psychologytoday.com

How to Build Ethical Artificial Intelligence

6-7 minutes

The field of artificial intelligence is exploding with projects such as IBM Watson, DeepMind's AlphaZero, and voice recognition used in virtual assistants including Amazon's Alexa, Apple's Siri, and Google's Home Assistant. Because of the increasing impact of AI on people's lives, concern is growing about how to take a sound ethical approach to future developments. Building ethical artificial intelligence requires both a moral approach to building AI systems and a plan for making AI systems themselves ethical. For example, developers of self-driving cars should be considering their social consequences including ensuring that the cars themselves are capable of making ethical decisions.

Ethical Questions

Here are some major issues that need to be considered.

1. Should we be worried about the prospect of machines becoming more intelligent than humans, and what can we do about it?

2. What needs to be done to prevent new AI applications from creating mass unemployment?

3. How can an AI application such as face recognition be used

for social control in ways that restrict the privacy and freedom of human beings?

4. How can AI systems increase or possibly decrease social biases and inequality?

5. What are the harms associated with the development of killer robots?

We need a general approach to ethics that can help to answer such questions.

Ethical Challenges

Applying ethics to artificial intelligence is difficult because of the lack of a generally accepted ethical framework. Here are some of the challenges that need to be dealt with to come up with ethical AI.

1. Ethical theories are highly controversial. Some people prefer ethical principles established by religious texts such as the Bible or the Quran. Philosophers argue about whether ethics should be based on rights and duties, on the greatest good for the greatest number of people, or on acting virtuously.

2. Acting ethically requires satisfying moral values, but there is no agreement about which values are appropriate or even about what values are. Without an account of the appropriate values that people use when they act ethically, it is impossible to align the values of AI systems with those of humans.

3. To build an AI system that behaves ethically, ideas about values and right and wrong need to be made sufficiently precise that they can be implemented in algorithms, but precision and algorithms are sorely lacking in current ethical deliberations.

Ethical Plan

Fortunately, my book Natural Philosophy presents an account of ethics that can meet these challenges.

1. I argue that the most plausible ethical theory is one that evaluates actions based on the extent to which they satisfy the vital needs of human beings. Vital needs are ones that are required for human lives and are distinguished from casual wants such as desiring a fancy car. Vital needs include not only biological needs such as food, water, and shelter, but also evidence-based psychological needs such as autonomy, relatedness to other people, and competence to achieve personal and social goals.

2. Accordingly, the appropriate values to be taken into account in ethical decisions are these vital human needs. The justification of such values comes not from religious texts or pure reason, but from empirical research that shows that these needs are in fact crucial to human lives.

3. Evaluating different actions with respect to how well they accomplish different needs for different people is an extraordinarily complex process, but it can be performed by algorithms that balance a multitude of constraints based on which actions satisfy most needs for most people. Such algorithms can be efficiently computed by neural networks and other methods.

Ethical Procedure

Accordingly, I propose the following ethical procedure to be carried out by people making decisions about the development of AI. Moreover, this procedure could be implemented in actual machines

1. List the alternative actions that are worth considering in a

particular situation. The ethical deliberation will assess these actions and choose based on moral considerations, not just on personal preferences. For example, government officials can consider whether or not to make military robots more intelligent and autonomous.

2. Identify all the people affected by these actions, including future generations as well as people currently alive. For killer robots, consider people who might be saved as well as ones that would be killed.

3. For each action, assess the extent to which it helps to promote or impede the satisfaction of human vital needs. For killer robots, the consequences to be considered include the survival and other needs of all the people potentially affected by intelligent weapons.

4. Translate the promotion of needs by actions into positive constraints and translate incompatibilities between actions into negative constraints. The result is a large constraint satisfaction network that can be evaluated computationally.

5. Maximize constraint satisfaction and choose the actions that do the best job of satisfying human needs.

Need Not Greed

Mahatma Gandhi said, “The world has enough for everyone’s needs, but not everyone’s greed." Like other ethical decisions, decisions about AI and ones made by intelligent machines should operate in the service of human needs. Currently, too many decisions are made in the service of greed for wealth and greed for power. Ethical AI can flourish if it puts universal human needs ahead of personal greed.