CHAPTER 5:
Ethical Challenges of AI Applications

Artificial Intelligence
Index Report 2021

CHAPTER 5:

Chapter Preview

ACCESS THE PUBLIC DATA

Overview

As artificial intelligence–powered innovations become ever more prevalent in our lives, the ethical challenges of AI applications are increasingly evident and subject to scrutiny. As previous chapters have addressed, the use of various AI technologies can lead to unintended but harmful consequences, such as privacy intrusion; discrimination based on gender, race/ethnicity, sexual orientation, or gender identity; and opaque decision-making, among other issues. Addressing existing ethical challenges and building responsible, fair AI innovations before they get deployed has never been more important.

This chapter tackles the efforts to address the ethical issues that have arisen alongside the rise of AI applications. It first looks at the recent proliferation of documents charting AI principles and frameworks, as well as how the media covers AI-related ethical issues. It then follows with a review of ethics-related research presented at AI conferences and what kind of ethics courses are being offered by computer science (CS) departments at universities around the world.

The AI Index team was surprised to discover how little data there is on this topic. Though a number of groups are producing a range of qualitative or normative outputs in the AI ethics domain, the field generally lacks benchmarks that can be used to measure or assess the relationship between broader societal discussions about technology development and the development of the technology itself.
One datapoint, covered in the technical performance chapter, is the study by the National Institute of Standards and Technology on facial recognition performance with a focus on bias. Figuring out how to create more quantitative data presents a challenge for the research community, but it is a useful one to focus on.
Policymakers are keenly aware of ethical concerns pertaining to AI, but it is easier for them to manage what they can measure, so finding ways to translate qualitative arguments into quantitative data is an essential step in the process.

CHAPTER HIGHLIGHTS

- The number of papers with ethics-related keywords in titles submitted to AI conferences has grown since 2015, though the average number of paper titles matching ethics-related keywords at major AI conferences remains low over the years.

- The five news topics that got the most attention in 2020 related to the ethical use of AI were the release of the European Commission's white paper on AI, Google's dismissal of ethics researcher Timnit Gebru, the AI ethics committee formed by the United Nations, the Vatican's AI ethics plan, and IBM's exiting the facial-recognition businesses.

5.1 AI PRINCIPLES AND FRAMEWORKS

Since 2015, governments, private companies, intergovernmental organizations, and research/professional organizations have been producing normative documents that chart the approaches to manage the ethical challenges of AI applications. Those documents, which include principles, guidelines, and more, provide frameworks for addressing the concerns and assessing the strategies attached to developing, deploying, and governing AI within various organizations. Some common themes that emerge from these AI principles and frameworks include privacy, accountability, transparency, and explainability.

The publication of AI principles signals that organizations are paying heed to and establishing a vision for AI governance. Even so, the proliferation of so-called ethical principles has met with criticism from ethics researchers and human rights practitioners who oppose the imprecise usage of ethics-related terms. The critics also point out that they lack institutional frameworks and are non-binding in most cases. The vague and abstract nature of those principles fails to offer direction on how to implement AI-related ethics guidelines.

Researchers from the AI Ethics Lab in Boston created a ToolBox that tracks the growing body of AI principles. A total of 117 documents relating to AI principles were published between 2015 and 2020. Data shows that research and professional organizations were among the earliest to roll out AI principle documents, and private companies have to date issued the largest number of publications on AI principles among all organization types (Figure 5.1.1). Europe and Central Asia have the highest number of publications as of 2020 (52), followed by North America (41), and East Asia and Pacific (14), according to Figure 5.1.2. In terms of rolling out ethics principles, 2018 was the clear high-water mark for tech companies—including IBM, Google, and Facebook—as well as various U.K., EU, and Australian government agencies.

Europe and Central Asia have the highest number of publications as of 2020 (44), followed by North America (30), and East Asia and Pacific (14). In terms of rolling out ethics principles, 2018 was the clear high-water mark for tech companies—including IBM, Google, and Facebook—as well as various U.K., EU, and Australian government agencies.

