Issue: AI and Ethics

AI and Ethics

By: Hannah H. Kim

Pub. Date: June 11, 2018
Access Date: April 25, 2019
DOI: 10.1177/237455680418.n1
Source URL: http://businessresearcher.sagepub.com/sbr-1946-106816-2890725/20180611/ai-and-ethics

Can machines learn to explain their decisions?

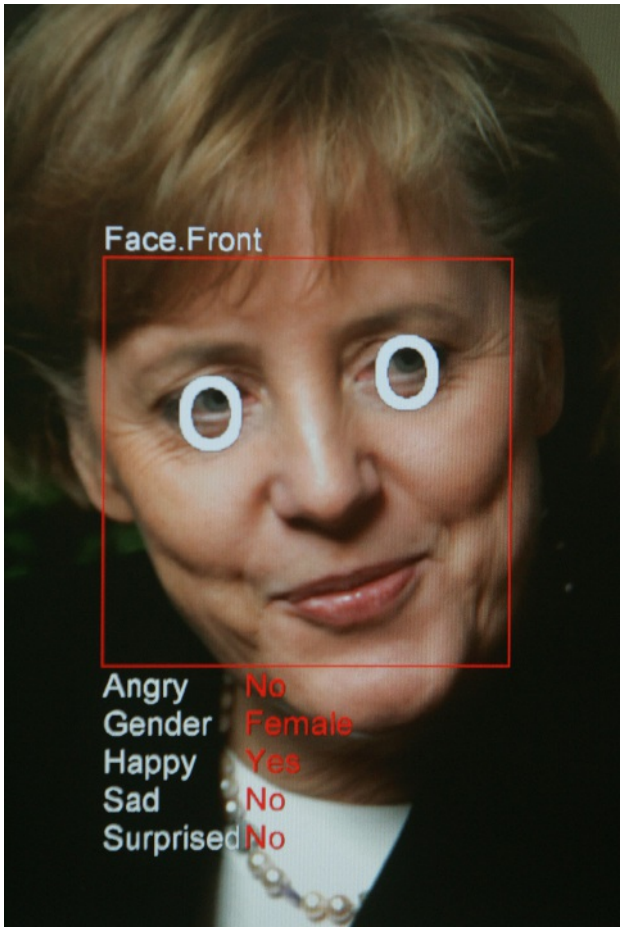
Executive Summary

Artificial intelligence, or AI – machine functions that exhibit abilities of the human mind such as learning and problem solving – is in a phase of rapid development. U.S.-based and international technology companies are in an intense competition to develop AI systems that can automate tasks in a diverse range of industries, from entertainment to marketing to human resources. However, AI systems are so complex that even their developers are sometimes uncertain about how they make decisions. This has raised unique ethical and regulatory challenges, as researchers point to evidence that AI systems can be biased in favor of certain groups of people for jobs, services or loans. Many researchers are calling for greater transparency and for a legal "right to explanation" for those affected by AI decisions.

Here are some key takeaways:

- Investments in AI are being fueled by tech giants, such as Google and Baidu; companies invested more than $20 billion in AI-related mergers and acquisitions globally last year.
- There has also been an upswing in research into AI systems such as deep learning, with most of it published in China and the United States.
- AI's decision-making in high-stakes fields such as hiring and criminal justice has elevated ethical concerns about bias and accountability.
- Click here to listen to an interview with author Hannah H. Kim or click here for the transcript.

Full Report



AI powers facial recognition software, such as this system that analyzed a photo of German Chancellor Angela Merkel. (John MacDougall/AFP/Getty Images)

Henry Gan, a software engineer at the short-video hosting company Gfycat, tested new software on his coworkers last year that used artificial intelligence to identify their faces. While the software correctly matched most of Gan's coworkers with their names, it confused Asian faces. [1]

Gan, who is Asian-American, attempted to remedy the issue by adding more photos of Asians and darker-skinned celebrities into the database that "trained" the software. His efforts produced little improvement. So Gan and his colleagues built a feature that forced the software to apply a more rigorous standard for determining a match when it recognized a face sharing features similar to Asians in its database. Gan acknowledged that allowing the software to look for racial differences to counteract prejudice may seem counterintuitive, but he said it "was the only way to get it to not mark every Asian person as Jackie Chan or something." [2]

Artificial intelligence, or AI, is in a phase of rapid development. Tech giants such as Google and Baidu are pouring money into it; in 2017, companies globally invested around $21.3 billion in AI-related mergers and acquisitions, according to PitchBook, a private market data provider. [3] Yet amid this investment growth, AI systems have presented unique ethical and regulatory challenges as their real-life applications, such as in facial recognition software, have shown evidence of bias.

