BROOKINGS

Report

The role of corporations in addressing AI's ethical dilemmas

Darrell M. WestThursday, September 13, 2018

Editor's Note:

This report is part of "A Blueprint for the Future of AI," a series from the Brookings Institution that analyzes the new challenges and potential policy solutions introduced by artificial intelligence and other emerging technologies.

The world is seeing extraordinary advances in artificial intelligence. There are new applications in finance, defense, health care, criminal justice, and education, among other areas.[1] Algorithms are improving spell-checkers, voice recognition systems, ad targeting, and fraud detection.

Yet at the same time, there is concern regarding the ethical values embedded within AI and the extent to which algorithms respect basic human values. Ethicists worry about a lack of transparency, poor accountability, unfairness, and bias in these automated tools. With millions of lines of code in each application, it is difficult to know what values are inculcated in software and how algorithms actually reach decisions.

As they push the boundaries of innovation, technology companies increasingly are becoming digital sovereigns that set the rules of the road, the nature of the code, and their corporate practices and terms of service.[2] In the course of writing software, their coders make countless decisions that affect the way algorithms operate and make decisions.[3]

The world is seeing extraordinary advances in artificial intelligence. Yet at the same time, there is concern regarding the ethical values embedded within AI and the extent to which algorithms respect basic human values.

In this paper, I examine five AI ethical dilemmas: weapons and military-related applications, law and border enforcement, government surveillance, issues of racial bias, and social credit systems. I discuss how technology companies are handling these issues and the importance of having principles and processes for addressing these concerns. I close by noting ways to strengthen ethics in AI-related corporate decisions.

Briefly, I argue it is important for firms to undertake several steps in order to ensure that AI ethics are taken seriously:

1. Hire ethicists who work with corporate decisionmakers and software developers
2. Develop a code of AI ethics that lays out how various issues will be handled
3. Have an AI review board that regularly addresses corporate ethical questions
4. Develop AI audit trails that show how various coding decisions have been made
5. Implement AI training programs so staff operationalizes ethical considerations in their daily work, and
6. Provide a means for remediation when AI solutions inflict harm or damages on people or organizations.

AI ethics

The growing sophistication and ubiquity of AI applications has raised a number of ethical concerns. These include issues of bias, fairness, safety, transparency, and accountability. Without systems compatible with these principles, the worry is that AI will be biased, unfair, or lack proper transparency or accountability.[4]

Concerns over possible problems have led many nongovernment, academic, and even corporate organizations to put forward declarations on the need to protect basic human rights in artificial intelligence and machine learning. These groups have outlined principles for AI development and processes to safeguard humanity.

In 2017, participants at a Future of Life conference held at Asilomar published a statement summarizing issues being raised by artificial intelligence and machine learning. They argued that "highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation." In addition, they proclaimed that "AI technologies should benefit and empower as many people as possible," and "the economic prosperity created by AI should be shared broadly, to benefit all of humanity."[5]

This was followed in 2018 by the "Toronto Declaration" organized by Amnesty International, Access Now, and other organizations. It focused on machine learning and laid out the basic principle that "states and private actors should promote the development and use of these technologies to help people more easily exercise and enjoy their human rights." Among the specific rights enumerated were to "protect individuals against discrimination, promote inclusion, diversity and equity, and safeguard equality."[6] To these ends, its signatories asked developers to identify risks, ensure transparency, enforce oversight, and hold private actors to account for their actions.

The growing sophistication and ubiquity of AI applications has raised a number of ethical concerns, including issues of bias, fairness, safety, transparency, and accountability.

A number of university projects also have focused on AI concerns. This includes efforts at Harvard University, the University of Oxford, Cambridge University, Stanford University, University of Washington, and elsewhere. Academic experts have pinpointed particular areas of concern and ways both government and business need to promote ethical considerations in AI development.[7]

Nonprofit organizations have been active in this space. For example, the Royal Society has undertaken a machine-learning project that analyzes the field's opportunities as well as barriers to be overcome. Its goal is "to ensure that machine learning can bring the maximum benefit to the greatest number of people."[8] The IEEE's Standards Association is working on an initiative for ethical considerations in the design of autonomous systems.