NUMBER of NEW AI ETHICS PRINCIPLES by ORGANIZATION TYPE, 2015-20

Source: AI Ethics Lab, 2020 | Chart: 2021 AI Index Report

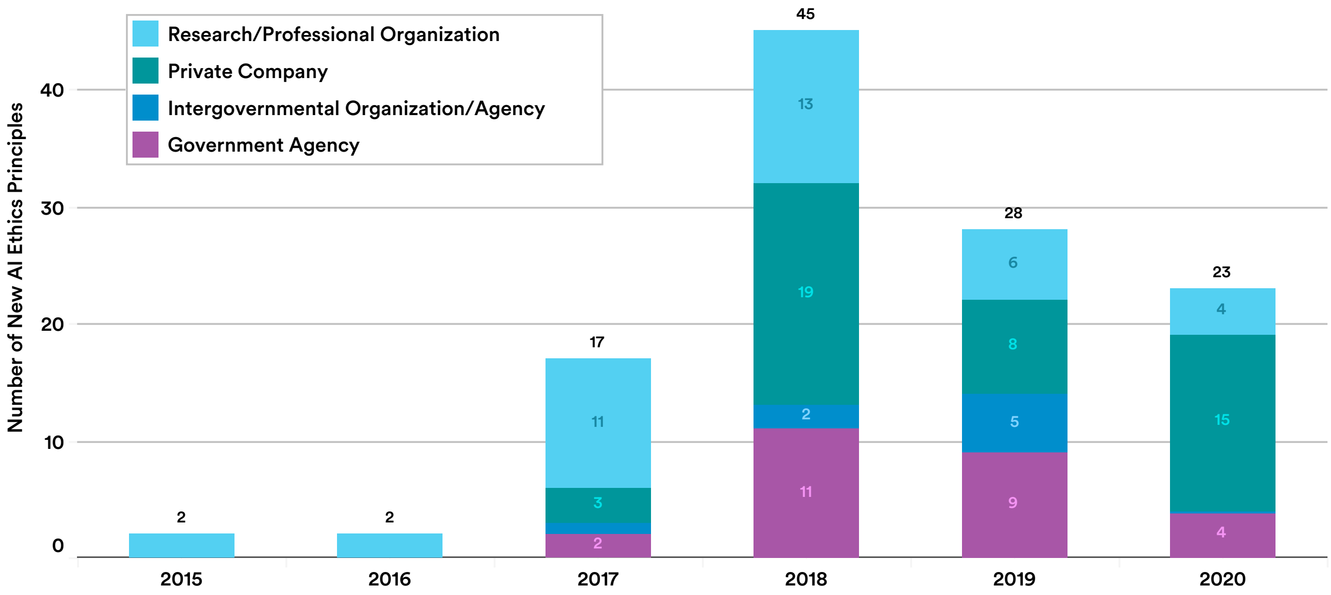

Figure 5.1.1

NUMBER of NEW AI ETHICS PRINCIPLES by REGION, 2015-20

Source: AI Ethics Lab, 2020 | Chart: 2021 AI Index Report

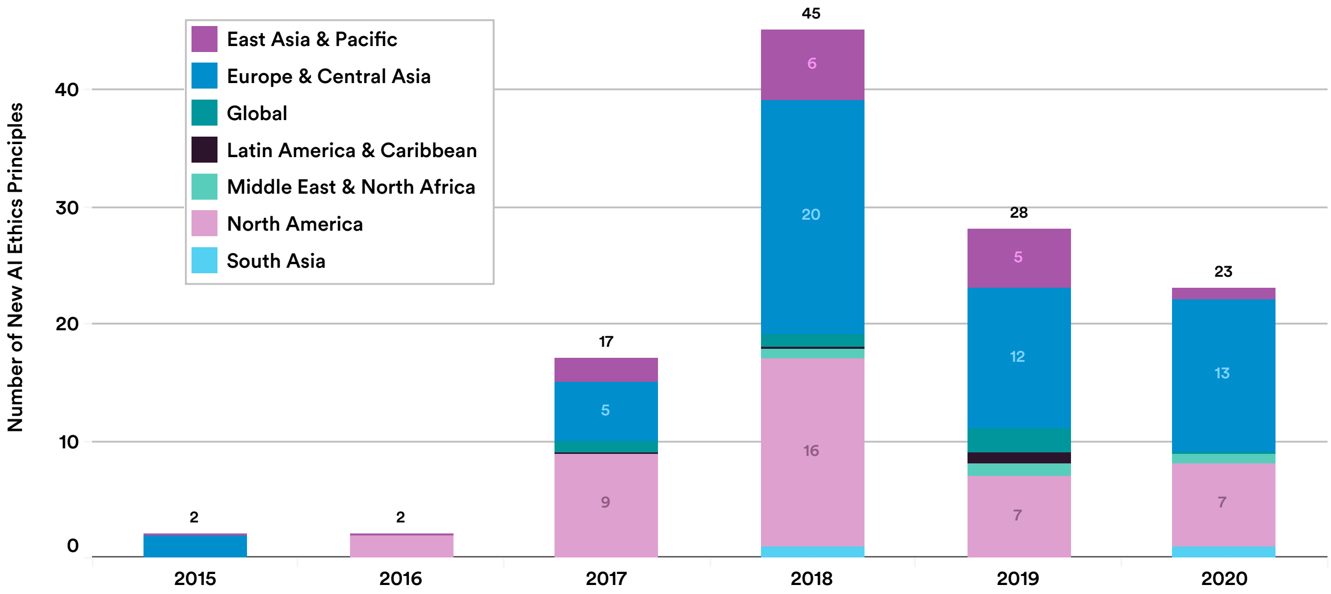

Figure 5.1.2

5.2 GLOBAL NEWS MEDIA

How has the news media covered the topic of the ethical use of AI technologies? This section analyzed data from NetBase Quid, which searches the archived news database of LexisNexis for articles that discuss AI ethics[1], analyzing 60,000 English-language news sources and over 500,000 blogs in 2020.

The search found 3,047 articles related to AI technologies that include terms such as "human rights," "human values," "responsibility," "human control," "fairness," "discrimination" or "nondiscrimination," "transparency," "explainability," "safety and security," "accountability," and "privacy." (See the Appendix for more details on search terms.) NetBase Quid clustered the resulting media narratives into seven large themes based on language similarity.

