Deloitte. Insights

Can AI be ethical? Why enterprises shouldn't wait for AI regulation

David Schatsky, Vivek Katyal, Satish Iyengar, and Rameeta Chauhan

With AI applications becoming ubiquitous in and out of the workplace, can the technology be controlled to avoid unintended or adverse outcomes? Organizations are launching a range of initiatives to address ethical concerns.

WITH GREAT POWER, the saying goes, comes great responsibility. As artificial intelligence (AI) technology becomes more powerful, many groups are taking an interest in ensuring its responsible use. The questions that surround AI ethics can be difficult, and the operational aspects of addressing AI ethics are complex. Fortunately, these questions are already driving debate and action in the public and commercial sectors. Organizations using AI-based applications should take note.

Signals

- In 2018, media mentions around AI and ethics doubled as compared to the previous year, with over 90 percent indicating positive or neutral sentiment.[1]

- About a third of executives in a recent Deloitte survey named ethical risks as one of the top three potential concerns related to AI.[2]

- Since 2017, more than two dozen national governments have released AI strategies, road maps, or plans that focus on developing ethics standards, policies, regulations, or frameworks.[3]

- Governments are setting up AI ethics councils or task forces and collaborating with other national governments, corporations, and other organizations on the ethics of AI.[4]

- Major technology companies such as Google, IBM, and Facebook have developed tools, designed guidelines, and appointed dedicated AI governance teams to address ethical issues, such as bias and lack of transparency.[5]

- Enterprises across industries such as financial services, life sciences and health care, retail, and media are joining consortia to collaborate with technology vendors, universities, governments, and other players in their respective industries to promote ethical AI standards and solutions.[6]

As adoption of increasingly capable AI grows, ethics concerns emerge

A growing number of companies see AI as critical to their future. But concerns about possible misuse of the technology are on the rise. In a recent Deloitte survey, 76 percent of executives said they expected AI to "substantially transform" their companies within three years,[7] while about a third of respondents named ethical risks as one of the top three concerns about the technology. The press has widely reported incidents in which AI has been misused or had unintended consequences.[8]

The conversation about responsible AI is hardly limited to concerns about controversial applications of the technology, such as automated weapons. It also considers how the infusion of AI into common activities such as social media interactions, credit decisions, and hiring can be controlled to avoid unintended or adverse outcomes for individuals and businesses. The discussion around AI and ethics has grown far more urgent in the last decade or so, and many initiatives to tackle ethical questions surrounding AI have taken shape in the last couple of years. This is primarily driven by recent

advancements in AI technologies, growing adoption, and increasing criticality of AI in business decision-making.

It's worth noting that concerns about the ethics of technology generally and AI specifically are nothing new. The topic was explored at least as far back as 1942 when science-fiction writer Isaac Asimov introduced his Three Laws of Robotics in a short story.[9] In 1976, a German-American computer scientist suggested that AI technology should not be used to replace people in positions that require abilities such as compassion, intuition, and creativity.[10] Still, today's AI presents enormous opportunities for businesses while introducing some novel risks that need to be managed.

AI systems may pose diverse ethical risks

Some of the ethical risks associated with AI use differ from those associated with conventional information technology. This is due to a variety of factors, including the role played by large datasets in AI systems, the novel applications of AI technology (such as facial recognition), and the capabilities that some systems demonstrate, from automatic learning to superhuman perception. As MIT professors Stefan Helmreich and Heather Paxson note, "Ethical judgments are built into our information infrastructures themselves. That's what AI does: It automates judgments—yes, no; right, wrong."[11] Prominent issues associated with ethical AI design, development, and deployment include the following:

Bias and discrimination. AI systems learn from the datasets with which they are trained. Depending on how a dataset is compiled or constructed, the potential exists that the data could reflect assumptions or biases—such as gender, racial, or income biases—that could influence the behavior of a system based on that data. These systems' developers intend no bias, but many have reported AI-driven instances of bias or discrimination in application areas such as recruiting, credit scoring,

and judicial sentencing.[12] Organizations need to ensure that their AI solutions make decisions fairly and do not propagate biases when providing recommendations.