Other nonprofits are focusing on how to develop artificial general intelligence and mold it toward beneficial uses. Individuals, such as Sam Altman, Greg Brockman, Elon Musk, and Peter Thiel, as well as firms, such as Y Research, Infosys, Microsoft, Amazon, and the Open Philanthropy Project have joined forces to develop OpenAI as a nonprofit AI research company. It defines its mission as "discovering and enacting the path to safe artificial general

intelligence." Its engineers and scientists use open-source tools to develop AI for the benefit of the entire community and has protocols "for keeping technologies private when there are safety concerns."[9]

Corporations have joined in the discussions as well. For example, Google has published a document calling for the "responsible development of AI." It said AI should be socially beneficial, not reinforce unfair bias, should be tested for safety, should be accountable to people, should incorporate privacy design, should uphold high standards of scientific excellence, and should be available for uses that accord with those principles.[10]

Microsoft meanwhile published an extensive report on "the future computed." It laid out the opportunities for AI, the need for "principles, policies and laws for the responsible use of AI," and noted the possible ramifications for the future of jobs and work.[11]

Several companies have joined together to form the Partnership for Artificial Intelligence to Benefit People and Society. They include Google, Microsoft, Amazon, Facebook, Apple, and IBM. It seeks to develop industry best practices to guide AI development with the goal of promoting "ethics, fairness and inclusivity; transparency, privacy, and interoperability; collaboration between people and AI systems; and the trustworthiness, reliability and robustness of the technology."[12]

Political polarization and dual-use technologies

In looking across AI activities, there are several applications that have raised ethical concerns. It is one thing to support general goals, such as fairness and accountability, but another to apply those concepts in particular domains and under specific political conditions. One cannot isolate ethics discussions from the broader political climate in which technology is being deployed.

The current polarization around politics and policymaking complicates the tasks facing decisionmakers. Republicans and Democrats have very different views of U.S. officeholders, policy options, and political developments. Ethical issues that might not be very controversial during a time of normal politics become much more divisive when people don't like or trust the officials making the decisions.

It is not the technology so much that dictates the moral dilemma as the human use case involved with the application. The very same algorithm can serve a variety of purposes, which makes the ethics of decisionmaking very difficult.

In addition, running through many ethical dilemmas is the problem of dual-use technologies. There are many algorithms and software applications that can be used for good or ill. Facial recognition can be deployed to find lost children or facilitate widespread civilian surveillance.[13] It is not the technology so much that dictates the moral dilemma as the human use case involved with the application. The very same algorithm can serve a variety of purposes, which makes the ethics of decisionmaking very difficult.[14]

For this reason, companies have to consider not just the ethical aspects of emerging technologies, but also their possible use cases. Indeed, the latter represents an interesting opportunity to explore AI ethics because it illustrates concrete aspects of ethical dilemmas. Having in-depth knowledge of those issues is important for AI development.

Dilemma one: Weapons development and military applications

One topic that has attracted considerable attention involves AI applications devoted to war or military activities. As technology innovation has accelerated, there have been discussions regarding whether AI should be used in war-related activities. In its code of ethics, for example, Google wrote that it will not design or deploy AI in: "weapons or other technologies designed to cause or directly facilitate injury to people; in technologies that gather or use information for surveillance violating internationally accepted norms; or technologies for any purpose that contravene widely accepted principles of international law and human rights." To clarify the situation, its document also added, "[For any AI applications] where there is a material risk of harm, we will proceed only where we believe that the benefits substantially outweigh the risks, and will incorporate appropriate safety constraints."[15]

Of course, many other firms have not adopted this position. For example, Palantir has garnered at least $1.2 billion in federal contracts since 2009 through products popular with defense, intelligence agencies, homeland security, and law enforcement. One of its primary applications known as Gotham imports "large reams of structured data (like spreadsheets) and unstructured data (like images) into one centralized database, where all of the information can be visualized and

analyzed in one workspace."[16] The goal is to use technology to make military applications more efficient and effective, and help defense planners achieve their objectives in the field.

Indeed, military leaders long have recognized the need to upgrade capabilities and incorporate the latest advances in their arsenals. The U.S. Department of Defense has set up a Joint Artificial Intelligence Center designed to improve "large-scale AI projects."[17] Its plan is to work with private companies and university researchers to make sure America takes advantage of the latest products for defense purposes.

