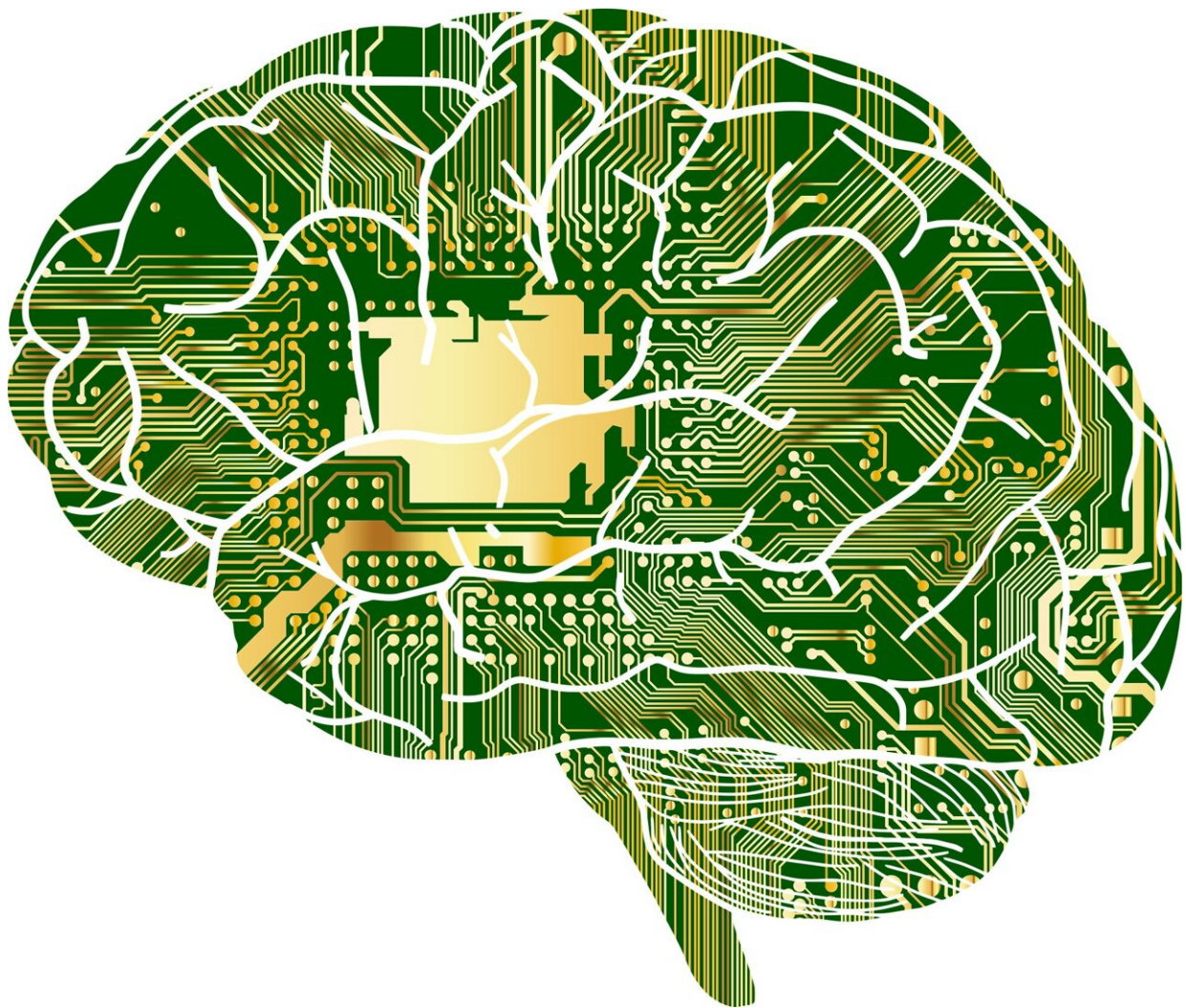


AI has social consequences, but who pays the price? Tech companies' problem with 'ethical debt'

April 19 2023, by Casey Fiesler



Credit: Pixabay/CC0 Public Domain

As public concern about the ethical and social implications of artificial intelligence keeps growing, it might seem like it's time to slow down. But inside tech companies themselves, the sentiment is quite the opposite. As Big Tech's AI race heats up, it would be an "absolutely fatal error in this moment to worry about things that can be fixed later," a Microsoft executive wrote in an internal email about generative AI, [as the New York Times reported](#).

In other words, it's time to "move fast and break things," [to quote Mark Zuckerberg's](#) old motto. Of course, when you break things, you might have to fix them later—at a cost.

In software development, the term "[technical debt](#)" refers to the implied cost of making future fixes as a consequence of choosing faster, less careful solutions now. Rushing to market can mean releasing software that isn't ready, knowing that once it does hit the market, you'll [find out what the bugs are](#) and can hopefully fix them then.