Figure 5.2.1 shows that articles relating to AI ethics guidance and frameworks topped the list of the most covered news topics (21%) in 2020, followed by research and education (20%), and facial recognition (20%).

The five news topics that received the most attention in 2020 related to the ethical use of AI were:

1. The release of the European Commission's white paper on AI (5.9%)
2. Google's dismissal of ethics researcher Timnit Gebru (3.5%)
3. The AI ethics committee formed by the United Nations (2.7%)
4. The Vatican's AI ethics plan (2.6%)
5. IBM exiting the facial-recognition businesses (2.5%).

NEWS COVERAGE on AI ETHICS (% of TOTAL) by THEME, 2020
Source: CAPIQ, Crunchbase, and NetBase Quid, 2020 | Chart: 2021 AI Index Report

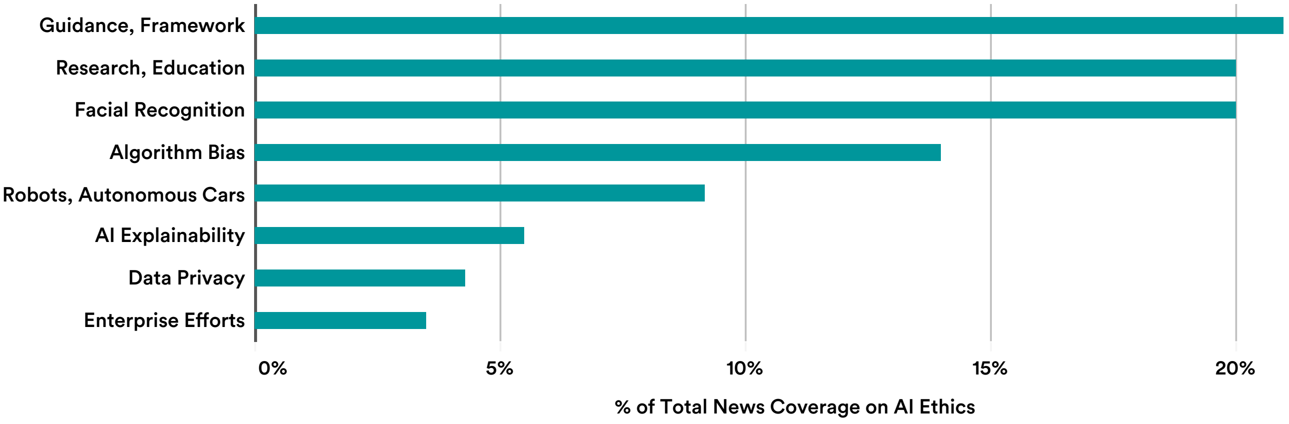

Figure 5.2.1

[1] The methodology for this is looking for articles that contain keywords related to AI ethics as determined by a Harvard research study.

