August 28, 2017
Blog, Opinion, Special Edition on Artificial Intelligence

Hold Artificial Intelligence Accountable

by Chamith Fonseka
figures by Rebecca Senft

The concept of artificial intelligence (AI) tends to evoke images of self-aware computers and robots – Knight Rider, Wall-E, the Terminator – but for the most part, this technology is, for now, restricted to fiction and film. In that time, however, artificial intelligence has already become embedded in everyday life, playing a role in everything from online shopping to criminal sentencing. Conversely, the development of AI policy has lagged far behind the development of AI. Without regulation or transparency, proprietary algorithms are used to make important, life-altering decisions without independent oversight. While a lack of transparency and regulation may not be a big deal when it comes to an algorithm incorrectly guessing what movie you might want to watch next, it can become a major issue in a world where AI technology is in use everywhere.

Invisible AI

The ability to easily integrate AI technology throughout the economy has led to the development of what I call "invisible AI." Unlike self-driving cars or automated personal assistants, systems that use invisible AI do not advertise that fact; instead, algorithms are used in the background to predict outcomes and make decisions. While the specific algorithms may differ, all AI systems can be thought of in a similar manner: A large amount of data is used to train the AI to recognize patterns, which is then used to classify new data. For example, an AI could be trained to differentiate between different fruits by giving it thousands of labeled images of apples and bananas. Then, when presented with an unlabeled image, the algorithm would compare it to the images it has seen before and determine whether the new image looks more like an apple or a banana. Although this may seem like a simple example, the general principle underlies the ability of AI programs to perform complex tasks like facial recognition or understanding language.

One common use of invisible AI is price steering, where different prices are shown to different consumers for the same product. A study conducted at Northeastern University at 2014 found that some hotel and travel booking sites use an algorithm that determine what to charge consumers based upon their personal data, such as showing lower prices to users browsing with a mobile device or showing higher prices to consumers who rented more expensive hotel rooms in the past. While the study showed that price steering was generally limited to a small number of products, invisible AI can also be used to display options in different orders based upon customer data, which can dramatically affect pricing, as people will generally pick the first or second option they see that fits their criteria for purchase.

It's possible to consider price steering as no different from a shopkeeper charging higher prices to a customer who tends to spend a lot of money – unethical, perhaps, but not illegal. In this

example, however, invisible AI allows the metaphorical shopkeeper to charge you based on the amount of money you spent somewhere else, or the places you like to hang out.

The problem of non-neutrality is further compounded by the opaqueness of machine learning and AI: it's often difficult or impossible to know *why* an algorithm has made a decision. Certain algorithms act as a "black box" where it is impossible to determine how the output was produced, while the details of other algorithms are considered proprietary trade secrets by the private companies who develop and market AI systems. Thus, we often find ourselves in situations in which neither the person using the algorithm nor the person being categorized by the algorithm can explain why certain determinations are made. A common claim among companies that use AI is that algorithms are effectively neutral actors that make decisions in an unbiased manner unlike humans who are affected by their prejudices. However, as machine-learning algorithms are typically trained to make decisions based on past data, the application of invisible AI is liable to reinforce existing inequalities.

Consider the job market, where hiring managers often rely on algorithms to identify potential candidates among a vast pool of applicants. Multiple studies have shown when presented with two applicants with identical resumes, people will preferentially select applicants with white-sounding names over those with black-sounding names. An algorithm trained such data might learn to prioritize applicants with white-sounding names without ever being explicitly directed to do so; and thus, judgments produced by this algorithm would be considered unbiased when they are anything but.

A particularly dangerous use of invisible AI is in the criminal justice system, where decisions are often life-changing. For example, after a person is convicted of a crime, a judge traditionally determines the length of the sentence based upon the facts of the case and the likelihood of reoffending. In some states, judges rely on risk assessments that are produced by algorithms to determine whether a given individual is low-risk or high-risk for continuing to commit crimes. Proponents argue that algorithms that use combination of factors – like the nature of the crime, records of previous offenses, personal history – present a fairer method of determining the likelihood of future criminal behavior. At face value, this appears to be reasonable, especially given the long history of racial disparity in the justice system.

Last year, reporters from ProPublica used court records in Florida to examine the accuracy of one of these systems, COMPAS (Correctional Offender Management Profiling for Alternative Sanctions,) and found that nearly 40% of people determined to be high-risk for reoffending did not do so, while others that were categorized as low-risk went on to commit other crimes. In fact, the algorithm was more likely to wrongly categorize white defendants as low-risk while black defendants were often incorrectly labeled as high-risk. Although the corporation that produces COMPAS disagreed with this analysis, they were unwilling to disclose their methods or explain how the algorithm took different factors into account when it generates risk scores.

The Right Debate

Popular discussions of ethical issues involving AI often focus on the potential of fully self-aware machines and how to ensure that AI will properly value human life. Technology leaders like

Elon Musk (Tesla) and Mark Zuckerberg (Facebook) have joined foundations and engaged in public debate about how to avoid creating self-aware AI that could be dangerous to human life; however, little debate has occurred on the ethical use of AI in its current form. Companies like Google and Amazon have used huge amounts of personal data to build advanced machine learning technology with little oversight or input from the public, while local, state, and federal governments have silently inserted AI into all kinds of decision-making processes. While the European Union has recently developed rules to regulate the transfer and use of personal data, no comparable legislation exists in the United States.

AI is often used as a substitute for human judgement and morality despite being fully unsuited for the purpose. Imagine a person driving a car down a highway who notices a turtle crossing the road at the last minute. Depending on personal judgement, some people might slow down and allow the animal to pass while others may swerve to avoid the animal and keep on driving. Now, consider the same scenario with a self-driving car. The AI that controls the car will handle this situation based on how it has been taught to value different objectives: is completing the trip on time more important than ensuring that the animal is unharmed? And how does it come to that decision? These are important ethical questions; yet, AI is an essentially unregulated technology.

AI will need be taught to make ethical and moral decisions, such as a self-driving car determining whether to avoid a turtle crossing the road.

New policies will not solve all of these challenges, but they can play an important role in ensuring that the benefits of AI are equitably distributed while acknowledging its limitations. An initial step in this direction could include the following suggestions:

- Require transparency and evaluation of AI algorithms. Understanding how an algorithm makes its decisions is critical for determining whether it is affected by bias or other factors; moreover, algorithms should be regularly tested to ensure that they are still working appropriately.
- Ensure that the objectives of AI technology are aligned with societal norms, such as making sure that AI systems used in criminal justice are optimized to give benefit of doubt to the innocent.
- Develop policies and regulations about how personal data can be used by government and businesses. Controlling how governments, companies, and individuals can use "big" data will protect against unwanted commercialization of their personal information. People should have the right to know how their activities – both on and off the internet – are being used by AI systems.

Although its limitations are significant, artificial intelligence has great promise for advancing the best interests of humanity if these technologies can be integrated in a fair and just manner.

Chamith Fonseka is a fifth year PhD candidate in the Biomedical and Biological Sciences program at Harvard University.

This article is part of a Special Edition on Artificial Intelligence.

For more information:

Cathy O'Neil

https://www.bloomberg.com/view/contributors/ATFPV0aLyJM/catherine-h-oneil

https://www.amazon.com/Weapons-Math-Destruction-Increases-Inequality/dp/0553418815

Zeynep Tufciki

https://www.nytimes.com/2016/05/19/opinion/the-real-bias-built-in-at-facebook.html

https://www.twitterandteargas.org/

Propublica COMPAS Investigation

https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Harvard Data Privacy Lab

https://dataprivacylab.org/