Ethical artificial intelligence: the most important movement you've never heard of

As science fiction becomes science fact, focusing on ethics will make sure AI benefits all of humanity – not entrench the advantages of a privileged few.

Fake news spreads six times faster than factual news. The algorithms built into social media software, in an effort to hold our attention, feed us information that triggers the basest instincts of our complex psychologies. These are some of the starkest realities exposed in the Netflix documentary "The Social Dilemma."

Holding our attention is what the robot (if you will) was programmed to do. And it's doing its job really, *really* well. "As long as social media companies profit from outrage, confusion, addiction, and depression," says Tristan Harris, founder of The Center for Humane Technology, "our well-being and democracy will continue to be at risk."

Of course, infinite-scroll news feeds are only a tiny piece of the artificial intelligence puzzle. AI and machine learning influence the way we shop, bank, commute – even the decisions we make about who earns an interview vs. whose resume goes into the trash bin.

Never miss an article!

Follow Ladders on Flipboard

Although the robots may or may not be coming to take our jobs, they're already playing an ever-more participatory role in our lives. The upshot? Whether software (and how

we build it) helps or harms us is important for everyone to understand – not just developers. That's where ethical AI comes in.

What is ethical AI?

There are a few different lenses for looking at the ethical considerations of AI. There are the complex philosophical issues and futurism predictions like the "singularity."

There are also science-fiction-like ideas of what would happen if an AI system became "conscious" and able to teach itself whatever it wanted – not just what it's been programmed to learn. And then there's the moral behavior of humans as they design and create intelligent machines.

It's this latter consideration that's been coming up recently, and there are some very smart people who've founded organizations with missions to make us more aware of the situation. One is Ethical AI, whose website includes this data point:

That number should make all of us sit up straighter and pay attention. The kicker is that sometimes the error isn't the result of buggy code, but the result of unconscious bias.

Take this heart attack detector app developed in 2019, for example. It wrongly registered the same symptoms as a panic attack in women and not, as it would've for men, as a heart attack. The AI was unintentionally harmful because it amplified biases that already existed – in this case, the notion that women are overly emotional.

Now consider the fact that facial recognition AI makes more accurate identifications when presented with a white face versus a person of color. The same goes for self-driving cars. They tend to be better at detecting (and therefore avoid hitting) pedestrians with pale skin. Such systems are already being used in policing, and yes: already producing erroneous outcomes.

This is why it's imperative to design AI with ethical principles in place. And the responsibility is on us to be ethical leaders at the grassroots level, rather than relying solely on government or tech giants. They, of course, need to do their part, too. But, in the case of tech companies, the harsh truth is many lack the financial incentive to affect real change.

The principles of ethical artificial intelligence and its importance for developers

Today, governments and other regulators are at least five years behind the curve in terms of understanding what AI is capable of and how to govern it, according to Dr. Catriona Wallace, founder and CEO of Ethical AI Advisory, and founder and director of Flamingo AI.

The Ethical AI Advisory states: "Adopting an ethical approach to the development and use of AI works to ensure organisations, leaders, and developers are aware of the potential dangers of AI, and, by integrating ethical principles into the design, development, and deployment of AI, seek to avoid any potential harm."

Dr. Wallace explains that there are different things that humans value, but that it ultimately comes down to the basics: do no harm, and prioritize health and safety. The Ethical AI website includes eight noteworthy principles and guidelines for ethical AI:

1. **Human, social, and environmental well-being**
   Throughout their lifecycle, AI systems should benefit individuals, society, and the environment.
2. **Human-centered** values
   Throughout their lifecycle, AI systems should respect human rights, diversity, and the autonomy of individuals.
3. **Fairness**
   Throughout their lifecycle, AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities, or groups.
4. **Privacy protection and security**
   Throughout their lifecycle, AI systems should respect and uphold privacy rights and data protection, and ensure the security of data.
5. **Reliability and safety**
   Throughout their lifecycle, AI systems should reliably operate in accordance with their intended purpose.
6. **Transparency and explainability**
   There should be transparency and responsible disclosure to ensure people know when they are being significantly impacted by an AI system and can find out when an AI system is engaging with them.
7. **Contestability**
   When an AI system significantly impacts a person, community, group, or environment, there should be a timely process to allow people to challenge the use or output of the AI system.
8. **Accountability**
   Those responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled.

Questions of ethics also extend to personal data and how it's used in AI algorithms when we interact with apps. Should we be made aware we're interacting with AI? Should we be allowed to opt out?

"The first product my company brought to market was sold in financial services," Dr. Wallace recalls. "One of our customers didn't want their customers to know that they were interacting with a robot. They wanted them to think it was a human. We told them

that was unethical. My strong belief is that customers need to know if they're dealing with a robot, and if the data the robot gathers will be going into a data set that will be used to train the algorithm even further. They should have both of these things listed on their website."

