

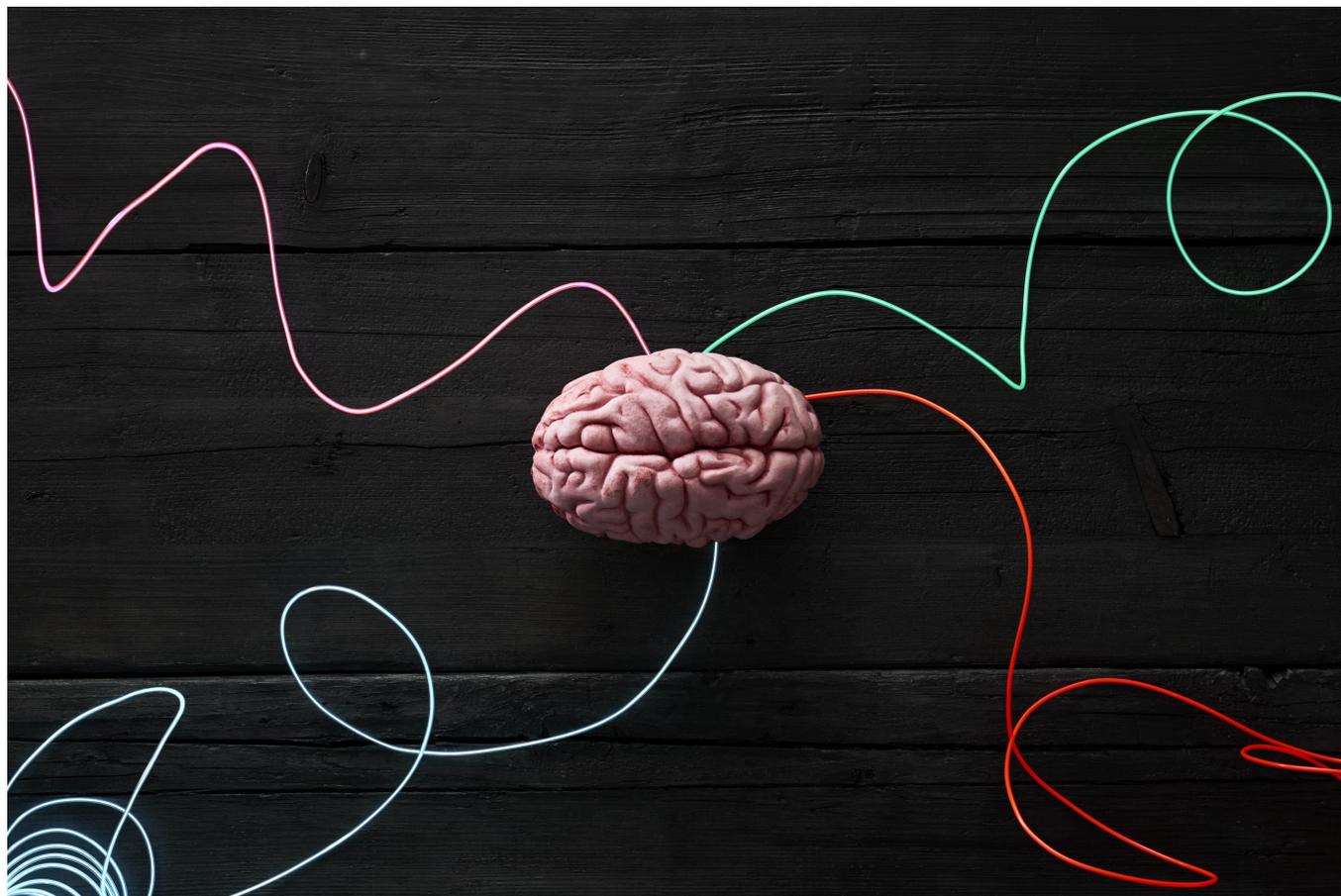
# A.I. Ethics Boards Should Be Based on Human Rights

Tech companies should ensure their ethics boards are guided by universal human rights and resist bad faith arguments about diversity and free speech



Evan Selinger [Follow](#)

Jun 20 · 8 min read ★



Credit: Westend61/Getty Images

Co-authored with [Brenda K Leong](#)



Who should be on the ethics board of a tech company that's in the business of artificial intelligence (A.I.)? Given the attention to the devastating failure of Google's proposed Advanced Technology External Advisory Council (ATEAC) earlier this year, which was announced and then canceled within a week, it's crucial to get to the bottom of this question. Google, for one, admitted it's "going back to the drawing board."