5.3 ETHICS AT AI CONFERENCES

Researchers are writing more papers that focus directly on the ethics of AI, with submissions in this area more than doubling from 2015 to 2020. To measure the role of ethics in AI research, researchers from the Federal University of Rio Grande do Sul in Porto Alegre, Brazil, searched ethics-related terms in the titles of papers in leading AI, machine learning, and robotics conferences. As Figure 5.3.1 shows, there has been a significant increase in the number of papers with ethics-related keywords in titles submitted to AI conferences since 2015.

Further analysis in Figure 5.3.2 shows the average number of keyword matches throughout all publications among the six major AI conferences. Despite the growing mentions in the previous chart, the average number of paper titles matching ethics-related keywords at major AI conferences remains low over the years.

Changes are coming to AI conferences, though. Starting in 2020, the topic of ethics was more tightly integrated into conference proceedings. For instance, the Neural Information Processing Systems (NeurIPS) conference, one of the biggest AI research conferences in the world, asked researchers to submit “Broader Impacts” statements alongside their work for the first time in 2020, which led to a deeper integration of ethical concerns into technical work. Additionally, there has been a recent proliferation of conferences and workshops that specifically focus on responsible AI, including the new Artificial Intelligence, Ethics, and Society Conference by the Association for the Advancement of Artificial Intelligence and the Conference on Fairness, Accountability, and Transparency by the Association for Computing Machinery.

There has been a significant increase in the number of papers with ethics-related keywords in titles submitted to AI conferences since 2015. Further analysis shows the average number of keyword matches throughout all publications among the six major AI conferences.