AI research was born as an academic discipline from a workshop at Dartmouth College in 1956. [4] After several cycles of high expectations followed by disappointment and loss of research funding, the current wave of progress accelerated in 2010 with the availability of big data, increased computer processing power and advancements in machine learning algorithms. [5]

"The most important general-purpose technology of our era is artificial intelligence, particularly machine learning … that is, the machine's ability to keep improving its performance without humans having to explain exactly how to accomplish all the tasks it's given," wrote Erik Brynjolfsson and Andrew McAfee, co-directors of the Massachusetts Institute of Technology's Initiative

on the Digital Economy. [6]

Machine learning differs from the previous approach to AI, in which developers had to explicitly code rules into software. Instead, machine learning allows a system to learn from examples, find patterns, make predictions and improve its own performance. Deep learning, which is a subfield of machine learning, enables a system to process information in layers to accurately recognize extremely complex data patterns. Researchers have been surprised by the recent successes of very large deep-learning networks, which is the main cause of current optimism for AI. [7]

In 2016, machine learning attracted $5 billion to $7 billion in investment, according to a report by the global management company McKinsey. [8] Machine learning can be combined with other technologies to enable a wide range of tasks. The entertainment company Netflix reported how it used machine learning algorithms to personalize video-streaming options for its subscribers and avoid cancelled subscriptions that saved the company $1 billion per year. [9] The marketing company Infinite Analytics used machine learning to improve online advertising placement for its clients, which resulted in a threefold return on investment for a global packaged goods company. [10]

Companies are also integrating machine learning into their business processes, including chatbots that automate customer service or smart robots that are designed to collaborate with human workers in factories and warehouses. [11] In 2012, the e-commerce giant Amazon acquired the robotics company Kiva, and used machine learning to improve the performance of Amazon's 80,000 robots in its fulfillment centers to reduce operating costs by 20 percent. [12]

"The methodologies and hardware to support machine learning have become unbelievably advanced," says Michael Skirpan, co-founder of the self-described ethical engineering consulting and research firm Probable Models, who provides training to companies' engineering teams about fairness in machine learning. "Most adopters are large-enterprise companies that are putting machine learning into their software or online platforms."

Amazon acquired Kiva to upgrade the robots in its fulfillment centers, such as this one in Tracy, California. (David Paul Morris/Bloomberg via Getty Images)

Most investment in AI has consisted of internal spending by large U.S.-based and international technology companies, which are engaged in intense competition to lead AI innovation. One measure of this interest is the sharp increase from 2013 to 2015 in journal articles mentioning "deep learning," with most of the research published in the United States and China. Another measure is that deep-learning patents increased sixfold in that time. [13] The excitement of AI comes from its immense potential to systemically transform industries, according to Brynjolfsson and McAfee. [14]

Currently, however, machine learning is still considered to be early in its development as a research field. Technology companies are

staking out their positions in this phase of AI deployment and securing the limited pool of AI talent by acquiring startups and recruiting top researchers from universities by offering them higher salaries, company stock options and the opportunity to work with large, privately held datasets. [15]

As competition in AI intensifies, academic institutions and researchers as well as various organizations are focusing more attention on the technology's ethical and social effects. While the performance of AI systems has improved rapidly in the closed world of the laboratory, the transition of AI into real-world applications has raised unforeseen challenges, such as how AI systems can make decisions that reflect and reinforce human prejudices. [16]

Fairness in AI and Machine Learning

An AI system can learn bias in a number of ways, according to researchers in a Microsoft team called FATE (for fairness, accountability, transparency and ethics). The system can mimic behaviors it learns from its users, it can be trained with incomplete or historically prejudiced datasets, or its code can reflect the biases of the developers who had built them, the researchers concluded. [17]

Hiring is a prime example of how an AI system can learn bias from the quality of the data on which it is trained. In this area, the system can automate the initial assessment of job applicants and make retention predictions to reduce turnover costs for a company. Solon Barocas, an assistant professor of information science at Cornell University, explains that the AI system would first have to be trained using historical data. By exposing the system to examples of past job applicants who have gone on to be high- or low-performing employees, the system will attempt to learn the distinguishing characteristics of each, in order to make performance predictions about future job applicants. "The hope is that computers can be much more powerful in finding those indicators of potentially high-performing employees," Barocas says.

However, because the data used to train the system come from managers' evaluations, "what the machine learning model is learning to predict is not who is going to perform well at the job, but … managers' evaluations of these people," Barocas says. If the managers' assessments were biased, the AI system will learn discriminatory practices inherent in a historic dataset and apply those to future candidates.

Bias can also occur when an AI system is trained on data that are not representative of the people who will be affected by the algorithms. In hiring, if an employer had categorically discriminated against certain groups of people for positions, the system will continue to reject those applicants, simply because it has no data about them. Will Byrne, director of strategy at Presence Product Group, a product studio specializing in digital technology, wrote, "As a result of societal bias and lack of equal opportunity, predictors of successful women engineers and predictors of successful male engineers are simply not the same. Creating different algorithms with unique training corpuses for different groups … leaves a minority group less disadvantaged." [18]