Lack of transparency. It is natural for customers or other parties affected by technology to want to know something about how the system that affected them works—what data it is using and how it is making decisions. However, much AI development entails building highly effective models whose inner workings are not well understood and cannot be readily explained—they are black boxes. Techniques are emerging that help shine light inside the black box of certain machine learning models, making them more interpretable and accurate, but they are not suitable for all applications.[13] Ethical AI use takes into account a responsibility to be transparent about the workings of systems and the use of data wherever possible.

Erosion of privacy. Many companies collect large quantities of personal data from consumers when they register for or use products or services. That data can be used to train AI-based systems for purposes such as targeted advertising and promotions and personalization. Ethical issues arise when that information is used for a different purpose—say, to train a model for making employment offers—without users' knowledge or consent. A recent study highlighted that 60 percent of customers are concerned about AI-based technology compromising their personal information.[14] To build customer trust, companies need to be transparent about how collected information is being used, create clearer mechanisms for consent, and better protect individual privacy.

Poor accountability. With AI technologies increasingly automating the decision-making process for a wide range of critical applications, such as autonomous driving, disease diagnosis, and wealth management, the question arises who should bear responsibility for the harm with which these AI systems may be associated. For instance, if a self-driving car doesn't stop after seeing a pedestrian and hits the individual, who should be held responsible: the car manufacturer, passenger, or owner?

Existing accountability mechanisms for IT systems fail to adequately address such scenarios. Businesses, governments, and the public need to work toward establishing proper accountability structures.

Workforce displacement and transitions. Companies are already using AI to automate tasks, with some aiming to take advantage of automation to reduce their workforces.[15] In the 2018 Deloitte executive survey, 36 percent of respondents saw job cuts from AI-driven automation rising to the level of an ethical risk. Even jobs that are not eliminated may be impacted in some way by AI. Employers should find ways to use AI to increase opportunities for employees while mitigating negative impacts.

The marketplace is taking action on AI ethics

The increasing adoption of AI technologies, and growing awareness of various ethical risks associated with them, calls for urgency in designing approaches and mechanisms to deal with those risks. Governments, technology vendors, corporates, academic institutions, and others have already started laying the foundation for ethical AI use.

TECH VENDORS AT THE FOREFRONT

Many of the technology vendors creating AI tools and platforms are also at the forefront of ethical AI development efforts. Major technology companies including Google and IBM have developed ethical guidelines to govern the use of AI internally as well as guide other enterprises.[16] For instance, while releasing its ethical guidelines, Google pledged to not develop AI specifically for weaponry, or for surveillance tools that would violate "internationally accepted norms."[17] Additionally, many technology vendors have launched or open-sourced tools to address ethical issues such as bias and lack of transparency in AI development and deployment. Examples include Facebook's Fairness Flow, IBM's AI Fairness 360 and AI OpenScale environment, and Google's What-if tool.[18]
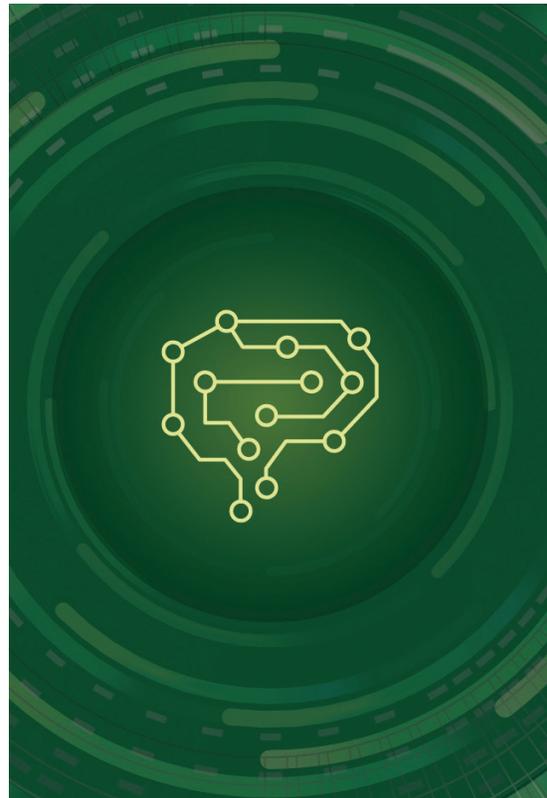
GOVERNMENTS AND REGULATORS ARE ALREADY VERY ACTIVE