During a period of considerable international turbulence and global threats, America has to be careful not to engage in unilateral disarmament when possible adversaries are moving full-speed ahead.

This is consistent with the urgings of Brookings President John Allen and business executive Amir Husain. They argue the world is moving towards "hyperwar," in which advanced capabilities will combine into rapid-style engagements based on physical and digital encounters.[18] As such, it is important for the United States to have the means to defend itself against possible AI-based attacks from adversaries.

Many commentators have noted that countries, such as Russia, China, Iran, and North Korea, have AI capabilities and are not refraining from deployment of high-tech tools. During a period of considerable international turbulence and global

threats, America has to be careful not to engage in unilateral disarmament when possible adversaries are moving full-speed ahead. Disputes over AI deployment demonstrate not all agree on an AI prohibition for national security purposes.

The American public understands this point. In an August 2018 survey undertaken by Brookings researchers, 30 percent of respondents believed the United States should develop AI technologies for warfare, 39 percent did not, and 31 percent were undecided. However, when told that adversaries already are developing AI for war-related purposes, 45 percent thought America should develop these kinds of weapons, 25 percent did not, and 30 percent were undecided.[19]

There are substantial demographic differences in these attitudes. Men (51 percent) were much more likely than women (39 percent) to support AI for warfare if adversaries develop these kinds of weapons. The same was true for senior citizens (53 percent) compared to those aged 18 to 34 (38 percent).

Dilemma two: Law and border enforcement

In the domestic policy area, there are similar concerns regarding the militarization of policing practices and shootings of unarmed black men in communities across the U.S. Those tendencies have led some to decry AI applications in law enforcement. Critics worry that emerging technologies, such as facial recognition software, unfairly target minorities and lead to biased or discriminatory enforcement, sometimes with tragic consequences.

Some business leaders have been quite outspoken on this topic. For example, Brian Brackeen, the chief executive officer of facial recognition firm Kairos, argues that its usage "opens the door for gross misconduct by the morally corrupt." He discusses the history of law enforcement against American minorities and concludes, "There is no place in America for facial recognition

that supports false arrests and murder." Speaking on behalf of his company, Brackeen says his firm will not work with government agencies and says, "Any company in this space that willingly hands this software over to a government, be it America or another nation's, is willfully endangering people's lives."[20]

The same logic applies to border enforcement under the current administration. With President Donald Trump's crackdown on undocumented arrivals, employees at some firms have complained about contracts with the Immigration and Customs Enforcement agency that is charged with enforcing administration decisions.[21] They object to Trump's policies and argue technology firms should not enable that crackdown by providing technologies for border enforcement. McKinsey & Company already have announced it no longer will work with Immigration and Customs Enforcement and Customs and Border Protection due to employee objections to enforcement actions at those agencies.[22]

Dilemma three: Government surveillance

Government surveillance is a challenge in many places. A number of countries have turned toward authoritarianism in recent years. They have shut down the internet, attacked dissidents, imprisoned reporters or NGO advocates, and attacked judges. All of these activities have fueled concerns regarding government use of technology to surveil or imprison innocent people.

As a result, some companies have disavowed any interest in selling to government agencies. As an illustration, CEO Rana el Kaliouby of Affectiva, an AI firm that works on image recognition, has turned down such opportunities. "We're not interested in applications where you're spying on people," he announced. This includes security agencies, airport authorities, or lie detection contracts.[23]

Government surveillance is a challenge in many places. A number of countries have turned toward authoritarianism in recent years, fueling concerns regarding government use of technology to surveil or imprison innocent people.