Applications of AI in Commercial and Public Institutions

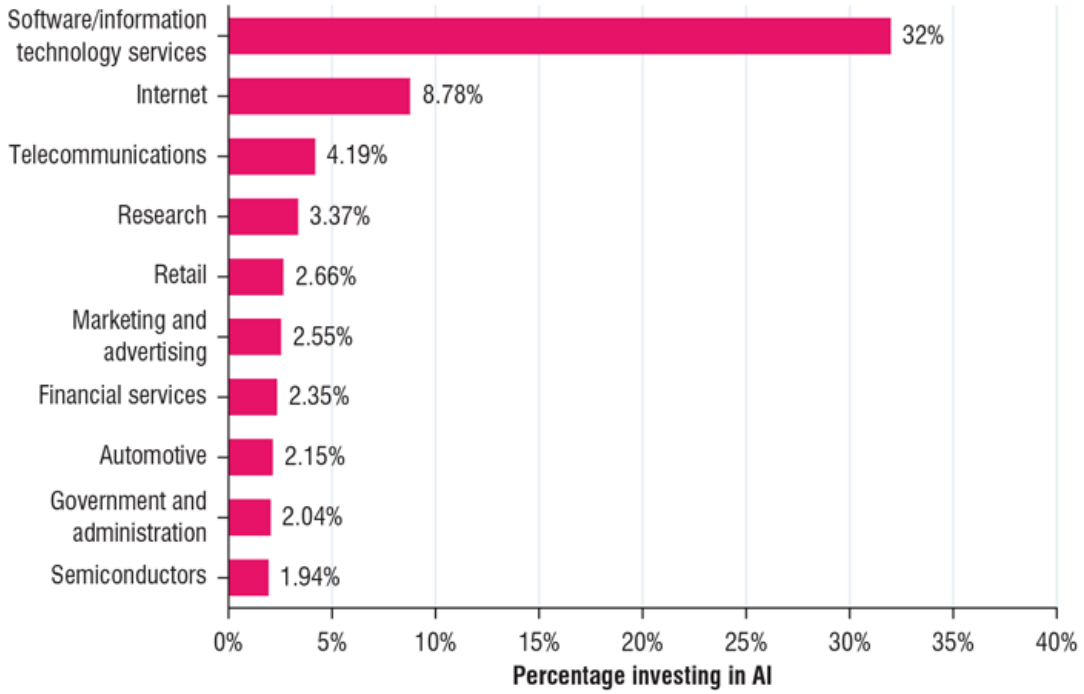
Algorithmic bias draws further concern from researchers as an increasing number of social, corporate and governmental institutions implement AI systems to automate decisions. [19] "Machine learning is increasingly used in high-stakes decision-making," Barocas says. "It's used in medicine, criminal justice, credits, insurance. It's now part of decision-making that implicates people's basic rights and liberties."

For example, hospitals such as the Mayo Clinic, UCLA Medical Center and Johns Hopkins Hospital are using AI systems to develop personalized cancer treatment plans and improve accuracy in medical diagnosis. [20] Insurance companies, including State Farm, Liberty Mutual and Progressive, are using AI systems to develop virtual assistants that use customers' social data to provide faster, personalized service and to customize insurance options. [21]

Similarly, government agencies, such as those in New York City, are using AI systems to make decisions on which schools students will be assigned to and to deploy police based on areas where a crime is most likely to occur. [22]

Software/IT Invests Most in AI

Percentage of companies investing in AI, by industry



Aman Naimat, "The New Artificial Intelligence Market," p. 7, O'Reilly Media, 2016, https://tinyurl.com/yad95bro

The software/information and technology services industry leads all others in investing in artificial intelligence.

However, the application of AI systems to automate such consequential decisions raises ethical concerns, because currently there is not a way to audit the machine's decision-making process. Many AI systems work as "black boxes," producing results without transparency as to how decisions were made. Although developers use accuracy assessments to ensure that an AI system works well, the algorithms that power machine learning are often so complex that the developers themselves cannot exactly understand how they work, according to a 2016 report by a White House interagency working group, the Subcommittee on Machine Learning and Artificial Intelligence of the National Science and Technology Council (NSTC). [23]

Because of this opacity, AI's ethical and social issues have become apparent after the deployment of automated decision systems in real-world situations. [24] In 2016, the nonprofit news organization ProPublica examined AI systems called "risk assessments" that were used in courts across the United States to aid in sentencing decisions. According to ProPublica, the risk assessments exhibited racial bias, labeling white defendants as "low risk" to commit future crimes more often than black defendants. [25]

A software company called Northpointe had created the algorithms that were among the most widely used in these risk assessments. Its central product derived scores from 137 questions about a defendant's criminal history, education and behavior, and did not ask about race. [26] Nevertheless, the AI system had made correlations among a number of factors, such as whether the defendant was "a suspected or admitted gang member," that resulted in consistent racial bias, according to ProPublica. [27]

Northpointe rejected ProPublica's report, saying in a letter to the news organization that its conclusions did not "accurately reflect the outcomes from the application of the model." [28]

Since 2016, researchers have made efforts to study algorithmic bias, according to the AI Now Institute, which is affiliated with New York University. However, the exact reasons for algorithmic bias are not clearly known, the institute said in a 2017 report. Machine learning algorithms consider an enormous quantity of variables, as well as nuanced correlations among these variables, and this complexity can obscure the methodology of its decision-making. [29]

Researchers have expressed significant concern that black-box AI systems can undermine the public's ability to understand or contest the reasons for decisions that can have a significant effect on people's lives. "Imagine that you're rejected from a job, and the answer is, 'Well, we know that the model performs well, but we don't really understand why you were rejected,' " Barocas says. "Or, you apply for credit, and the rejection decision's explanation is, 'the model said so.' "

Cathy O'Neil, a mathematician and data scientist, wrote in her book, "Weapons of Math Destruction," that "many companies go out of their way to hide the results of the models or even their existence. One common justification is that the algorithm constitutes a 'secret sauce' crucial to their business." [30]

The inability to verify an AI system's results played a role in a ruling by a federal judge that the Houston school district's use of a system that assessed teachers' performances based on their students' test scores may violate the teachers' civil rights. The software company that designed the system, the SAS Institute, refused to share the algorithms powering its Educational Value-Added Assessment System (EVAAS), saying they were trade secrets. U.S. Magistrate Judge Stephen Smith wrote in his May 2017 opinion: "The EVAAS score might be erroneously calculated for any number of reasons, ranging from data-entry mistakes to glitches in the computer code itself. Algorithms are human creations, and subject to error like any other human endeavor." [31]

"There's a sense that an algorithm just came into being. Well, humans wrote the algorithm."