Governments and regulators have already begun to play a crucial role in establishing policies and guidelines to tackle AI-related ethical issues. For instance, the European Union's General Data Protection Regulation (GDPR) requires organizations to be able to explain decisions made by their algorithms.[19] This is just an example from the growing list of national governments—such as the United States, the United Kingdom, Canada, China, Singapore, France, and New Zealand—that have released AI strategies, road maps, or plans focusing on developing ethical standards, policies, regulations, or frameworks.[20] Other notable government initiatives include setting up AI ethics councils or task forces, and collaborating with other national governments, corporates, and other organizations.[21] Though most of these efforts are still in initial phases and do not impose binding requirements on companies (with GDPR a prominent exception), they signal growing urgency about AI ethical issues.

ACADEMIA IS DRIVING RESEARCH AND EDUCATION

Universities and research institutions are playing an important role as well. Not only do they educate those who design and develop AI-based solutions—they are researching ethical questions and auditing algorithms for the public good. A number of universities, including Carnegie Mellon and MIT, have launched courses dealing specifically with AI and ethics.[22] MIT also created a platform called Moral Machine[23] to crowdsource data and effectively train self-driving cars to respond to a variety of morally fraught scenarios. Indeed, ethics was a central theme at the recent launch of MIT's new Schwartzman College of Computing.[24] More-over, academics are getting seats on AI governance teams at many technology companies and other enterprises as external advisers to help guide the responsible development of AI applications.[25]

NONPROFITS AND CORPORATES ARE ENGAGING ON AI ETHICS

Consortia and think tanks are bringing together technology companies, governments, nonprofit organizations, and academia to collaborate on a complex and evolving set of AI-related ethical issues, leverage each other's expertise and capabilities, and simultaneously build the AI ecosystem. One such consortium is the Partnership on AI, which counts 80-plus partner organizations.[26] Companies across sectors are working to adopt ethical AI practices such as establishing ethics boards and retraining employees, and professional services firms are guiding clients on these issues.[27]

How companies can make AI ethics a priority

Technological progress tends to outpace regulatory change, and this is certainly true in the field of AI. But organizations may not want to wait for AI-related regulation to catch up. To protect their stakeholders and their reputations, and to fulfill their ethical commitments, organizations can do many things now as they design, build, and deploy AI-powered systems.[28]

ENLIST THE BOARD, ENGAGE STAKEHOLDERS

Since any AI-related ethical issue may carry broad and long-term risks—reputational, financial, and strategic—it is prudent to engage the board to address AI risks. Ideally, the task should fall to a technology or data committee of the board or, if no such committee exists, the entire board.

Designing ethics into AI starts with determining what matters to stakeholders such as customers, employees, regulators, and the general public. Companies should consider setting up a dedicated AI governance and advisory committee including cross-functional leaders and external advisers that would engage with stakeholders, including multi-stakeholder working groups, and establish and oversee governance of AI-enabled solutions including their design, development, deployment, and use. With regulators specifically, organizations need to stay engaged to not only to track evolving regulations but to shape them.