However, negative news stories about generative AI tend not to be about these kinds of bugs. Instead, much of the concern is about AI systems amplifying [harmful biases and stereotypes](#) and [students using AI](#) deceptively. We hear about [privacy concerns](#), [people being fooled by misinformation](#), [labor exploitation](#) and fears about how quickly [human jobs may be replaced](#), to name a few. These problems are not software glitches. Realizing that a technology reinforces oppression or bias is very different from learning that a button on a website doesn't work.

As [a technology ethics educator and researcher](#), I have thought a lot about these kinds of "bugs." What's accruing here is not just technical debt, but [ethical debt](#). Just as technical debt can result from limited testing during the development process, ethical debt results from not considering possible negative consequences or societal harms. And with ethical debt in particular, the people who incur it are rarely the people

who pay for it in the end.

Off to the races

As soon as OpenAI's ChatGPT [was released in November 2022](#), the starter pistol for today's AI race, I imagined the debt ledger starting to fill.

Within months, Google and Microsoft released their own generative AI programs, which seemed rushed to market in an effort to keep up. Google's stock prices fell when its chatbot Bard confidently [supplied a wrong answer](#) during the company's own demo. One might expect Microsoft to be particularly cautious when it comes to chatbots, considering Tay, its Twitter-based bot that was almost immediately shut down in 2016 after spouting misogynist and white supremacist talking points. Yet early conversations with the AI-powered Bing [left some users unsettled](#), and it [has repeated known misinformation](#).

When the social debt of these rushed releases comes due, I expect that we will hear mention of unintended or unanticipated consequences. After all, even with ethical guidelines in place, it's not as if OpenAI, Microsoft or Google can see the future. How can someone know what societal problems might emerge before the technology is even fully developed?

The root of this dilemma is uncertainty, which is a common side effect of many technological revolutions, but [magnified in the case of artificial intelligence](#). After all, part of the point of AI is that its actions are not known in advance. AI may not be designed to produce negative consequences, but it is designed to produce the unforeseen.

However, it is disingenuous to suggest that technologists cannot accurately speculate about what many of these consequences might be.

By now, there have been countless examples of how AI can reproduce bias and exacerbate social inequities, but these problems are rarely publicly identified by tech companies themselves. It was external researchers who found racial bias in widely used commercial [facial analysis systems](#), for example, and in a [medical risk prediction algorithm](#) that was being applied to around 200 million Americans. Academics and advocacy or [research organizations](#) like the [Algorithmic Justice League](#) and the [Distributed AI Research Institute](#) are doing much of this work: identifying harms after the fact. And this pattern doesn't seem likely to change if [companies keep firing ethicists](#).

Speculating—responsibly

I sometimes describe myself as [a technology optimist](#) who thinks and prepares like a pessimist. The only way to decrease ethical debt is to take the time to think ahead about things that might go wrong—but this is not something that technologists are necessarily [taught to do](#).

[Scientist and iconic science fiction writer Isaac Asimov](#) once said that sci-fi authors "foresee the inevitable, and although problems and catastrophes may be inevitable, solutions are not." Of course, science fiction writers do not tend to be tasked with developing these solutions—but right now, the technologists developing AI are.

So how can AI designers learn to think more like science fiction writers? [One of my current research projects](#) focuses on developing ways to support this process of ethical speculation. I don't mean designing with far-off robot wars in mind; I mean the ability to consider future consequences at all, including in the very near future.

This is a topic [I've been exploring](#) in my teaching for some time, encouraging students to think through the ethical implications of sci-fi technology in order to prepare them to do the same with technology they

might create. One exercise I developed is called the [Black Mirror Writers Room](#), where students [speculate about possible negative consequences](#) of technology like social media algorithms and self-driving cars. Often these discussions are based on [patterns from the past or the potential for bad actors](#).

Ph.D. candidate [Shamika Klassen](#) and I evaluated this teaching exercise in a [research study](#) and found that there are pedagogical benefits to encouraging computing students to imagine what might go wrong in the future—and then brainstorm about how we might avoid that future in the first place.

However, the purpose isn't to prepare students for those far-flung futures; it is to teach speculation as a skill that can be applied immediately. This skill is especially important for helping students [imagine harm to other people](#), since technological harms often disproportionately impact marginalized groups that are underrepresented in computing professions. The next steps for my research are to translate these ethical speculation strategies for real-world technology design teams.

Time to hit pause?

In March 2023, [an open letter](#) with thousands of signatures advocated for pausing training AI systems more powerful than GPT-4. Unchecked, AI development "might eventually outnumber, outsmart, obsolete and replace us," or even cause a "loss of control of our civilization," its writers warned.

As [critiques of the letter](#) point out, this focus on hypothetical risks ignores actual harms happening today. Nevertheless, I think there is little disagreement among AI ethicists that AI development needs to slow down—that developers throwing up their hands and citing "unintended

consequences" is not going to cut it.

We are only a few months into the "AI race" picking up significant speed, and I think it's already clear that ethical considerations are being left in the dust. But the debt will come due eventually—and history suggests that Big Tech executives and investors may not be the ones paying for it.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: AI has social consequences, but who pays the price? Tech companies' problem with 'ethical debt' (2023, April 19) retrieved 6 December 2023 from <https://techxplore.com/news/2023-04-ai-social-consequences-pays-price.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.