In May 2018, the European Union's General Data Protection Regulation (GDPR) took effect. The GDPR contains a set of rules that includes a "right to explanation" for those affected by decisions made solely by AI systems and imposes penalties for noncompliance of up to 4 percent of a company's annual global revenues, or 20 million euros, whichever amount is greater. [32]

Since 2017, researchers at Microsoft and Google have been paying increasing attention to developing methods that show how AI systems work. [33] In a 2017 article in the machine learning academic journal Distill, three Google research scientists wrote about a tool they had developed that allowed people to see how an AI system builds its understanding of images in layers. [34] "There's some sense in which we don't know what it means to see," wrote Chris Olah, the lead researcher. "We don't understand how humans do it. We want to understand something not just about neural nets but something deeper about reality." [35]

Another effort was launched by the Defense Advanced Research Projects Agency (DARPA), an agency of the U.S. Department of Defense. DARPA's Explainable AI (XAI) initiative, which provides $75 million in funding to 12 new research programs, is aimed at producing techniques that will allow humans to understand how AI systems make decisions. [36] "The real secret is finding a way to put labels on the concepts inside a deep neural net," said David Gunning, a program manager at DARPA. "If the concepts inside can be labeled, then they can be used for reasoning." [37]

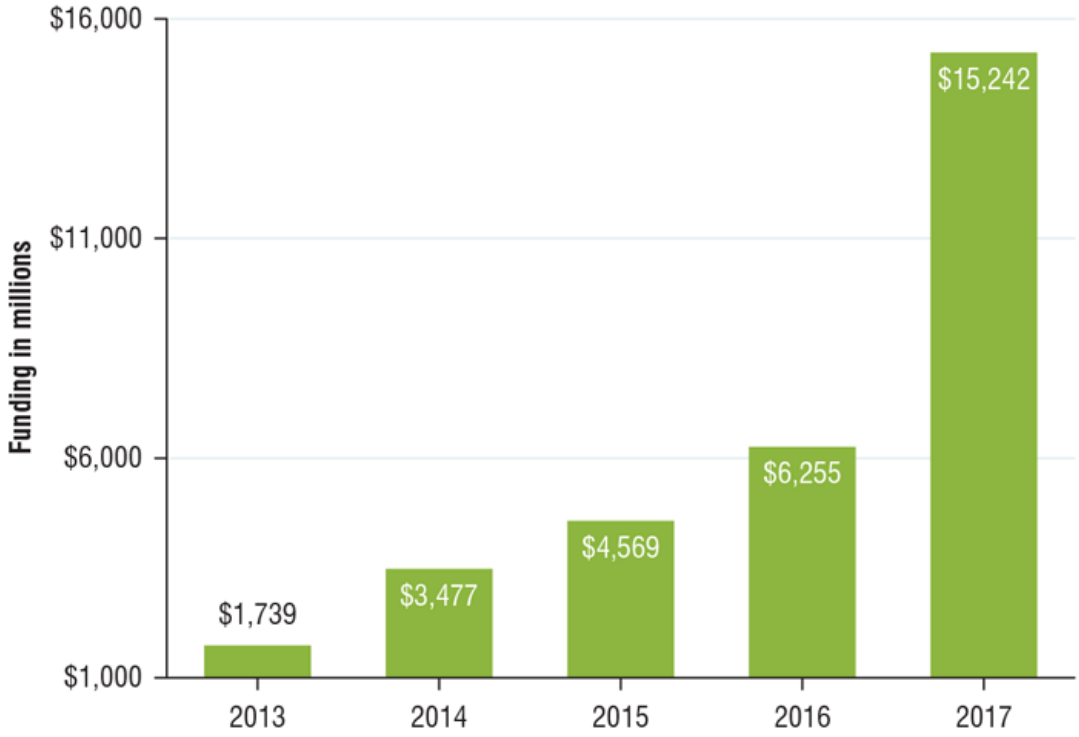
Industry experts say there is also a need to create a system of documentation that developers can use to track how AI systems are tested and made and to enable auditing. Developing industry standards to document the quality of datasets, trial tests and analysis methods can mitigate algorithmic bias by providing a framework of analysis for testing AI systems, or otherwise make a system's biases more apparent, according to a 2016 report published by NSTC's Subcommittee on Networking and Information Technology Research and Development. [38]

Global consulting firms such as Deloitte and Accenture offer companies advice on algorithmic risk management. However, external auditing of companies' AI systems to check for bias issues is at an experimental phase because of the difficulty in verifying black-box AI systems. [39] The apartment rental search engine startup, Rentlogic, is one example of a company that has hired an external mathematician to test the fairness of its algorithms. [40]

In December 2017, New York became the first U.S. city to establish a task force to study potential bias in automated decision systems used in municipal agencies. [41] As a method to provide accountability for black-box AI systems used in public agencies, the AI Now Institute has proposed algorithmic impact assessments, which include agencies' self-auditing of automated decision systems to evaluate potential fairness issues, as well as external review processes for researchers to measure the social impact of AI systems over time. The goal of algorithmic impact assessments is to create a practical framework to track and measure AI systems that will allow the public to question and appeal automated decisions. [42]

AI Funding Spiked in 2017

Funding for AI startups, 2013–17



Note: Funding excludes hardware-focused robotic startups.