LEVERAGE TECHNOLOGY AND PROCESS TO AVOID BIAS AND OTHER RISKS

AI developers need to be trained to test for and remediate systems that unintentionally encode bias and treat users or other affected parties unfairly. Researchers and companies are introducing tools and techniques that can help. These include analytics tools that can automatically detect how data variables may be correlated with sensitive variables such as age, sex, or race; tests for algorithmic bias that may generate decisions that are unfair to certain populations; and methods for auditing and explaining how machine learning algorithms generate their outputs. Companies will need to integrate new technologies, control structures, and processes to manage these risks.[29] Organizations

should stay informed of developments in this area and ensure they have processes in place to use them appropriately.

BUILD TRUST THROUGH TRANSPARENCY

In an era of opaque, automated systems and deepfakes (realistic but synthetic images, videos, speech, or text generated by AI systems),[30] companies can help build trust with stakeholders by being transparent about their use of AI. For instance, rather than masquerade as humans, intelligent agents or chatbots should identify themselves as such. Companies should disclose the use of automated decision systems that affect customers. Where possible, companies should clearly explain what data they collect, what they do with it, and how that usage affects customers.

HELP ALLEVIATE EMPLOYEE ANXIETY

Whether AI will eliminate jobs or transform them, it is likely that the technology eventually will affect many, if not most, jobs in some way. An ethical response for companies is to begin advising employees on how AI may affect their jobs in the future. This could include retaining workers whose tasks are expected to get automated or whose work will likely entail using automated systems—or giving them time to seek new employment. Companies in technology, financial services, energy and resources, and telecom have already started preparing their employees to stay relevant in AI-driven future.[31]

Balance the benefits and risks of AI

New technology always brings benefits and risks, and AI is no different. Wise leaders seek to balance risks and benefits to achieve their goals and fulfill their responsibilities to their diverse stakeholders. Even as they seek to take advantage of AI technology to improve business performance, companies should consider the ethical questions raised by this technology and begin to develop their capacity to leverage it effectively and in an ethically responsible way.

Endnotes

1. Deloitte's Quid analysis.

2. Jeff Loucks, Tom Davenport, and David Schatsky, *State of AI in the enterprise, 2nd edition*, Deloitte Insights, October 22, 2018.

3. Tim Dutton, "An overview of national AI strategies," Medium, June 29, 2018.

4. Zoey Chong, "New AI ethics council in Singapore will give smart advice," CNet, June 5, 2018; Sydney J. Freedberg, "Joint Artificial Intelligence Center created under DoD CIO," Breaking Defense, June 29, 2018; Zoë Bernard, "The first bill to examine 'algorithmic bias' in government agencies has just passed in New York City," *Business Insider*, December 19, 2017; Tim Sandle, "France and Canada collaborate on ethical AI," *Digital Journal*, June 10, 2018; Amanda Russo, "United Kingdom partners with World Economic Forum to develop first artificial intelligence procurement policy," World Economic Forum, September 20, 2018; CIFAR, "AI & society," 2017.

5. Alex Hern, "DeepMind announces ethics group to focus on problems of AI," *Guardian*, October 4, 2017; Sundar Pichai, "AI at Google: Our principles," Google, June 7, 2018; Kyle Wiggers, "Google's What-If tool for TensorBoard helps users visualize AI bias," *VentureBeat*, September 11, 2018; Adam Cutler, Milena Pribić, and Lawrence Humphrey, "Everyday ethics for artificial intelligence," IBM, September 2018; Stephan Shankland, "Facebook starts building AI with an ethical compass," CNet, May 2, 2018.

6. Peter High, "Bank of America and Harvard Kennedy School announce the Council on the Responsible Use of AI," *Forbes*, April 23, 2018; Finextra, "Singapore's MAS preps guidelines for ethical use of AI and data analytics," April 4, 2018; Big Data Institute, "Oxford secures £17.5 million to lead national programmes in AI to improve healthcare," November 6, 2018; Partnership on AI, "Zalando," accessed March 28, 2019; Partnership on AI, "BBC," accessed March 28, 2019.

7. Thomas H. Davenport, Jeff Loucks, and David Schatsky, *The 2017 Deloitte state of cognitive survey*, Deloitte, November 2017.