Tech companies are realizing that artificial intelligence changes power dynamics and as providers of A.I. and machine learning systems, they should proactively consider the ethical impacts of their inventions. That's why they're publishing vision documents like "Principles for A.I." when they haven't done anything comparable for previous technologies. (Google never published a "Principles for Web Search.") But what version of ethics should they choose? Ethical norms, principles, and judgments differ between time, place, and culture, and might be irreconcilable even within local communities. There's so much disagreement that red lines aren't even easily drawn around truly alarming A.I. applications, like lethal autonomous weapons and government scoring systems such as the one China is experimenting with.

Further complications arise because businesses, unlike individuals or governments, are accountable to shareholders. Fulfilling their fiduciary obligations can mean prioritizing growth, emphasizing profit, and working with international clients whose political allegiances vary along the democratic-authoritarian continuum.

This has led, understandably, to skepticism about the sincerity of corporate ethics. Whenever tech companies talk about ethics, critics worry that it's a strategy for avoiding stronger government regulations and gaining goodwill, consisting of empty slogans followed by minimal legal compliance. Hence, when tech companies establish external A.I. ethics boards, they'll probably be viewed as self-serving, "ethics washing" facades.

The cynicism is often justified. After all, these companies tout transparency but promote "trade language that prevents it," are accused of supporting A.I. ethics researchers who don't always disclose their corporate funding, and, in one prominent case, keep the composition of an ethics board shrouded in secrecy. If tech companies want to create meaningful ethical practices, they need to invite the right people to their ethics boards and empower these folks to make publicly available recommendations that hold businesses accountable to transparent standards.

## Human rights and A.I. ethics

Through design decisions, implementation strategies, and unintended consequences, A.I. has the potential to impact lives across the globe. In military, law enforcement, banking, criminal justice, employment, and even product delivery contexts, algorithmic systems can threaten human rights by automating discrimination, chilling public speech and assembly, and limiting access to information.

Recognizing these problems, leading frameworks for A.I. governance base their approach on human rights, including the European Commission's Ethics Guidelines for Trustworthy A.I., the Organisation for Economic Co-operations and Development's Principles on A.I., Business for Social Responsibility's A.I.: A Rights-Based Blueprint for Business, Data and Society's Governing A.I.: Upholding Human Rights and Dignity, and Access Now's Human Rights in the Age of A.I.

Human rights are imperfect ideals, subject to conflicting interpretations, and embedded in agendas with "outsized expectations." Though supposedly global, human rights aren't honored everywhere. Nevertheless, the United Nations Universal Declaration of Human Rights is the best statement ever crafted for establishing all-around social and legal equality and fundamental individual freedoms. The Institute of Electrical and Electronics Engineers rightly notes that human rights are a viable benchmark, even among diverse ethical traditions.

"Whether our ethical practices are Western (Aristotelian, Kantian), Eastern (Shinto, Confucian), African (Ubuntu), or from a different tradition, by creating autonomous and intelligent systems that explicitly honor inalienable human rights and the beneficial values of their users, we can prioritize the increase of human well-being as our metric for progress in the algorithmic age."

Technology companies should embrace this standard by explicitly committing to a broadly inclusive and protective interpretation of human rights as the basis for corporate strategy regarding A.I. systems. They should only invite people to their A.I. ethics boards who endorse human rights for everyone.

Diversity matters when ethically evaluating A.I. impacts, particularly on the most marginalized populations. But it shouldn't include representing those who engage in such marginalization.

Once board members are selected, tech companies must require these experts to maintain a demonstrated commitment to human rights throughout their tenure. If due process reveals that a board member says or does something that is substantively out of line with human rights, she should be removed no matter how high her profile or how significant her past contributions. It's that simple. The penalty is strong but appropriate, and it disincentivizes "digital ethics shopping," which is the corporate malpractice of appealing to malleable ethical benchmarks to justify status quo behavior.

### **Whose human rights?**

Article 4 of the UN Declaration of Human Rights states no one shall be held in slavery. Surely, at this moment in history, that should be an uncontroversial ethical proposition. Yet the International Labour Organization (ILO) estimates that as of 2017, 25 million people are victims of forced labor or slavery. These include cases of sex trafficking, child soldiers, and labor for domestic work, construction, and agriculture. So if such an obvious human right remains aspirational, consider the challenges of more complex issues, such as Article 16's right to marry.

The LGBTQ+ community is one of the most vulnerable worldwide, subject to everything from limited employment (Article 23) and social opportunities (Article 22) to conversion therapy and even torture (Article 5). Countries that criminalize same-sex relationships and gender nonconforming identities have higher rates of mental health issues and suicide within these groups. And gay marriage remains a controversial topic almost everywhere, one that's often linked to religiously-based, restrictive conceptions of human rights.