NUMBER of PAPER TITLES MENTIONING ETHICS KEYWORDS at AI CONFERENCES, 2000-19

Source: Prates et al., 2018 | Chart: 2021 AI Index Report

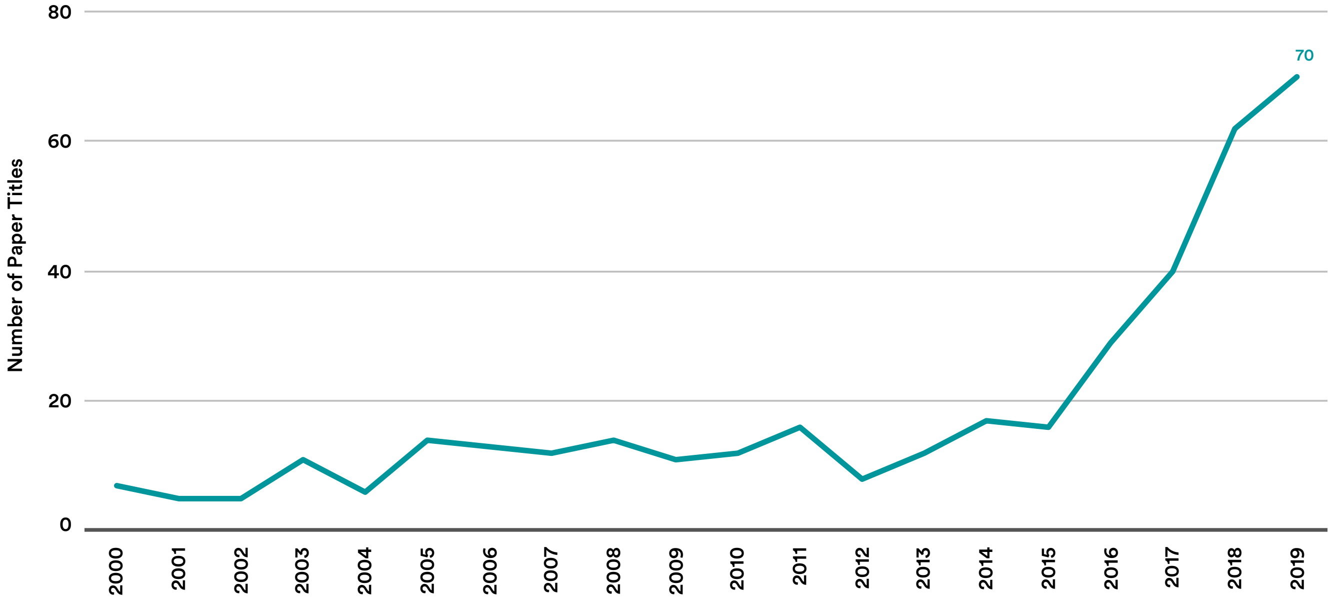

Figure 5.3.1

AVERAGE NUMBER of PAPER TITLES MENTIONING ETHICS KEYWORDS at SELECT LARGE AI CONFERENCES, 2000-19

Source: Prates et al., 2018 | Chart: 2021 AI Index Report

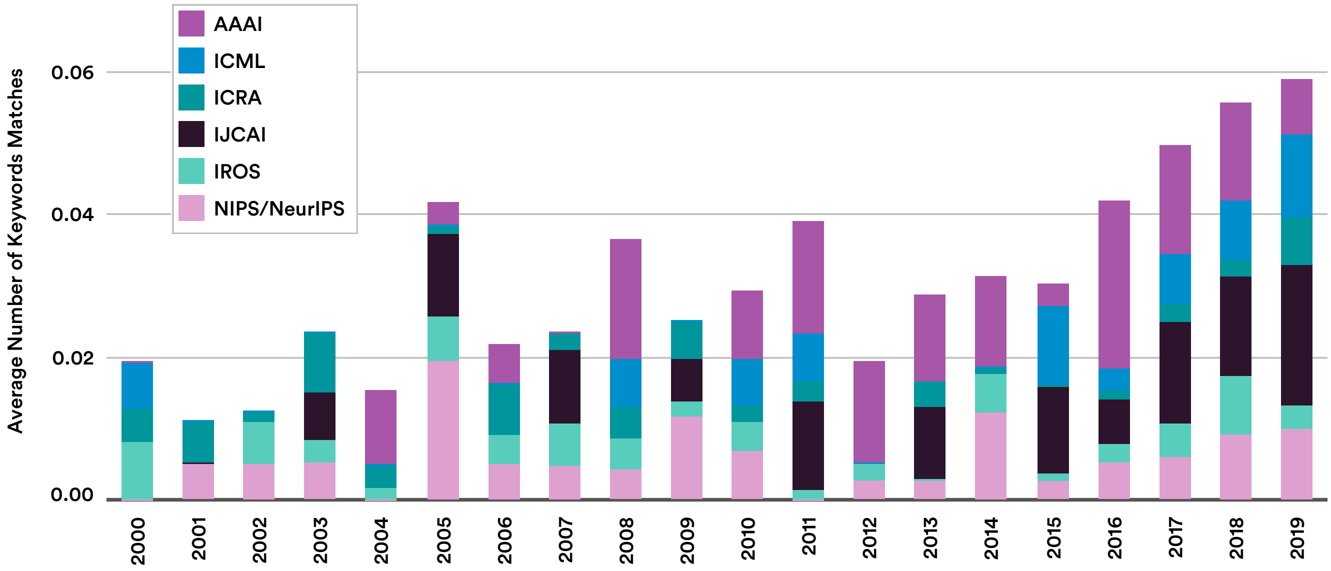

Figure 5.3.2

5.4 ETHICS OFFERINGS AT HIGHER EDUCATION INSTITUTIONS

Chapter 4 introduced a survey of computer science departments or schools at top universities around the world in order to assess the state of AI education in higher education institutions.[2] In part, the survey asked whether the CS department or university offers the opportunity to learn about the ethical side of AI and CS. Among the 16 universities that completed the survey, 13 reported some type of relevant offering.

Figure 5.4.1 shows that 11 of the 18 departments report hosting keynote events or panel discussions on AI ethics, while 7 of them offer stand-alone courses on AI ethics in CS or other departments at their university. Some universities also offer classes on ethics in the computer science field in general, including stand-alone CS ethics courses or ethics modules embedded in the CS curriculum offering.[3]

11 of the 18 departments report hosting keynote events or panel discussions on AI ethics, while 7 of them offer stand-alone courses on AI ethics in CS or other departments at their university.

