Harvard Business Review

ARTICLE
ETHICS

Create an Ethics Committee to Keep Your AI Initiative in Check

by Steven Tiell

Harvard Business Review

Create an Ethics Committee to Keep Your AI Initiative in Check

by Steven Tiell

NOVEMBER 15, 2019

HBR STAFF/GEOPAUL/GETTY IMAGES

WITF-FM, a public radio, television, and online news broadcaster in central Pennsylvania, includes the following statement above select online news coverage: "WITF strives to provide nuanced perspectives from the most authoritative sources. We are on the lookout for biases or assumptions in our own work, and we invite you to point out any we may have missed."

It's not uncommon for news organizations to invite comments and feedback from their audience; in fact, most encourage it. But WITF has gone above and beyond a general invitation for engagement. This statement highlights the potential for bias in their own reporting — and their attempt to avoid it.

Contemporary sensitivities to bias are growing, and this will only increase with the proliferation and ubiquity of Artificial Intelligence (AI). Most of today's AI systems are built via machine learning, a technique that requires any one of thousands of potential algorithms to "learn" patterns from extremely large stockpiles of data. This should produce a model that is predictive of future real-world scenarios, but bias skews the accuracy of these models. Organizations using AI are starting to recognize the role that strong, organization-wide values must play in fostering responsible innovation, and like WITF, many leaders are going above and beyond. For example, Marc Benioff, founder and co-CEO of Salesforce, vocally advocates for companies to take responsibility for their contributions to society. Those values are highly aligned with Salesforce's company culture, which prioritized the creation of two first-of-their-kind roles: Chief Equality Officer (Tony Prophet) and a Chief Ethical and Humane Use Officer (Paula Goldman), to elevate the responsibility for protecting these organizational values to the C-suite.

With a strong commitment across the organization, businesses can align, distribute, and scale values-led decision-making, which builds tremendous trust with both internal and external stakeholders. This is the bedrock of a responsible organization.

But maintaining an organization's values across products and services — particularly as organizations start to use AI to help make or inform decisions — will require strong internal governance. Consider a situation where AI technology is perceived as unfair or malfunctioning. Can the creator be held accountable? Is there recourse for consumers in the form of an accountable engineer, an internal governance board, or even external governance? To use the Salesforce example again, the company is governing itself via a stakeholder-rich ethics committee that oversees ethics-related decisions at the organization.

Every organization can create a strong internal governance framework to address how they design and implement AI. In collaboration with the Ethics Institute at Northeastern University, Accenture has released a report on *Building Data & AI Ethics Committees*, which can serve as a manual. It provides a roadmap and identifies key decisions that organizations will need to make: What does success look like? What are the values the committee is meant to promote and protect? What types of expertise are needed? What is the purview of the committee? What are the standards by which judgments are made?

Establishing this level of ethical governance is critical to helping executives mitigate downside risks, because addressing AI bias can be extremely complex. Data scientists and software engineers have biases just like everyone else, and when they allow these biases to creep into algorithms or the data sets used to train them — however unintentionally — it can leave those subjected to the AI feeling

like they have been treated unfairly. But eliminating bias to make fair decisions is not a straightforward equation.

While many colloquial definitions of "bias" involve "fairness," there is an important distinction between the two. Bias is a feature of statistical models, while fairness is a judgment against the values of a community. Shared understandings of fairness are different across cultures. But the most critical thing to understand is their relationship. The gut feeling may be that fairness requires a lack of bias, but in fact, data scientists must often introduce bias in order to achieve fairness.

Consider a model built to streamline hiring or promotions. If the algorithm learns from historic data, where women have been under-represented in the workforce, myriad biases against women will emerge in the model. To correct for this, data scientists might choose to introduce bias — balancing gender representation in historic data, creating synthetic data to fill in gaps, or correcting for balanced treatment (fairness) in the application of data-informed decisions. In many cases, there's no possible way to be both unbiased and fair.

An Ethics Committee can help to not only maintain an organization's values-based intentions, but can increase transparency into how they use AI. Even when it's addressed, AI bias can still be maddening and frustrating for end users, and most companies deploying AIs today are subjecting people to it without giving them much agency in the process. Consider the experience of using a mapping app. When travelers are simply told which route to take, it is an experience stripped of agency; but when users are offered a set of alternate routes, they feel more confident in the selected route because they enjoyed more agency, or self-determination, in choosing it. Maximizing agency when AI is being used is another safeguard strong governance can help to ensure.

Organizations that choose to establish an ethics committee aren't simply addressing human bias, statistical bias, and fairness; their scope of concerns is typically much wider, and often includes a desire to increase organizational maturity concerning how their products and services impact stakeholders, including civil societies. Committee-based governance provides an institutional feedback loop for how AIs are performing in the real world, giving valuable insights to designers, engineers, and executive teams. The committee should be made up of individuals who represent different views, have internal and external perspectives, and have different technical and non-technical backgrounds. Who sits on these committees, why they were selected, to whom they are accountable, and what their purpose is, should be clearly thought through and articulated.

To lead in this space, organizations should focus on three areas for growth:

- **Establish governance for ethics & AI**. Unintended consequences of AI can be severe and can present existential risks to organizations. Robust and stakeholder-rich governance is the best chance organizations have to identify and manage potential risks. Some organizations find it more feasible to implement "essential governance" — team-based process integrations that introduce minimal friction and have the least potential to delay a project. While this might be an expedient way to get started, its strategic value to the organization will be limited.
- **Describe how/when fairness happens and how/what biases have been accounted for.** Responsible organizations and data scientists invest extraordinary amounts of time and resources deliberating over what biases might exist in data, how those are handled, and what might get amplified or mitigated by algorithm selection. These biases might impact the perceived fairness of AI models, so they should be considered carefully upfront.
- **Provide mechanisms for recourse.** Every organization that deploys AIs should be able to articulate why and by what measure the solution is better than what already exists. This establishes what those subjected to the AIs should expect. When the AI system fails to meet end-user expectations, they need to have immediate and easy-to-access forms of recourse; organizations should consider ways to provide recourse to a responsible engineer (or team that owns the product), internal governance (such as an AI or ethics committee), and external governance (such as a regulator).

How organizations choose to address bias, fairness, governance, and recourse for AI will certainly vary, but the need to create robust governance will only increase for everyone. Those who get started early will enjoy market advantages in both B2B and B2C contexts. Laggards may find themselves boxed out — by competitors managing their digital risks, and by consumers who seek fairness and accountability in the technologies they choose to use.

Steven Tiell leads Responsible Innovation for Accenture Labs.