"Top AI Trends to Watch in 2018," p. 25, CB Insights, 2018, https://tinyurl.com/y96rh38t

Investor funding for artificial intelligence startups spanning multiple industries jumped more than 140 percent in 2017 from the year before.

While there has been a public backlash against biased machine learning algorithms as they are understood, Skirpan of Probable Models says, "Consumers are a long way away from fully even knowing where machine learning, AI or data-driven technologies are touching their lives."

The increasing use of AI systems in daily life, and the degree of technical complexity by which they operate, put more responsibility on the developers to design AI systems with ethically and socially responsible considerations, according to a 2016 report by NSTC's Subcommittee on Machine Learning and Artificial Intelligence. [43] At an event in San Francisco in February 2018 called Data For Good Exchange, a group of data scientists presented a code of ethics for their field to attendees including employees from Microsoft, Pinterest, AT&T and Google. The goal of the event was to raise awareness of ethical concerns, such as bias that may be embedded in data. [44]

The Institute of Electrical and Electronics Engineers (IEEE), a New York-based professional organization, is among several groups working to create industry standards to help AI engineers build systems while taking into consideration both technical and social effects. John C. Havens, executive director of the IEEE Global Initiative on Ethics of Autonomous and Intelligence Systems (A/IS), says, "With autonomous and intelligent systems, there's a whole new realm of issues: everything from algorithmic bias to human agency to how personal data is used in these systems, that … require new types of methodologies and priorities for engineers."

Barocas, the Cornell professor, says, "The people often best positioned to understand these issues [of bias] are the engineers building those systems." He says progress surrounding ethics and AI systems should prioritize "providing the front-line engineer with the guidance to make sure that when they are building systems, they have [social impact] considerations in mind."

Progress Toward Equality and Diversity in AI

Some experts argue that biased AI systems can reflect the lack of diversity within the industry. A 2017 report by the AI Now Institute concluded that "AI developers are mostly male, generally highly paid, and similarly technically educated. Their interests, needs, and life experiences will necessarily be reflected in the AI they create." [45] Encouraging diversity among the creators of AI technologies, especially more women and people of color, may mitigate bias issues by reflecting a broader variety of perspectives in AI systems themselves, according to several industry diversity organizations, such as Black in AI and Women in AI. [46]

"There's a sense that an algorithm just came into being," Havens says. "Well, humans wrote the algorithm. You have to have technological systems that mirror and reflect the full population."

In a 2018 study, researchers at the Massachusetts Institute of Technology and Microsoft's research arm tested three commercial AI systems that identified the gender of people from their faces, and found error rates of 0.8 percent for lighter-skinned men and 34.7 percent for darker-skinned women. [47] The researchers expressed concern regarding the gender and racial inequity evident in these AI systems, as automated facial recognition software is used for commercial applications, such as in smartphones, and governmental agencies, including to identify criminal suspects in policing.


Rana el Kaliouby

The study examined commonly used image datasets and found that lighter-skinned people constituted the overwhelming majority of examples. The researchers concluded that the imbalance caused the disparity in error rates, and introduced a new face dataset that may mitigate bias issues in facial recognition systems. [48]

The makers of two of the systems in question, Microsoft and IBM, said in response to the study that they were working to improve the accuracy of their products. [49] IBM said its system "now uses different training data and different recognition capabilities than the service evaluated in this study." [50]

In 2016, a technology industry consortium called Partnership on AI to Benefit People and Society was created with founding members Microsoft, Apple, Google, IBM, Facebook, Amazon and the British AI company DeepMind to create industry best practices, including the development of accountability methods in AI systems. [51]

Havens says a common argument he hears regarding the implementation of standards focused on ethical aspects of A/IS is that they will hinder innovation at a time of unprecedented growth. "Often, anyone talking ethics or values-based design gets put into a compartment where they are assumed to be keeping progress from moving forward," Havens says.

However, some leading AI companies say ethics may be the focus of the next wave of industry development. "Today, AI ethics is the new 'Green,' " wrote Rana el Kaliouby, CEO of the emotion measurement technology company, Affectiva. In an article for the business magazine Inc., el Kaliouby drew a parallel to companies such as Walmart that have moved to reduce their overall environmental impact.

"Spurring controversy on whether this was a smart business decision for its shareholders, Walmart committed to minimizing waste and optimizing products and operations while preserving natural resources," el Kaliouby wrote. "On the path to ubiquity of AI, there will be many ethics-related decisions that we, as AI leaders, need to make. We have a responsibility to drive those decisions, not only because it is the right thing to do for society but because it is the smart business decision." [52]

As global competition in AI intensifies, Havens says that initiatives to enable transparency and accountability in A/IS are intended to help the technologies' creators speak to the public and other stakeholders and answer the question: "How can I show you what I've built in ways that you will understand?"

About the Author

Hannah H. Kim is an independent business journalist and ghostwriter. She has written for Vice, Broadly, Korea Daily Newspaper, Isthmus Newspaper and Kenyon Review. She has also ghostwritten several business books. She is from Los Angeles and is an Iowa Writers' Workshop graduate and a member of the American Society of Journalists and Authors. Learn more about her work at www.hannah-h-kim.com. Her previous report for Business Researcher was on the meditation industry.