When Google appointed Kay Cole James, president of the Heritage Foundation, to its technology advisory council, many critics — including a number from within Google — objected that the company was pandering to conservatives. Google's A.I. Principles were

central to the debate because they include a corporate commitment to avoid creating or using “technologies whose purpose contravenes widely accepted principles of international law and human rights.” Since James is known for being anti-LGBTQ+ concerning trans individuals who don’t fit within her personal views on human sexuality, and leads the Heritage Foundation, long a proponent of “traditional” marriage, how could she be expected to hold Google accountable to its stated ideals?

Likewise, because Google implicitly validated Cole’s position by inviting her to join the council, the company inadvertently harmed marginalized communities. Their suffering wouldn’t be negated even if, somehow, James set aside her conflicting opinions in order to hold the company to its ideals during board meetings — ideals that, at least in part, clashed with James’ own convictions.

The advisory council controversy demonstrated that Google’s A.I. Principles are too vaguely worded to count as a clear policy statement, and contain caveats that may function as loopholes to cooperate with governments or businesses that aren’t fully committed to human rights. If a company doesn’t acknowledge and accept the inevitable financial impacts of establishing artificial intelligence principles based on human rights, its professed values are meaningless. Critics will rightly judge the appeal to ethics as PR cover for doing business as usual.

That said, Google deserves some credit. Amazon shareholders recently forced a vote on the company selling facial recognition technology to government agencies out of concern over potential human rights violations. While the motion to discontinue sales failed, the shareholders should be commended for appealing to human rights principles. Imagine if Amazon already had a properly configured A.I. ethics board and leadership took its advice seriously. Perhaps, then, the company’s strategy would have been like the one Google chose: Google said it’s too soon to sell facial recognition systems to the governments, leading the ACLU’s Nicole Ozer to praise its moral compass.

## **Diversity, speech, and human rights**

Critics will object by saying that if tech companies insist upon an inclusive interpretation of human rights, the board’s ability to represent diverse opinions will be diminished. Diversity matters when ethically evaluating A.I. impacts, particularly on the most

marginalized and vulnerable populations. But it shouldn't include representing those whose views support such marginalization.

Simply culling outliers who can't agree on the innate worth and equality of every human shouldn't be considered a radical step of exclusion.

People who are committed to human rights already represent a full spectrum of perspectives and experiences. They will disagree with each other about how to establish priorities and apply principles. Diversity and human rights are complementary because the human rights community isn't a monolithic group of like-minded people. Nor are those impacted mere stereotypes, but real humans with their quality of life — or even life itself — at stake. Simply culling outliers who can't agree on the innate worth and equality of every human shouldn't be considered a radical step of exclusion any more than understanding the need to censor hate speech should equate to being against the fundamental right to free speech.

To neutralize another objection, insisting that ethics board members consistently support these rights doesn't in any way impact free speech. Corporate speech, after all, isn't government speech, and outside of government restraint there are many situations where individuals aren't allowed to speak freely, at least not without punitive consequences.

Corporate social media policies restrict employee speech by prohibiting negative representations of brands or employee satisfaction. Properly functioning media limit advocacy-based arguments (like this industry spin) to the opinion section and have guidelines for avoiding false balance on issues like climate change and vaccine safety. Even universities, institutions widely viewed as places that should protect speech with the same broad boundaries as governments, don't shield all opinions in every context.

In fact, choosing individuals who are strongly pro-human rights might increase free speech. That's because many human rights principles explicitly include protections for free expression. Conversely, including people who don't toe the line on human rights invites chilling attacks.

When critics complain that human rights are a pretext for suppressing conservative views, they're being disingenuous. The only way for tech companies working on A.I. to be responsible global actors is to reject intolerance and the abuses of language that prop it up, including weaponized appeals to "ideological diversity" that really mean permission for ethics boards to reduce or redefine the fundamental human rights that every person deserves to have protected.

A foundational commitment to human rights should lead to better ethical decisions about A.I.-based systems. As a start, it puts companies on notice. They shouldn't be in the business of lethal autonomous weapons, government scoring systems, and government facial recognition systems if they can't make a robust case for how these endeavors can coexist with human rights protections. And that doesn't even begin to address the less obvious gray areas where A.I. will create a myriad of unforeseen consequences. Before more lives are impacted, we all deserve assurance that tech companies will roll out A.I. services with the aim of protecting essential rights and liberties.

. . .

*Evan Selinger is Professor of Philosophy at Rochester Institute of Technology.*

*Brenda Leong is Senior Counsel & Director of Strategy at the Future of Privacy Forum.*

[Human Rights](#)   [Artificial Intelligence](#)   [Ethics](#)   [Industry](#)   [Machine Learning](#)

[About](#)   [Help](#)   [Legal](#)