In addition, Microsoft has argued facial recognition is to "be left up to tech companies." Company President Brad Smith says this software "raises issues that go to the heart of fundamental human rights protections like privacy and freedom of expression." As a result, he supports "a government initiative to regulate the proper use of facial recognition technology, informed first by a bipartisan and expert commission."[24]

Other companies, however, have not taken this stance. Amazon sells its Rekognition facial recognition software to police agencies and other kinds of government units, even though some of its employees object to the practice.[25] It has the view that government authorities should have access to the latest technologies. But the firm has announced that "it will suspend ... customer's right to use ... services [like Rekognition] if it determines those services are being abused."[26]

In China, there is growing use of facial recognition combined with video cameras and AI to keep track of its own population. There, law enforcement scans people at train stations to find wanted people or identifies jaywalkers who cross the street illegally. It is estimated that the country has deployed 200 million video

cameras, which makes possible surveillance on an unprecedented scale.[27] When combined with AI analysis that matches images with personal identities, the capacity for in-depth population control is enormous.

In his analysis of the ethics of facial recognition software, Brookings scholar William Galston points out there should be "a reasonable expectation of anonymity." Government authorities should not deploy such technology unless there is "a justification weighty enough to override the presumption against doing so," and that "this process should be regulated by law ... [and] the equivalent of a search warrant."[28] In his view, having clear legal standards is vital in order to prevent likely abuses.

Dilemma four: Racial bias

There is considerable evidence of racial biases in facial recognition software. Some systems have "misidentified darker-skinned women as often as 35 percent of the time and darker-skinned men 12 percent of the time," much higher than the rates for Caucasians.[29]

Most systems operate by comparing a person's face to a range of images in a large database. As pointed out by Joy Buolamwini of the Algorithmic Justice League, "If your facial recognition data contains mostly Caucasian faces, that's what your program will learn to recognize."[30] Unless the databases have access to diverse data, these programs perform poorly when attempting to recognize African-American or Asian-American features.

There is considerable evidence of racial biases in facial recognition software.

Many historical data sets reflect traditional values, which may or may not represent the preferences wanted in a current system. As Buolamwini notes, such an approach risks repeating inequities of the past:

The rise of automation and the increased reliance on algorithms for high-stakes decisions such as whether someone gets insurance or not, your likelihood to default on a loan or somebody's risk of recidivism means this is something that needs to be addressed. Even admissions decisions are increasingly automated—what school our children go to and what opportunities they have. We don't have to bring the structural inequalities of the past into the future we create.

This is one of the reasons why it is important to increase data openness so AI developers have access to large data sets for training purposes. They need unbiased information in order to instruct AI systems properly on how to recognize certain patterns and make reasonable decisions. Governments can help

in this regard by promoting greater access to their information.[31] They have some of the largest data sets, and this information can be a valuable resource for training AI and overcoming past problems.

In addition, in sensitive areas, such as criminal justice—where inaccuracies can lead to higher incarceration rates—there need to be minimum standards of accuracy for facial recognition software to be deployed. Systems should certify what their rates are so officials understand what possible biases come with AI deployment. Jennifer Lynch of the Electronic Frontier Foundation argues that "an inaccurate system will implicate people for crimes they didn't commit and shift the burden to innocent defendants to show they are not who the system says they are."[32]

Dilemma five: Social credit systems

China is expanding its use of social credit systems for daily life. It compiles data on people's social media activities, personal infractions, and paying taxes on time, and uses the resulting score to rate people for credit-worthiness, travel, school enrollment, and government positions.[33] Those with high scores are accorded special discounts and privileges, while those who fare more poorly can be banned from travel, refused enrollment at favored schools, or restricted from government employment.

The problem with these systems depends on their opacity. As noted by Jack Karsten and me in a blog post, "It is not clear what factors affect someone's score, and so those with a low score may face exclusion without knowing why."[34] In addition, given inequitable access to activities that promote higher scores, such systems can increase disparities based on socio-economic background, ethnic category, or education level. Authoritarian regimes may turn to AI to support their interest in controlling the population.

Recommendations for going forward

It is not easy to resolve any of the ethical issues surrounding the topics discussed above. Each of them raises important ethical, legal, and political concerns, and therefore are not amenable to easy resolution. Leaders dealing with these challenges will have to take considerable time and energy to work through the substantive issues.

But there are organizational and procedural mechanisms that help with some of these ethical dilemmas. Having clear processes and avenues for deliberation would help deal with particular problems. There are a number of steps that would help firms ensure fair, safe, and transparent AI applications.

As William Galston suggests, if these reforms prove inadequate, there may need to be government legislation to mandate appropriate safeguards.[35] Improving protections in the areas of racial bias and discrimination are especially important. In addition, resolving how the United States wants to handle technology for warfare is crucial.