Chronology

1950–2007	**AI develops in fits and starts.**
1950	Alan Turing, an English computer scientist and mathematician who helped break German codes during World War II, writes his seminal paper, "Computing Machinery and Intelligence," and asks, "Can machines think?"
1956	Computer scientist John McCarthy coins the term "artificial intelligence" for a workshop at Dartmouth College, where AI is founded as an academic discipline.
1963	The Defense Advanced Research Projects Agency (DARPA), a Defense Department agency originally known as ARPA, provides millions of dollars in grants to fund academic AI research.

1974–80	The first "AI winter," a term coined by the Association for the Advancement of Artificial Intelligence, starts as the U.S. and British governments drastically cut research funding. AI's initial advances had raised unrealistic expectations, which led to criticisms in the face of fundamental problems such as limited computing power.
1980–87	AI-powered "expert systems" – computer systems that solve complex problems – are adopted globally by many corporations, including two-thirds of the Fortune 1000 companies.
1987–93	The industry falls into a second AI winter as expert systems fail to compete with desktop computers. Research continues, and AI algorithms begin to appear in real-world systems such as data mining and medical diagnosis.
1997	AI development reaches a milestone when IBM's chess-playing computer, Deep Blue, defeats world champion Garry Kasparov.
2007	DARPA creates the Personalized Assistant that Learns (PAL) to make information systems more effective for users. PAL is commercialized by the startup Siri, Inc., and is acquired by Apple in 2010.
2010–Present	**Investment and technology advances fuel progress.**
2010	The availability of big data, increased computer processing power and advancements in machine learning algorithms launch a new wave of progress for AI.
2014	Venture capital investments in AI startups jump to more than $300 million from about $75 million the previous year, according to the market research firm CBInsights.
2015	Deep-learning research increases as the number of journal articles that are indexed in the Web of Science, a citation indexing service, and mention "deep learning" grows exponentially from 2013 to 2015. Most of the research is published in the United States and China.
2016	The White House Office of Science and Technology Policy leads a series of public workshops on AI.… A technology industry consortium called the Partnership on AI to Benefit People and Society is created with founding members Microsoft, Apple, Google, IBM, Facebook, Amazon and the British AI company DeepMind.
2017	New York becomes the first U.S. city to establish a task force to study potential bias in AI systems used in municipal agencies.… DARPA launches its Explainable AI (XAI) initiative, which provides $75 million in funding to 12 new research programs with the goal of producing techniques that will allow humans to understand how AI systems make decisions.… China unveils a plan to become the world leader in AI by 2030.
2018	The European Union's General Data Protection Regulation, which includes a "right to explanation" for those affected by decisions made solely by AI systems, takes effect (May).

Resources for Further Study

Bibliography

Books

Domingos, Pedro, "The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World," Basic Books, 2015. A computer science professor at the University of Washington provides an introduction to machine learning and its applications in business.

Eubanks, Virginia, "Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor," St. Martin's Press, 2018. An associate professor of political science at the University at Albany, SUNY, investigates the effects of data tracking and automated decision-making on poor and working-class people.

O'Neil, Cathy, "Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy," Crown, 2016. A mathematician and data scientist who earned her doctorate in math from Harvard University examines how unregulated and uncontestable black-box AI systems can reinforce discrimination.

Articles

Angwin, Julia, et al., "Machine Bias," ProPublica, May 23, 2016, http://tinyurl.com/jzdpyz5. Reporters uncover racial bias evident in algorithmic-based scoring systems that are used to inform decisions in courtrooms nationwide.

Bass, Dina, and Ellen Huet, "Researchers Combat Gender and Racial Bias in Artificial Intelligence," Bloomberg, Dec. 4, 2017, http://tinyurl.com/ycbkrore. An article profiles AI researchers who are aiming to combat machine-learning bias affecting women and minorities.

Simonite, Tom, "Artificial Intelligence Seeks an Ethical Conscience," Wired, Dec. 7, 2017, http://tinyurl.com/y7crvtqv. A journalist summarizes discussions about ethics among leading AI researchers at a conference in Long Beach, Calif.

Reports and Studies

"Artificial Intelligence: The Next Digital Frontier?" McKinsey & Company, June 2017, http://tinyurl.com/yapjawqd. A report by a global consulting firm examines AI investments and commercial applications.

"The National Artificial Intelligence Research and Development Strategic Plan," Executive Office of the President, National Science and Technology Council, Committee on Technology, October 2016, http://tinyurl.com/qxrwqtf. A report by a White House interagency working group provides national strategy recommendations for AI research and development.

"Preparing for the Future of Artificial Intelligence," Executive Office of the President, National Science and Technology Council, Committee on Technology, October 2016, http://tinyurl.com/h4ekpt2. Another White House interagency report surveys the current progress of AI and addresses concerns regarding its impact on society, business and public policy.

Campolo, Alex, et al., "AI Now 2017 Report," AI Now Institute, January 2018, http://tinyurl.com/ybqr6z74. A nonprofit research institute at New York University provides an overview of current ethics issues in AI and offers recommendations for future research.

The Next Step

AI Bias

Delaney, John K., "France, China, and the EU All Have an AI Strategy. Shouldn't the US?" Wired, May 20, 2018, https://tinyurl.com/ydcxvfo4. The United States needs to catch up to other countries in AI regulation, including oversight aimed at guaranteeing fair and unbiased implementation, says a Maryland congressman.