8. Abigail Beall, "It's time to address artificial intelligence's ethical problems," *Wired*, August 24, 2018.

9. arXiv, "Do we need Asimov's laws?" *MIT Technology Review*, May 16, 2014.

10. Christoph Schulze, "Ethics and AI," University of Maryland, 2012.

11. Stefan Helmreich and Heather Paxson, "Computing is deeply human," MIT School of Humanities, Arts, and Social Sciences, February 18, 2019.

12. Jeffrey Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women," Reuters, October 10, 2018; Kaveh Waddell, "How algorithms can bring down minorities' credit scores," *Atlantic*, December 2, 2016; Ed Yong, "A popular algorithm is no better at predicting crimes than random people," *Atlantic*, January 17, 2018.

13. David Schatsky and Rameeta Chauhan, *Machine learning and the five vectors of progress*, Deloitte Insights, November 29, 2017.

14. Salesforce Research, *Trends in customer trust*, accessed March 27, 2019.

15. Loucks, Davenport, and Schatsky, *State of AI in the enterprise, 2nd edition*.

16. Pichai, "AI at Google: Our principles;" Adam Cutler, Milena Pribić, and Lawrence Humphrey, *Everyday ethics for artificial intelligence*, IBM, September 2018.

17. Dave Gershgorn, "Google's new ethics rules forbid using its AI for weapons," Quartz, June 8, 2018.

18. Kush R. Varshney, "Introducing AI Fairness 360," IBM Research blog, September 19, 2018; Natasha Lomas, "IBM launches cloud tool to detect AI bias and explain automated decisions," *TechCrunch*, September 19, 2018; Wiggers, "Google's What-If tool for TensorBoard helps users visualize AI bias."

19. David Meyer, "AI has a big privacy problem and Europe's new data protection law is about to expose it," *Fortune*, May 25, 2018.

20. Dutton, "An overview of national AI strategies."

21. Chong, "New AI ethics council in Singapore will give smart advice;" Freedberg, "Joint Artificial Intelligence Center created under DoD CIO;" Bernard, "The first bill to examine 'algorithmic bias' in government agencies has just passed in New York City;" Sandle, "France and Canada collaborate on ethical AI;" Russo, "United Kingdom partners with World Economic Forum to develop first artificial intelligence procurement policy."

22. Jeremy Hsu, "College AI courses get an ethics makeover," *Discover*, April 26, 2018; Natalie Saltiel, "The ethics and governance of artificial intelligence," MIT Media Lab course, November 16, 2017.

23. MIT, "Moral Machine," accessed March 27, 2019.

24. MIT School of Humanities, Arts, and Social Sciences, "Ethics, computing, and AI," February 18, 2019.

25. Alex Hern, "DeepMind announces ethics group to focus on problems of AI," *Guardian*, October 4, 2017; Aurel Dragan, "SAP launches the first Guide to Artificial Intelligence and an external commission of AI ethics consultants," *Business Review*, September 20, 2018; *PR Newswire*, "Axon launches first artificial intelligence ethics board for public safety; promotes responsible development of AI technologies," April 26, 2018.

26. Partnership on AI, "Frequently asked questions," accessed March 27, 2019.

27. *PR Newswire*, "Axon launches first artificial intelligence ethics board for public safety;" Rachel Louise Ensign, "Bank of America's workers prepare for the bots," *Wall Street Journal*, June 19, 2018; Thomas H. Davenport and Vivek Katyal, "Every leader's guide to the ethics of AI," *MIT Sloan Management Review*, December 6, 2018; Ali Hashmi, *AI ethics: The next big thing in government*, Deloitte, February 2019.

28. For a fuller treatment of some of these topics, see Davenport and Katyal, "Every leader's guide to the ethics of AI."

29. For more on managing algorithmic risk, see Dilip Krishna, Nancy Albinson, and Yang Chu, *Managing algorithmic risks*, Deloitte, 2017.

