

The ethics of artificial intelligence

27 October 2021, by Joe Petrik



Maura R. Grossman. Credit: University of Waterloo

Maura R. Grossman, JD, Ph.D., is a Research Professor in the Cheriton School of Computer Science, an Adjunct Professor at Osgoode Hall Law School, and an affiliate faculty member of the Vector Institute for Artificial Intelligence. She is also Principal at Maura Grossman Law, an eDiscovery law and consulting firm in Buffalo, New York.

Maura is best known for her work on technology-assisted review, a supervised machine learning approach that she and her colleague, Computer Science Professor Gordon V. Cormack, developed to expedite review of documents in high-stakes litigation. She teaches Artificial Intelligence: Law, Ethics, and Policy, a course for graduate computer science students at Waterloo and upper-class law students at Osgoode, as well as the ethics workshop required of all students in the master's programs in [artificial intelligence](#) and [data science](#) at Waterloo.

What is AI?

Artificial intelligence is an umbrella term first used at a conference in Dartmouth in 1956. AI means computers doing intelligent things—performing cognitive tasks such as thinking, reasoning, and predicting—that were once thought to be the sole province of humans. It's not a single technology or

function.

Generally, AI involves algorithms, machine learning, and natural language processing. By algorithms we simply mean a set of rules to solve a problem or perform a task.

There are basically two types of AI, though some people believe there are three. The first is narrow or weak AI. This kind of AI does some task at least as well as, if not better than, a human. We have AI technology today that can read an MRI more accurately than a radiologist can. In my field of law, we have technology-assisted review—AI that can find legal evidence more quickly and accurately than a lawyer can. Other examples are programs that play chess or AlphaGo better than top players.

The second type is general or strong AI; this kind of AI would do most if not all things better than a human could. This kind of AI doesn't yet exist and there's debate about whether we'll ever have strong AI. The third type is super intelligent AI, and that's really more in the realm of science fiction. This type of AI would far outperform anything humans could do across many areas. It's obviously controversial though some see it as an upcoming existential threat.

Where is AI being used?

AI is used in countless areas.

In healthcare, AI is used to detect tumors in MRI scans, to diagnose illness, and to prescribe treatment. In education, AI can evaluate teacher performance. In transportation, it's used in autonomous vehicles, drones, and logistics. In banking, it's determining who gets a mortgage. In finance, it's used to detect fraud. Law enforcement uses AI for facial recognition. Governments use AI for benefits determination. In law, AI can be used to examine briefs parties have written and look for missing case citations.

AI has become interwoven into the fabric of society

and its uses are almost endless.

What is ethical AI?

AI isn't ethical, just as a screwdriver or a hammer isn't ethical. AI may be used in ethical or unethical ways. What AI does, however, is raise several ethical issues.

AI systems learn from past data and apply what they have learned to new data. Bias can creep in if the old data that's used to train the [algorithm](#) is not representative or has systemic bias. If you're creating a skin cancer detection algorithm and most of the training data was collected from White males, it's not going to be a good predictor of skin cancer in Black females. Biased data leads to biased predictions.

How features get weighted in algorithms can also create bias. And how the developer who creates the algorithm sees the world and what that person thinks is important—what features to include, what features to exclude—can bring in bias. How the output of an algorithm is interpreted can also be biased.

How has AI been regulated, if at all?

Most regulation so far has been through "soft law"—ethical guidelines, principles, and voluntary standards. There are thousands of soft laws and some have been drafted by corporations, industry groups, and professional associations. Generally, there's a fair degree of consensus as to what would be considered proper or acceptable use of AI—for example, AI shouldn't be used in harmful ways to perpetuate bias, AI should have some degree of transparency and explainability, it should be valid and reliable for its intended purpose.

The most comprehensive effort to date to generate a law to govern AI was proposed in April 2021 by the European Union. This draft EU legislation is the first comprehensive AI regulation. It classifies AI into risk categories. Some uses of AI are considered unacceptably high risk and they tend to be things like using AI to manipulate people psychologically. Another prohibited use is AI to determine social scores, where a person is

monitored and gets points for doing something desirable and loses points if doing something undesirable. A third prohibited use is real-time biometric surveillance.

The next category are high-risk AI tools like those used in medicine and self-driving vehicles. A company must meet all sorts of requirements, conduct risk assessments, keep records, and so on before such AI can be used. Then there are low-risk uses, such as web chatbots that answer questions. Such AI requires transparency and disclosure, but not much else.

Can AI conform to human values or social expectations?

It's very difficult to train an algorithm to be fair if you and I cannot agree on a definition of fairness. You may think that fairness means the algorithm should treat everyone equally. I might believe that fairness means achieving equity or making up for past inequities.

Our human values, cultural backgrounds, and social expectations often differ, leaving it difficult to determine what an algorithm should optimize. We simply don't have consensus yet.

In machine learning, we often don't know what the system is doing to make decisions. Are transparency and explainability in AI important?

That's a difficult question to answer. There is definitely something to be said for transparency and explainability, but in many circumstances it may be good enough if the AI has been tested sufficiently to show that it works for its intended purpose. If a doctor prescribes a drug, the biochemical mechanism of action may be unknown, but if the medication has been proven in clinical trials to be safe and effective, that may be enough.

Another way to look at this is, if we choose to use less sophisticated AI that we can more easily explain, but it is not as accurate or reliable than a more opaque algorithm, would that be an acceptable tradeoff? How much accuracy are we willing to give up in order to have more transparency and explainability?

It may depend on what the algorithm is being used for. If it's being used to sentence people, perhaps explainable AI matters more. In other areas, perhaps accuracy is the more important criterion. It comes down to a value judgment.

Provided by University of Waterloo

APA citation: The ethics of artificial intelligence (2021, October 27) retrieved 17 January 2022 from <https://techxplore.com/news/2021-10-ethics-artificial-intelligence.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.