

Why AI Bias Can Hurt Your Business?

How can business leaders ensure AI bias doesn't harm their company?



FRAN PULIDO

The algorithm was designed with good intentions: identify patients who would benefit from extra medical care. This would help them avoid future complications and, in turn, relieve strain on the medical system. The tool had become widely used by hospitals in the United States—but then a

2019 study in *Science* unearthed a problem. The algorithm was assigning unfairly low risk scores to Black patients, meaning that those people were less likely to be singled out for high-risk care management. This was because the algorithm used patients' past healthcare costs as a way to gauge their medical needs, which functioned as a proxy

for race. In the United States, unequal access to medical care means that Black patients typically incur lower costs than white patients. Simply put, it was perpetuating a problem, not solving one. If the algorithm were to deliver more equitable results, the report argued, it would raise the percentage of Black patients benefiting from extra care from 17.7 percent to 46.5 percent.

It's a notable example of algorithmic bias, which is a serious concern in the algorithm-driven world of artificial intelligence. AI is a powerful tool and has led to advancements in everything from computer vision and translation to cybersecurity and drug discovery. For businesses, its potential to unlock cost efficiencies, save time, and improve profitability across a wide gamut of activities is enormous. But AI is fueled by data. If that data reflects human bias or contains undesirable correlations, the systems can make decisions that disadvantage certain groups. This can include discrimination on the basis of protected characteristics such as age, gender, or race.

As AI enters ever more areas of our lives—the global AI market size is forecast to be \$1.8tn by 2030—the stakes are only going to get higher. In May, Geoffrey Hinton—a computer scientist widely hailed as “the Godfather of AI”—resigned from his role at a major tech company in order to have the freedom to warn about the potential perils of the tools his work helped create.

This is a live issue for policymakers. The EU's Ar-

tificial Intelligence Act, expected to come in later this year, or early next, seeks to reduce the threat that AI poses to human rights. Some believe that more drastic action is needed. A group of AI experts recently signed an open letter calling on governments to step in and issue a moratorium if labs don't pause development of AI models more powerful than OpenAI's GPT-4.

The issue of AI bias has entered the popular conversation, too, thanks to landmark developments in generative AI. Easy-to-use and startlingly capable tools such as ChatGPT, which can generate text or images, have drawn widespread attention to how sophisticated AI has become. The associated opportunities and risks are now high on the business agenda. “ChatGPT was a wake-up call for boards of directors to be far better educated as to what AI is, and what it means for society,” says Alison Kay, UK&I Managing Partner for Client Service at professional services firm EY. “I think it's vital we're asking these questions now, because although I think AI will be truly transformative, and can be a force for good, we've got to be equally conscious of the potential downsides and the impact on real lives”.

Tackling AI bias is a commercial imperative

Bias can creep into AI through various routes. The most direct is when the data simply holds a mirror up to society and reflects human biases. It may under- or over-represent certain groups, it may contain prejudiced viewpoints, or it may have been

prepared by people who are themselves biased. Sometimes the data may contain correlations that are true, but which should be ignored for ethical or legal reasons. Excluding them can be easier said than done, however, as the algorithms may factor them in by proxy—as happened in the healthcare prediction example.

A less direct way that a system can develop bias is through identifying spurious correlations. One famous instance involves a tech company that devised an AI tool for screening CVs. The engineers behind it trained it on data about previous applicants, to teach it what counted as a “good” hire. The problem was, the company’s staff were mostly male, so the AI deduced that “being female” was an undesirable trait. Because AI has no structural understanding of the world—it simply finds patterns—it can’t consciously counteract these kinds of obvious errors. “It’s not a mind in the way that we think about the mind,” says Kay. “We aren’t comparing like for like. Because of its ability to compute and to make inferences and to pick up subtleties around the data, we can end up with a very different outcome to what was intended.” Worse still, if the AI tool is a “black box” that can’t show its working—something which historically has been commonplace—these biases could go undetected and have devastating results.

Organizations have an ethical obligation to root out algorithmic biases and put humans first. “The possibility that it can build further bias into what is

an already biased society is significant,” says Kay. “But the scale at which AI is appearing in society means this risk is growing exponentially. It can already be pretty harmful now—it may be disastrous in the future.” A recent study from University of California, Berkeley, for example, found that algorithmic bias in fintech tools meant that Latino and African-American borrowers were paying higher rates of interest on their mortgages. Collectively, this was costing those groups an extra \$765 million per year. What’s particularly insidious is that this form of discrimination is invisible—it can take an academic study to bring it to light. What’s more, it entrenches prejudice on a systemic level. A 2020 paper for the Association for Computational Linguistics highlighted how AI models often used for moderating online discussions wrongly censored comments mentioning disabilities. Education and information are a vital part of tackling prejudice, but these models erased the topic from the conversation altogether.

For businesses, addressing AI bias is not just a matter of social responsibility. It’s also a commercial imperative. If a company is found to be using biased tools, the consequences can be severe. “The biggest risk is that our customers lose trust,” says Kay. “It doesn’t take much these days to lose faith, especially with social media and with regulatory frameworks being heightened.” Consumer expectations are towering, and it’s easier than ever for them to take their custom elsewhere.

Bias issues can also erode trust of AI systems within the business. Even if the problem is solved, employees or executives may be wary of using AI in the future. This can hamstring operations, as choosing not to use this technology can put the organization at a competitive disadvantage. For Kay, a careful balance is required. "As much as possible you need to mitigate the downsides while embracing the positives."

So, where should businesses start?

Adherence to new AI regulations will be important, but solving this problem will require a collective effort. Businesses need to be proactive, and Kay recommends a two-pronged approach. "You need both technical and cultural solutions to manage your way through this," she says. Technical solutions might encompass designing policies and standards for the development of AI; creating protocols for reviewing training data for bias; and setting up an AI ethics board to review the firm's use of these tools. "It sounds awfully boring in the context of the very exciting AI world," she says. "But unless businesses actually put those frameworks into place, it won't work. Trust can be lost very quickly—and it's almost impossible to regain." The cultural solutions are less concrete, but no less important. These could include fostering an ethos of transparency and responsibility around how AI is used, implementing schemes to inform employees and customers about the opportunities and risks, and bringing the topic firmly onto the or-

ganization's purpose and responsibility agenda. Efforts could also be made to improve data literacy across the organization, so that a greater diversity of people can be involved in evaluating and interrogating how data is used. "I think data literacy is really critical," says Kay. "I still think organizations leave this to a small number of tech-focused people—there needs to be knowledge at all levels, but especially the board of directors."

But it's not enough simply to come up with strategies. There also have to be controls—checks and balances—to ensure that the theory is being put into practice. Audits are a sensible option. This could involve an external audit, says Kay, but should begin with an internal one. "An external audit really should act as a final check—it first needs to be internally faced up to."

Indeed, if the boardroom isn't taking this issue seriously, the whole strategy is likely to be undermined. Buy-in at the most senior level is critical. "You can put in place all the governance, frameworks, and controls," says Kay. "But it's that tone from the top that will ultimately make a difference." Businesses have an opportunity to lead the way on the AI bias issue—and it's vital they seize it. After all, if history has taught us anything, it's that disruptive technologies can have profound consequences—for good and ill—and trying to manage the downsides after the fact can be a fool's errand. "We are in danger of being led by the technology," says Kay. "We need to ensure we are led by people."