30. John Villasenor, "Artificial intelligence, deepfakes, and the uncertain future of truth," Brookings, February 14, 2019

31. Mark MacCarthy, "Planning for artificial intelligence's transformation of 21st Century jobs," *CIO*, March 6, 2018; Ensign, "Bank of America's workers prepare for the bots;" Genpact, "New ways of working with artificial intelligence," accessed March 27, 2019.

Acknowledgments

About the authors

DAVID SCHATSKY is a managing director at Deloitte LLP, based in New York. He analyzes emerging technology and business trends for Deloitte's leaders and clients. Before joining Deloitte, Schatsky led two research and advisory firms. He is on Twitter @dschatsky.

VIVEK KATYAL is a principal at Deloitte & Touche LLP, based in Minneapolis. As the Global and US Risk Analytics leader and data risk leader for Cyber Risk Services, he primarily serves his clients' various risk needs and manages professionals who support those areas. He also represents Risk in the Deloitte Analytics Integrated Market Offering. He is on LinkedIn at www.linkedin.com/in/vivek-vic-katyal-482166/ and on Twitter @DeloitteFinSvcs.

SATISH IYENGAR is a senior manager at Deloitte & Touche LLP, based in Minneapolis. He is an advisory professional with more than 16 years of management consulting experience in the areas of data and AI risk management. As part of the cyber data risk team at Deloitte Risk & Financial Advisory, Iyengar works on the assessment, design, development, and implementation of data, analytics, and AI strategies and solutions. He is on LinkedIn at www.linkedin.com/in/satish-iyengar-170210/ and on Twitter @iyengar33.

RAMEETA CHAUHAN is an assistant manager at Deloitte Services India Pvt. Ltd., based in Mumbai. She tracks and analyzes emerging technology and business trends, with a primary focus on cognitive technologies, for Deloitte's leaders and its clients. Prior to Deloitte, Chauhan worked with multiple companies as a part of technology and business research teams. She is on LinkedIn at www.linkedin.com/in/rameeta-chauhan-15b0311b/.

Contacts

David Schatsky
Managing director
US Innovation
Deloitte LLP
+1 646 582 5209
dschatsky@deloitte.com

Vivek Katyal
Global Analytics & Data Risk leader
Principal
Deloitte and Touche LLP
+1 612 397 4772
vkatyal@deloitte.com

Deloitte & Touche LLP's Data Risk Services practice has helped organizations address their data risks for more than 10 years. Our work in helping clients to preserve, maintain, and create value with their data assets includes the design, implementation, and management of data-driven strategies to mitigate risks and drive growth and efficiencies. Contact the authors for more information or read more about value-based data risk management on Deloitte.com.

Deloitte.
Insights

Sign up for Deloitte Insights updates at www.deloitte.com/insights.

Follow @DeloitteInsight

Deloitte Insights contributors
Editorial: Matthew Budman, Blythe Hurley, and Rupesh Bhat
Creative: Molly Woodworth and Anoop K R
Promotion: Nikita Garia
Cover artwork: Molly Woodworth

About Deloitte Insights

Deloitte Insights publishes original articles, reports and periodicals that provide insights for businesses, the public sector and NGOs. Our goal is to draw upon research and experience from throughout our professional services organization, and that of coauthors in academia and business, to advance the conversation on a broad spectrum of topics of interest to executives and government leaders.

Deloitte Insights is an imprint of Deloitte Development LLC.

About this publication

This publication contains general information only, and none of Deloitte Touche Tohmatsu Limited, its member firms, or its and their affiliates are, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your finances or your business. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser.

None of Deloitte Touche Tohmatsu Limited, its member firms, or its and their respective affiliates shall be responsible for any loss whatsoever sustained by any person who relies on this publication.

About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee ("DTTL"), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as "Deloitte Global") does not provide services to clients. In the United States, Deloitte refers to one or more of the US member firms of DTTL, their related entities that operate using the "Deloitte" name in the United States and their respective affiliates. Certain services may not be available to attest clients under the rules and regulations of public accounting. Please see www.deloitte.com/about to learn more about our global network of member firms.