Computers only do what they're programmed to do. Because AI is trained on existing behaviors, they can magnify gender and race biases. "The example of Amazon AI assessing resumes was an AI that used historical data with biases," notes Dr. Wallace. The program had been trained on a data set that massively over-represented men's resumes, which led it to conclude that men were preferable. Women's resumes were automatically downgraded.

In other words, even Amazon – one of the top tech companies in the world – struggles to get it right.

AI can get better and better at feeding you what you seem to like, but it doesn't know *why* you like it. It's pure quantitative analysis to the Nth degree, and value judgments don't factor in at run-time. The only chance to add values to the equation is when selecting the data set used for training. If the engineers are careful to ensure the data set is free of bias, there's a fighting chance the resulting AI will behave fairly and equitably. If not, existing biases will only become more entrenched.

AI vs. machine learning

Artificial intelligence is when machines carry out tasks in a smart way, where "smart" is defined by the engineers who built it. When my phone recognizes that I get up every day at 6:30 a.m. and begins to show me reminders to set my alarm at that time, that's not necessarily machine learning. A human engineer could easily build a rule into my phone's operating system that says "Remind the user to set their wake-up alarm; find the time they most commonly set their alarm for, and suggest that time."

Machine learning is a subset of AI. It involves giving machines access to example data and examples of desirable outcomes, then letting them decide for themselves what's smart and what's not.
An example of machine learning would be if my phone's OS was designed to analyze *all* usage patterns (not just the alarm) and notify me when I deviate from them. It would be as if the machine were thinking "Hmm. Jamey always sets his alarm on Sunday evenings, but it's almost his usual bedtime and he hasn't done that yet. I'd better remind him."

How explicit are the instructions given to the machine? The less explicit, the more likely you're looking at machine learning.

What do AI engineers think about "bad robots?"

Let's address the elephant in the room: Atlassian is a tech company, and we use AI in our software.

At a basic level, our software is designed to help people work more efficiently and effectively. Machine learning can enhance that. Let's say you're about to share a Confluence page with a colleague. Our algorithm tracks who you interact with in Confluence most often, and will suggest people based on that data set as soon as you type the first letter of a name. This machine-learning engine, called "Smarts," delivers millions of notifications, suggestions, and reminders like this every day.

Simple though these examples may be, they give our Search & Smarts team something to measure and observe as they build more advanced functionality. Shihab Hamid and Timothy Clipsham, both senior technologists involved with "Smarts", say they're aware of the possibilities of bias in the data and the implications of their work overall. Thus, they're always on the lookout for unintended consequences.

Hamid illustrates a hypothetical scenario that, without attention to bias, could arise in a system like Jira. "Maybe you always assign high-severity issues to the most senior engineers. And across the data set, those senior engineers are overwhelmingly male. That will introduce a certain bias." Then, when the system encounters a more gender-balanced set of senior engineers, it may initially lag behind in suggesting critical work be assigned to the women in that group.

For the moment, Atlassian's use of AI is far removed from hiring decisions and equitable lending. So what keeps techies like Hamid and Clipsham up at night? "The way I see it, AI is all about predicting what's going to happen to the future," says Clipsham. "And machine learning is one technique that can help you predict what's going to happen." He emphasized, however, that the stakes would go up if we were to, say, auto-generate in-product avatars for users, because the data needed to accomplish that reaches into the dicey territory of gender and ethnicity.

How can we accept the good and filter out the bad?

According to Dr. Wallace, artificial intelligence should deliver three primary benefits:

1. Make things faster, bigger, and cheaper.
2. Be more accurate than a human brain when it comes to analytics and decision making.
3. Be more reliable than humans, both in terms of error rates and uptime. (Bots don't need to take sick days, after all.)

But, as we've seen in the examples above and many others, AI can also break things faster and in bigger ways by orders of magnitude with literally life-and-death implications. It can't factor values or other qualitative considerations into the decisions it makes. And although it's often better on accuracy, AI can be far worse on discrimination. Clearly, there's some work to do – and without delay.

If the first step toward change is awareness, then it's our collective responsibility to be as educated on this topic as we can. We use computers, we use social media. We must understand how they are, in many senses, using us, too.

AI is all around us, and it has fundamentally changed our lives in many positive ways. But it has the potential to be a major threat to the world, too. AI mimics human intelligence, and, in the most advanced forms, can learn on its own with minimal programming. Many engineers attest to the fact that, with many programs, there's a "sit back and see what the machine does" element to it all.

There's also growing recognition that we need better representation of women and people of color on AI and robotics teams. Professional groups like Black In Computing and the Algorithmic Justice League are raising awareness of the impact this overwhelmingly white- and male-dominated field can have (and *is* having) on communities of color. Meanwhile, organizations like Black Girls Code and Code2040 are working to bring more engineers from underrepresented groups into the hiring pipeline.

The machines, and very intelligent ones at that, are here and here to stay. But we can do our due diligence, remain vigilant, and make peace with the robots. Bookmark these resources and check them out the next time you need a break from your Facebook feed.

This article first appeared on Atlassian.