1. Hiring company ethicists

It is important for companies to have respected ethicists on their staffs to help them think through the ethics of AI development and deployment. Giving these individuals a seat at the table will help to ensure that ethics are taken seriously and appropriate deliberations take place when ethical dilemmas arise, which is likely to happen on a regular basis. In addition, they can assist corporate leadership in creating an AI ethics culture and supporting corporate social responsibility within their organizations. These ethicists should make annual reports to their corporate boards outlining the issues they have addressed during the preceding year and how they resolved ethical aspects of those decisions.

2. Having an AI code of ethics

Companies should have a formal code of ethics that lays out their principles, processes, and ways of handling ethical aspects of AI development. Those codes should be made public on the firm's websites so that stakeholders and external parties can see how the company thinks about ethical issues and the choices its leaders have made in dealing with emerging technologies.

3. Instituting AI review boards

Businesses should set up internal AI review boards that evaluate product lines and are integrated into company decisionmaking. These boards should include a representative cross-section of firm stakeholders and be consulted on AI-related decisions. Their portfolio should include development of particular product lines, the procurement of government contracts, and procedures used in developing AI products.

4. Requiring AI audit trails

Companies should have AI audit trails that explain how particular algorithms were put together or what kinds of choices were made during the development process. This can provide some degree of "after-the-fact" transparency and explainability to outside parties. Such tools would be especially relevant in cases that end up under litigation and need to be elucidated to judges or juries in case of consumer harm. Since product liability law is likely to be the governing force in adjudicating AI harm, and it is necessary to have audit trails that provide both external transparency and explainability.

5. Implementing AI training programs

6. Providing a means of remediation for AI damages or harm

There should be a means of remediation in case AI deployment results in consumer damages or harm.[36] This could be through legal cases, arbitration, or some other negotiated process. This would allow those hurt by AI to address the problems and rectify the situation. Having clear procedures in place will help when disasters strike or there are unanticipated consequences of emerging technologies.

Public support for these recommendations

Survey data indicate there is substantial support for these actions. An August 2018 Brookings survey found: 55 percent of respondents supported the hiring of corporate ethicists; 67 percent favored companies having a code of ethics; 66 percent believed companies should have an AI review board; 62 percent thought software designers should compile an AI audit trail that shows how they made coding decisions; 65 percent favored the implementation of AI training programs for company staff; and 67 percent wanted companies to have mediation procedures when AI solutions inflict harm or damages on people.[37]

Individuals want companies to take meaningful action to protect them from unfairness, bias, poor accountability, inadequate privacy protection, and a lack of transparency. If those steps fail, legislation will become the likely recourse.

The strong public support for these steps indicates people understand the ethical risks posed by artificial intelligence and emerging technologies, as well as the need for significant action by technology-based organizations. Individuals want companies to take meaningful action to protect them from unfairness, bias, poor accountability, inadequate privacy protection, and a lack of transparency. If those steps fail, legislation will become the likely recourse.

Report Produced by **Center for Technology Innovation**

Footnotes

1. 1 Darrell M. West and John R. Allen, "How Artificial Intelligence is Transforming the World," Brookings Institution report, April 24, 2018.
2. 2 John R. Allen, "Remarks Delivered at the North America Think Tank Summit," April 12, 2018.
3. 3 Darrell M. West, *The Future of Work: Robots, AI, and Automation*, Brookings Institution Press, 2018.
4. 4 Amir Khosrowshahi, "Game Changers: Artificial Intelligence," testimony before the U.S. House of Representatives Committee on Oversight and Government Reform, Subcommittee on Information Technology, February 14, 2018.
5. 5 Asilomar Future of Life Institute, "AI Principles," 2017.
6. 6 Anna Bacciarelli, Joe Westby, Estelle Masse, Drew Mitnick, Fanny Hidvegi, Boye Adegoke, Frederike Kaltheuner, Malavika Jayaram, Yasodara Cordova, Solon Barocas, and William Isaac, "The Toronto Declaration: Protecting the Rights to Equality and Non-Discrimination in Machine Learning Systems," May 17, 2018.
7. 7 University of Oxford, "Towards a Code of Ethics for AI," undated.
8. 8 Peter Donnelly, "Machine Learning: The Power and Promise of Computers That Learn by Example," The Royal Society Working Group, undated.
9. 9 Open AI website at https://openai.com/about/.

