PROJECT SYNDICATE – 13 APRILE 2026

To Work for Us, AI Must Not Think for Us

di Dani Rodrik

The public discussion about AI's impact on society focuses largely on the potential displacement of workers and the loss of jobs. But an even greater risk is the displacement of human thought and the processes that produce the knowledge base on which AI models themselves rely.

CAMBRIDGE—Only a few years ago, AI seemed like merely a nice toy: a chatbot that simulated intelligence by assembling complete sentences in response to users' prompts but that ultimately wasn't much more sophisticated than an advanced search engine. Yet now, it has proven to be an incredible tool that can performs task I never thought would be possible in my lifetime.

For example, I have used AI to locate online datasets, manipulate them, carry out statistical tests, and produce polished tables and charts, complete with sensible commentary on what the results mean, how they relate to the academic literature, and the strengths and weaknesses of the analysis. In less than half an hour, AI can do a job that would take a research assistant several days.

Sometimes, the current crop of AI models seem almost capable of reading your mind. Unlike in programming or writing code, you do not have to specify very precisely what you are looking for, leaving no room for misinterpretation. The model will "intuit" what you are after and fill in the missing details (though you had better always check them, as law firms that have filed AI-generated briefs with fictional citations can attest). Or, barring that, the interface will prompt you until you have clarified your query.

It is comforting to think that AI could be a tool that will help us all become more productive and better at what we do. It has certainly made me more efficient at research.

It lowers entrepreneurs' costs by providing marketing and consulting services on the cheap. It allows junior customer-service agents to avail themselves of the skills and experience of more senior staff. And it enables gig workers or craftsmen to provide more sophisticated and technically demanding services.

Unlike many earlier technologies, AI is uniquely positioned to help those with fewer skills and less education—workers occupying the lower rungs of the economy. By endowing each one of us with greater capabilities, it offers advantages that are potentially most meaningful for those with the greatest initial disadvantages. That means it could function very differently from, say, automation, the main purpose of which is to replace workers on the assembly line or in sales/clerical work.

The worry, of course, is that AI will also do much more than that, with uncertain consequences. For now, I see choosing and framing research questions as my own prerogative, and the main source of my competitive advantage. But at some point, I can imagine feeling tempted to ask AI to generate the questions themselves. In fact, the AI tools I use are already nudging me to do that. At the end of an exercise of the kind I sketched above, they will gently suggest further avenues of fruitful scrutiny that I might follow up on.

AI substitutes for thought in other, more subtle ways. It is already shaping how I think of existing research. Not only does it summarize what's out there; it also tells me how adjacent research relates to my work and how I should think about it. It makes connections across different parts of the literature that had not occurred to me.

Therein lies the bigger danger. The public discussion about AI's impact on society focuses largely on the potential displacement of workers and the loss of jobs. But an even greater risk is the displacement of human thought. When we allow AI to do the job of thinking for us, we cross an important threshold. Our collective ability to think degrades, as does our incentive to learn to think. And because the line between applying thought to a problem and thought itself is already fuzzy, it is easily crossed.

In an interesting recent paper, Daron Acemoglu, Dingwen Kong, and Asuman Ozdaglar of MIT formalize an intuition about how such cognitive offloading can produce catastrophic results. They ask what happens when AI models get very good at providing the kind of context-specific knowledge that can help people perform

whatever specific tasks they are engaged in. Such outputs would allow people to achieve better outcomes, even with less learning.

But there is a problem here, because knowledge has an important externality. As I think about how to solve my problem, I also contribute to the general stock of knowledge about how others can solve theirs. When I invest less in my own learning, the general stock of knowledge suffers. In the limiting, dystopic case, general knowledge disappears completely.

True, this is only a theoretical possibility for now, and depending on what one assumes about the strength of competing effects, better outcomes are also possible. But the danger is real. When we allow AI to do our learning and thinking for us, we degrade our own human capabilities—and risk eventually destroying the knowledge base on which AI itself relies.

Addressing these issues will require the development of social and professional norms on the appropriate use of AI. For example, researchers may need to include detailed disclosures about how they have used AI—a process that could be automated by AI tools themselves—with publication and promotion decisions weighted heavily toward products of the human mind. Organizations like the Partnership on AI can help develop and disseminate general principles. We will need new forms of government regulation as well, as virtually every new technology has required.

A necessary condition for such remedies is a new way of thinking about AI. Above all, the public discourse needs a different framing. The question we should be discussing is not what AI will do to us, but what we want it to do *for us*.