IDEAS

AI's future worries us. So does AI's present.

The long-term risks of artificial intelligence are real, but they don't trump the concrete harms happening now.

By **Jacqueline Harding and Cameron Domenico Kirk-Giannini** Updated July 14, 2023, 3:00 a.m.
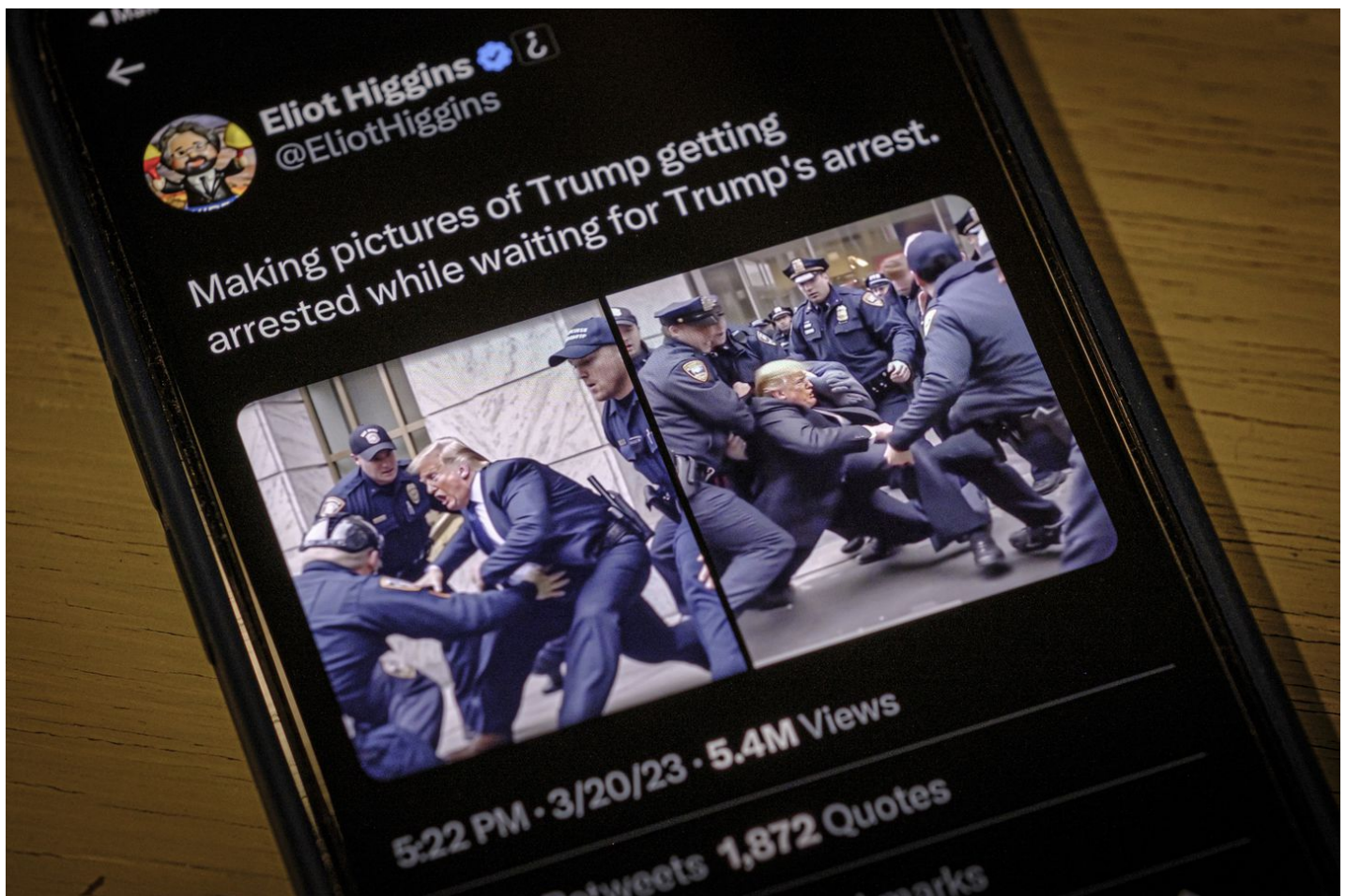
AI's potential for disinformation is high. Eliot Higgins used a widely accessible program to create realistic-looking images of a fictitious skirmish between Donald Trump and police officers. J. DAVID AKE/ASSOCIATED PRESS

"Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war." So say an impressively long list of academics and tech executives in a one-sentence statement released on May 30. We are independent research fellows at the Center for AI Safety, the interdisciplinary San Francisco-based nonprofit that coordinated the statement, and we agree that societal-scale risks from future AI systems are worth taking very seriously. But acknowledging the risks associated with future systems should not lead researchers and policymakers to overlook the all-too-real risks of the artificial intelligence systems that are in use now.

AI is already causing serious problems. It is facilitating disinformation, enabling mass surveillance, and permitting the automation of warfare. It disempowers both low-skill workers who are vulnerable to having their jobs replaced by automation and people in creative industries who have not consented for their work to be used as training data. The process of training AI systems comes at a high environmental cost. Moreover, the harms of AI are not equally distributed. Existing AI systems often reinforce societal structures that marginalize people of color, women, and LGBT+ people, particularly in the criminal justice system or health care. The people developing and deploying AI technologies are rarely representative of the population at large, and bias is baked into large models from the get-go via the data the systems are trained on.

All too often, future risks from AI are presented as though they trump these concrete present-day harms. In a recent CNN interview, AI pioneer Geoffrey Hinton, who recently left Google, was asked why he didn't speak up in 2020 when Timnit Gebru, then co-leader of Google's Ethical AI team, was fired from her position after raising awareness of the sorts of harms discussed above. He responded that her concerns weren't "as existentially serious as the idea of these things getting more intelligent than us and taking over." While we applaud Hinton's resignation from Google to draw attention to the future risks of AI, rhetoric like this should be avoided. It is crucial to speak up about the present-day harms of AI systems, and talk of "larger-scale" risks should not be used to divert attention away from them.

Downplaying the present effects of AI systems benefits nobody but the companies developing the technology. These companies have said they want AI to be regulated in theory, but they have resisted proposals that would interfere with their business models, such as "minimizing" training data to mitigate bias and allowing recourse for data theft. Regulation can't be relegated to the future. Talk of large-scale AI risk, however sincere, may amount to little more than free advertising for AI companies, unless there is regulatory follow-through to tackle both present and future harms.

Because of an erroneous identification by a facial recognition system, Randal Quran Reid of Atlanta was falsely accused of stealing purses in a state he had never visited. He spent nearly a week in jail. NICOLE CRAINE/NYT

The idea that attending to the risks of future AI systems precludes attending to the harms of present systems strikes us as both unfortunate and mistaken. The same structural factors that could make AI a threat to civilization at large also underlie AI-powered surveillance, propaganda, misinformation, discrimination, and job displacement. In all these cases, the basic problem is that corporations are incentivized to develop and deploy AI systems in a way that prioritizes their interests or those of their political stakeholders

over those of the rest of us. These technologies are being steered in directions that put them into conflict with human flourishing. When corporations compete to package research models into consumer-facing products, they will not ensure that the systems they develop are safe and free of bias. Data about consumers is a commodity that can be bought and sold, so companies will develop systems that collect and aggregate it. When a chatbot — [however flawed](#) — will work for free, companies will force workers out of their jobs.

Although it is an open question whether (and when) significantly more capable AI systems will be developed, discussing such systems is no longer the province of sci-fi. It is clear that our species is inventing technologies that could cause irreparable harm if they are not deployed in a way that is sensitive to human needs. Where some see two areas of inquiry — the harms of present models and those of future ones — we see one: the study of how to integrate AI into the fabric of human societies in a way that enhances rather than endangers our collective well-being.

Jacqueline Harding, a PhD student at Stanford University, and Cameron Domenico Kirk-Giannini, assistant professor of philosophy at Rutgers University, are philosophy fellows at the Center for AI Safety, based in San Francisco.

Show comments