Knight, Will, "Microsoft is creating an oracle for catching biased AI algorithms," MIT Technology Review, May 25, 2018, https://tinyurl.com/y75q2ybf. Microsoft and Facebook are working on an AI algorithm designed to detect instances of bias in response to growing concern over the technology's tendency to mirror societal prejudices.

Locascio, Robert, "Thousands of Sexist AI Bots Could Be Coming. Here's How We Can Stop Them," Fortune, May 10, 2018, https://tinyurl.com/y8nhwvmu. AI developers should work harder to diversify their workforces to combat potentially biased, sexist programming seeping into the AI-user relationship, says the CEO of a tech company.

Facial Recognition

Chutel, Lynsey, "China is exporting facial recognition software to Africa, expanding its vast database," Quartz, May 25, 2018, https://tinyurl.com/yabohwqb. As part of its partnership with a China-based facial recognition company, Zimbabwe has agreed to supply the company with personal data and to help improve the technology's ability to recognize differences among people of different ethnicities.

Greig, Jonathan, "Welsh police facial recognition software has 92% fail rate, showing dangers of early AI," TechRepublic, May 8, 2018, https://tinyurl.com/yb7w56v7. The high failure rate of the Welsh police force's facial recognition program highlights growing concerns among watchdog groups that the technology lacks government oversight and regulation.

Wren, Ian, and Scott Simon, "Body Camera Maker Weighs Adding Facial Recognition Technology," NPR, May 12, 2018, https://tinyurl.com/ycf4jyvh. The leading supplier of body cameras for law enforcement is considering adding a facial recognition feature in an effort to stay competitive in the market.

Organizations

AI Now Institute
60 5th Ave., 7th Floor, New York, NY 10011
1-212-998-1212
https://ainowinstitute.org
Research institute at New York University that examines the social implications of artificial intelligence.

Berkman Klein Center for Internet & Society
23 Everett St., #2, Cambridge, MA 02138

1-617-495-7547
https://cyber.harvard.edu/research/ai
Research center at Harvard University that serves, along with the MIT Media Lab, as an anchor institution of the Ethics and Governance of Artificial Intelligence Fund, a $27 million fund created in 2017 to advance AI research for the public interest.

Center for Human-Compatible Artificial Intelligence
University of California, Berkeley, CA 94720-1234
1-510-642-4964
http://humancompatible.ai
Research center whose mission is creating beneficial AI systems by incorporating elements from the social sciences.

Institute of Electrical and Electronics Engineers
3 Park Ave., 17th Floor, New York, NY 10016-5997
1-212-419-7900
http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html
Professional organization that developed an ethics initiative for autonomous and intelligent systems.

MIT Media Lab
77 Massachusetts Ave., E14/E15, Cambridge, MA 02139-4307
1-617-253-5960
https://www.media.mit.edu
Research laboratory at the Massachusetts Institute of Technology.

Partnership on AI to Benefit People and Society
215 2nd St., Suite 200, San Francisco, CA 94105
https://www.partnershiponai.org
Nonprofit technology industry consortium established to formulate best practices on AI technologies.

White House Office of Science and Technology Policy
1650 Pennsylvania Ave., Washington, DC 20504
1-202-456-4444
www.whitehouse.gov/ostp
Federal office established by Congress in 1976 to advise the president and others within the Executive Office of the President on science and technology.

Notes

[1] Tom Simonite, “How Coders Are Fighting Bias in Facial Recognition Software,” Wired, March 29, 2018, http://tinyurl.com/y8lcx2yn/.

[2] Ibid.

[3] “Google leads in the race to dominate artificial intelligence,” The Economist, Dec. 7, 2017, http://tinyurl.com/y9erbqft.

[4] “The National Artificial Intelligence Research and Development Strategic Plan,” National Science and Technology Council, Networking and Information Technology Research and Development Subcommittee, October 2016, p. 5, http://tinyurl.com/qxrwqtf.

[5] “Preparing for the Future of Artificial Intelligence,” Executive Office of the President, National Science and Technology Council, Committee on Technology, October 2016, pp. 5–6, http://tinyurl.com/h4ekpt2.

[6] Erik Brynjolfsson and Andrew McAfee, “The Business of Artificial Intelligence: What it can – and cannot – do for your organization,” Harvard Business Review, July 2017, http://tinyurl.com/y9xmtdk2.

[7] Ibid.

[8] Jacques Bughin, et al., “Artificial Intelligence: The Next Digital Frontier?” McKinsey & Company, June 2017, p. 12, http://tinyurl.com/yapjawqd.

[9] Nathan McAlone, “Why Netflix thinks its personalized recommendation engine is worth $1 billion per year,” Business Insider, June 14, 2016, http://tinyurl.com/y75hmg4s.

[10] Brynjolfsson and McAfee, op. cit.

[11] Bughin, et al., op. cit., p. 11; “Smart robots: Machine learning drives collaboration, continual improvement,” ElectronicDesign, May 24, 2018, http://tinyurl.com/yam7jrnn.

[12] “Machine Learning on AWS,” Amazon.com, undated, accessed April 23, 2018, http://tinyurl.com/ov3ulp8; “Google leads in the race to dominate artificial intelligence,” The Economist, Dec. 7, 2017, http://tinyurl.com/y7lkn6xq; Bughin, et al., op. cit.

[13] “The National Artificial Intelligence Research and Development Strategic Plan,” op. cit., pp. 13–14.

[14] Brynjolfsson and McAfee, op. cit.