10. 10 Google, "Responsible Development of AI," 2018.
11. 11 Microsoft, "The Future Computed: Artificial Intelligence and Its Role in Society," 2018.
12. 12 Alex Hern, "'Partnership on AI' Formed by Google, Facebook, Amazon, IBM and Microsoft," *The Guardian*, September 28, 2016.
13. 13 Natasha Singer, "Facebook's Push for Facial Recognition Prompts Privacy Alarms," *New York Times*, July 9, 2018.
14. 14 John Deutch, "Is Innovation China's New Great Leap Forward?", *Issues in Science and Technology*, Summer, 2018.
15. 15 Google, "Responsible Development of AI," 2018, p. 5.
16. 16 Sam Biddle, "How Peter Thiel's Palantir Helped The NSA Spy on the Whole World," *The Intercept*, February 22, 2017.
17. 17 Jade Leung and Sophie-Charlotte Fischer, "Pentagon Debuts Artificial Intelligence Hub, *Bulletin of the Atomic Scientists*, August 8, 2018.
18. 18 John R. Allen and Amir Husain, "On Hyperwar," *Naval Institute Proceedings*, July 17, 2017.
19. 19 Darrell M. West, "Brookings Survey Finds 62 Percent Believe Artificial Intelligence Should Be Guided by Human Values," Brookings Institution, August 29, 2018.
20. 20 Brian Brackeen, "Facial Recognition Software is Not Ready for Use by Law Enforcement," *Tech Crunch*, June 25, 2018.
21. 21 Farhad Manjoo, "Employee Uprisings Sweep Many Tech Companies. Not Twitter," *New York Times*, July 4, 2018.
22. 22 Michael Forsythe and Walt Boghanich, "Citing 'Moral Objections,' Employees Call on Deloitte to Sever Ties With ICE," *New York Times*, July 13, 2013.
23. 23 Matt O'Brien, "How Much All-Seeing AI Surveillance Is Too Much?", *Associated Press*, July 3, 2018.
24. 24 Rob Pegoraro, "Microsoft Argues Facial-Recognition Tech Could Violate Your Rights, *Yahoo*, July 27, 2018.
25. 25 Matt O'Brien, "How Much All-Seeing AI Surveillance Is Too Much?", *Associated Press*, July 3, 2018.
26. 26 Kyle Wiggers, "The Growing Importance of Clear AI Ethics Policies," *Venture Beat*, June 22, 2018.
27. 27 Paul Mozur, "Inside China;'s Dystopian Dreams: A.I., Shame and Lots of Cameras," *New York Times*, July 8, 2018.
28. 28 William Galston, "Why the Government Must Help Shape the Future of AI," Brookings Institution paper, September 13, 2018.
29. 29 Lizette Chapman and Joshua Brustein, "A.I. Has a Race Problem," *Bloomberg Businessweek*, July 2, 2018.
30. 30 "Joy Buolamwini," *Bloomberg Businessweek*, July 3, 2017, p. 80.
31. 31 Natasha Lomas, "AI Report Fed By DeepMind, Amazon, Uber Urges Greater Access to Public Sector Data Sets," *Tech Crunch*, April 24, 2017.
32. 32 Lizette Chapman and Joshua Brustein, "A.I. Has a Race Problem," *Bloomberg Businessweek*, July 2, 2018.
33. 33 Jack Karsten and Darrell M. West, "China's Social Credit System Spreads to More Daily Transactions," Brookings Institution TechTank blog, June 18, 2018.
34. 34 Ibid.
35. 35 William Galston, "Why the Government Must Help Shape the Future of AI," Brookings Institution paper, September 13, 2018.
36. 36 Intel, "Artificial Intelligence: The Public Policy Opportunity," 2017.

37. 37 Darrell M. West, "Brookings Survey Finds 62 Percent Believe Artificial Intelligence Should Be Guided by Human Values," Brookings Institution, August 29, 2018.