

# What is 'ethical AI' and how can companies achieve it?

May 25 2023, by Dennis Hirsch and Piers Turner

---



Credit: Pixabay/CC0 Public Domain

The rush to deploy powerful new generative AI technologies, such as ChatGPT, has raised alarms about [potential harm and misuse](#). The law's

glacial response to such threats has [prompted demands](#) that the companies developing these technologies implement AI "ethically."

But what, exactly, does that mean?

The straightforward answer would be to align a business's operations with one or more of the [dozens of sets of AI ethics principles](#) that governments, multistakeholder groups and academics have produced. But that is easier said than done.

We and our colleagues spent [two years interviewing and surveying AI ethics professionals](#) across a range of sectors to try to understand how they sought to achieve ethical AI—and what they might be missing. We learned that pursuing AI ethics on the ground is less about mapping [ethical principles](#) onto corporate actions than it is about implementing management structures and processes that enable an organization to spot and mitigate threats.

This is likely to be disappointing news for organizations looking for unambiguous guidance that avoids gray areas, and for consumers hoping for clear and protective standards. But it points to a better understanding of how companies can pursue ethical AI.

## **Grappling with ethical uncertainties**

[Our study](#), which is the basis for a [forthcoming book](#), centered on those responsible for managing AI ethics issues at major companies that use AI. From late 2017 to early 2019, we interviewed 23 such managers. Their titles ranged from privacy officer and privacy counsel to one that was new at the time but increasingly common today: data ethics officer. Our conversations with these AI ethics managers produced four main takeaways.

First, along with its many benefits, business use of AI poses substantial risks, and the companies know it. AI ethics managers expressed concerns about [privacy](#), [manipulation](#), bias, opacity, inequality and labor displacement. In one well-known example, [Amazon developed an AI tool to sort résumés](#) and trained it to find candidates similar to those it had hired in the past. Male dominance in the tech industry meant that most of Amazon's employees were men. The tool accordingly learned to reject female candidates. Unable to fix the problem, Amazon ultimately had to scrap the project.

Generative AI raises additional worries about [misinformation and hate speech at large scale](#) and misappropriation of [intellectual property](#).

Second, companies that pursue ethical AI do so largely for strategic reasons. They want to sustain trust among customers, business partners and employees. And they want to preempt, or prepare for, emerging regulations. The [Facebook-Cambridge Analytica scandal](#), in which Cambridge Analytica used Facebook user data, shared without consent, to [infer the users' psychological types](#) and target them with manipulative political ads, showed that the unethical use of advanced analytics [can eviscerate a company's reputation](#) or even, as in the case of Cambridge Analytica itself, bring it down. The companies we spoke to wanted instead to be viewed as responsible stewards of people's data.

The challenge that AI ethics managers faced was figuring out how best to achieve "ethical AI." They looked first to AI ethics principles, particularly those rooted in bioethics or human rights principles, but found them insufficient. It was not just that there are many competing sets of principles. It was that justice, fairness, beneficence, autonomy and other such principles are contested and subject to interpretation and can conflict with one another.

This led to our third takeaway: Managers needed more than high-level

AI principles to decide what to do in specific situations. One AI ethics manager described trying to translate human rights principles into a set of questions that developers could ask themselves to produce more ethical AI software systems. "We stopped after 34 pages of questions," the manager said.

Fourth, professionals grappling with ethical uncertainties turned to organizational structures and procedures to arrive at judgments about what to do. Some of these were clearly inadequate. But others, while still largely in development, were more helpful, such as:

- Hiring an AI ethics officer to build and oversee the program.
- Establishing an internal AI ethics committee to weigh and decide hard issues.
- Crafting data ethics checklists and requiring front-line data scientists to fill them out.
- Reaching out to academics, former regulators and advocates for alternative perspectives.
- Conducting algorithmic impact assessments of the type already in use in environmental and privacy governance.

## **Ethics as responsible decision-making**

The key idea that emerged from [our study](#) is this: Companies seeking to use AI ethically should not expect to discover a simple set of principles that delivers correct answers from an all-knowing, God's-eye perspective. Instead, they should focus on the very human task of trying to make responsible decisions in a world of finite understanding and changing circumstances, even if some decisions end up being imperfect.

In the absence of explicit legal requirements, companies, like individuals, can only do their best to make themselves aware of how AI affects people and the environment and to stay abreast of [public](#)

[concerns](#) and the latest research and expert ideas. They can also seek input from a large and diverse set of stakeholders and seriously engage with high-level ethical principles.

This simple idea changes the conversation in important ways. It encourages AI ethics professionals to focus their energies less on identifying and applying AI principles—though they remain part of the story—and more on adopting decision-making structures and processes to ensure that they consider the impacts, viewpoints and public expectations that should inform their business decisions.

Ultimately, we believe laws and regulations will need to provide substantive benchmarks for organizations to aim for. But the structures and processes of responsible decision-making are a place to start and should, over time, help to build the knowledge needed to craft protective and workable substantive legal standards.

Indeed, the emerging law and policy of AI focuses on process. [New York City passed a law](#) requiring companies to audit their AI systems for harmful bias before using these systems to make hiring decisions. Members of [Congress have introduced bills](#) that would require businesses to conduct algorithmic impact assessments before using AI for lending, employment, insurance and other such consequential decisions. These laws emphasize processes that address in advance AI's many threats.

Some of the developers of generative AI have taken a very different approach. Sam Altman, the CEO of OpenAI, initially explained that, in releasing ChatGPT to the public, the [company](#) sought to [give the chatbot "enough exposure to the real world"](#) that you find some of the misuse cases you wouldn't have thought of so that you can build better tools." To us, that is not responsible AI. It is treating human beings as guinea pigs in a risky experiment.

Altman's [call at a May 2023 Senate hearing for government regulation](#) of AI shows greater awareness of the problem. But we believe he goes too far in shifting to government the responsibilities that the developers of generative AI must also bear. Maintaining public trust, and avoiding harm to society, will require companies more fully to face up to their responsibilities.

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

Provided by The Conversation

Citation: What is 'ethical AI' and how can companies achieve it? (2023, May 25) retrieved 26 April 2024 from <https://techxplore.com/news/2023-05-ethical-ai-companies.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.