[15] Daniela Hernandez and Rachael King, “Universities’ AI Talent Poached by Tech Giants,” The Wall Street Journal, Nov. 24, 2016, http://tinyurl.com/y8852a8q.

[16] “Preparing for the Future of Artificial Intelligence,” op. cit.

[17] Dina Bass and Ellen Huet, “Researchers Combat Gender and Racial Bias in Artificial Intelligence,” Bloomberg, Dec. 4, 2017, http://tinyurl.com/ycbkrore.

[18] Will Byrne, “Now Is the Time to Act to End Bias in AI,” Fast Company, Feb. 28, 2018, http://tinyurl.com/y9b9kc8u.

[19] Alex Campolo, et al., “AI Now 2017 Report,” AI Now Institute at New York University, 2017, http://tinyurl.com/ybqr6z74.

[20] Kumba Sennaar, “How America’s 5 Top Hospitals Are Using Machine Learning Today,” techemergence, April 13, 2018, http://tinyurl.com/ycs5nvww.

[21] Kumba Sennaar, “How America’s Top 4 Insurance Companies Are Using Machine Learning,” techemergence, March 27, 2018, http://tinyurl.com/ybv37fhf.

[22] Dan Rosenblum, “The fight to make New York City’s complex algorithmic math public,” City & State New York, Nov. 21, 2017, http://tinyurl.com/y8hkdgh8.

[23] “Preparing for the Future of Artificial Intelligence,” op. cit., p. 32.

[24] Campolo, et al, op. cit.

[25] Julia Angwin, et al., “Machine Bias,” ProPublica, May 23, 2016, http://tinyurl.com/jzdpyz5.

[26] Ibid.

[27] “Risk Assessment,” Northpointe, Inc., 2011, http://tinyurl.com/y7o99q26.

[28] Angwin, et al., op. cit.

[29] Campolo, et al., op. cit.

[30] Cathy O’Neil, “Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy,” Crown Random House, 2016, p. 29.

[31] Cameron Langford, “Houston Schools Must Face Teacher Evaluation Lawsuit,” Courthouse News Service, May 8, 2017, http://tinyurl.com/y73j3zrm.

[32] “Recital 71 EU GDPR,” EU General Data Protection Regulation (EU-GDPR), April 27, 2016, http://tinyurl.com/y99l48qm; “GDPR Key Changes,” GDPR: Portal, undated, accessed April 30, 2018, http://tinyurl.com/y7vs48bu.

[33] Cliff Kuang, “Can A.I. Be Taught to Explain Itself?” The New York Times, Nov. 21, 2017, http://tinyurl.com/yamspwc6.

[34] Chris Olah, et al., “Feature Visualization,” Distill, Nov. 7, 2017, http://tinyurl.com/y7f7xwdt.

[35] Kuang, op. cit.

[36] David Gunning, “Explainable Artificial Intelligence (XAI),” Defense Advanced Research Projects Agency, undated, accessed April 30, 2018, http://tinyurl.com/yayu9wtx.

[37] Kuang, op. cit.

[38] “The National Artificial Intelligence Research and Development Strategic Plan,” op. cit.

[39] Jessi Hempel, “Want to Prove Your Business Is Fair? Audit Your Algorithm,” Wired, May 9, 2018, http://tinyurl.com/y7z8sgly.

[40] Laura Kusisto, “Is Your Landlord a Building-Code Violator? New Service Aims to Tell You,” The Wall Street Journal, Nov. 14, 2017, http://tinyurl.com/y9g9komv.

[41] Brett Wilkins, “NYC passes bill to study algorithm bias in city agencies,” Digital Journal, Dec. 20, 2017, http://tinyurl.com/y7tuydsn.

[42] Dillon Reisman, et al., “Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability,” AI Now Institute at New York University, April 2018, http://tinyurl.com/y8e3mql5.

[43] “Preparing for the Future of Artificial Intelligence,” op. cit.

[44] Tom Simonite, “Should Data Scientists Adhere to a Hippocratic Oath?” Wired, Feb. 8, 2018, http://tinyurl.com/y8oslvat.

[45] Campolo, et al., op. cit.

[46] Jackie Snow, “ ‘We’re in a diversity crisis’: Cofounder of Black in AI on what’s poisoning algorithms in our lives,” MIT Technology Review, Feb. 14, 2018, http://tinyurl.com/ybygu9q6; Moojan Asghari, “Why we started women in AI,” Medium, Dec. 23, 2017, http://tinyurl.com/ybp8a57y.

[47] Joy Buolamwini and Timnit Gebru, “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification,” MIT Media Lab, Feb. 13, 2018, http://tinyurl.com/yb6unuor.

[48] Ibid.

[49] Josh Horwitz, “If you’re a dark-skinned woman, this is how often facial-recognition software decides you’re a man,” Quartz, Feb. 13, 2018, http://tinyurl.com/y7ov9wj9.

[50] “IBM Response to ‘Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification,’ ” IBM, January 2018, http://tinyurl.com/y89z7yz6.

[51] “Our Work (Thematic Pillars),” Partnership on AI, undated, accessed May 28, 2018, https://tinyurl.com/ybttd8vq.

[52] Rana el Kaliouby, “Ethics in Artificial Intelligence Could Be the Next Big Movement. 5 Ways to Make it Happen,” Inc., Nov. 15, 2017, http://tinyurl.com/ybnvkuck.