. Nobody knows what future we're headed for. In fact, it's extremely difficult for even the best forecasters to predict more than two years out¹⁵. Rather than focus on where AI is taking society, let's focus on measuring the here and now and use that to guide our progress towards an outcome that we'll be satisfied with. In this way, wherever society ends up, we can be sure that we're happy with how we got there.

Siddharth Suri  

Microsoft Research, One Microsoft Way, Redmond, WA, USA.

 e-mail: suri@microsoft.com

Published online: 24 September 2024

References

1. Suri, S. et al. Preprint at <https://doi.org/10.48550/arXiv.2404.04268> (2024).
2. Peng, S., Kalliamvakou, E., Cihon, P. & Demirer, M. Preprint at <https://doi.org/10.48550/arXiv.2302.06590> (2023).
3. Noy, S. & W. Zhang, W. *Science* **381**, 187–192 (2023).
4. F. Dell'Acqua, E. et al. Navigating the jagged technological frontier: field experimental evidence of the effects of AI on knowledge worker productivity and quality. Working Paper 24-013 (Harvard Business School Technology & Operations Management Unit, 2023).
5. Brynjolfsson, E., Li, D. & Raymond, L. R. Generative AI at work. NBER Working Paper 31161 (National Bureau of Economic Research, 2023).
6. Case, A. & Deaton, A. *Deaths of Despair and the Future of Capitalism* (Princeton Univ. Press, 2020).
7. Gray, M. L. & Suri, S. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass* (Harper Collins, 2019).
8. Frey, C. & Osborne, M. *Technol. Forecast. Social Change* **114**, 254–280 (2017).
9. Acemoglu, D. & Restrepo, P. *Econometrica* **90**, 1973–2016 (2022).
10. Hofman, J. M., Goldstein, D. G. & Rothschild, D. M. A sports analogy for understanding different ways to use AI. *Harvard Business Review* (4 December 2023).
11. Ouyang, L. et al. Training language models to follow instructions with human feedback. In *36th Conference on Neural Information Processing Systems (NeurIPS, 2022)*.
12. Posch, L. et al. *Human Comput.* **9**, 22–57 (2022).
13. Kingsley, S. C., Gray, M. L. & Suri, S. *Policy Internet* **7**, 383–400 (2015).
14. Maslej, N. et al. *The AI Index 2024 Annual Report* (AI Index Steering Committee, Institute for Human-Centered AI, Stanford University 2024).
15. Tetlock, P. E. & Gardner, D. *Superforecasting: the Art and Science of Prediction* (Random House, 2015).

Acknowledgements

The author thanks S. Counts, D. Goldstein, J. Hofman, S. Jaffe, P. Yu, and especially M. Daepf, for helpful feedback.

Competing interests

The author is a full-time employee of and has a financial interest in Microsoft.