AI ETHICS OFFERING at CS DEPARTMENTS of TOP UNIVERSITIES around the WORLD, AY 2019-20
Source: AI Index, 2020 | Chart: 2021 AI Index Report

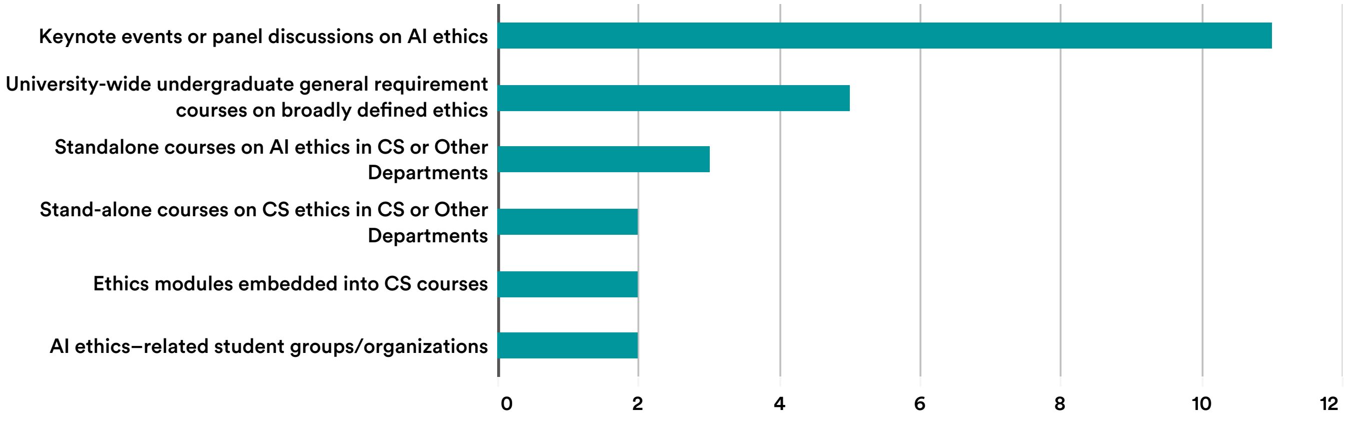

Figure 5.4.1

[2] The survey was distributed to 73 universities online over three waves from November 2020 to January 2021 and completed by 18 universities, a 24.7% response rate. The 18 universities are—Belgium: Katholieke Universiteit Leuven; Canada: McGill University; China: Shanghai Jiao Tong University, Tsinghua University; Germany: Ludwig Maximilian University of Munich, Technical University of Munich; Russia: Higher School of Economics, Moscow Institute of Physics and Technology; Switzerland: École Polytechnique Fédérale de Lausanne; United Kingdom: University of Cambridge; United States: California Institute of Technology, Carnegie Mellon University (Department of Machine Learning), Columbia University, Harvard University, Stanford University, University of Wisconsin–Madison, University of Texas at Austin, Yale University.

[3] The survey did not explicitly present “Ethics modules embedded into CS courses” as an option. Selections were filled in the “Others” option. This will be included in next year’s survey.

APPENDIX

NETBASE QUID

Prepared by Julie Kim

Quid is a data analytics platform within the NetBase Quid portfolio that applies advanced natural language processing technology, semantic analysis, and artificial intelligence algorithms to reveal patterns in large, unstructured datasets and generate visualizations to allow users to gain actionable insights. Quid uses Boolean query to search for focus areas, topics, and keywords within the archived news and blogs, companies, and patents database, as well as any custom uploaded datasets. Users can then filter their search by published date time frame, source regions, source categories, or industry categories on the news; and by regions, investment amount, operating status, organization type (private/public), and founding year within the companies' database. Quid then visualizes these data points based on the semantic similarity.

Network

Searched for [AI technology keywords + Harvard ethics principles keywords] global news from January 1, 2020, to December 31, 2020.

Search Query: (AI OR ["artificial intelligence"]("artificial intelligence" OR "pattern recognition" OR algorithms) OR ["machine learning"]("machine learning" OR "predictive analytics" OR "big data" OR "pattern recognition" OR "deep learning") OR ["natural language"]("natural language" OR "speech recognition") OR NLP OR "computer vision" OR ["robotics"]("robotics" OR "factory automation") OR "intelligent systems" OR ["facial recognition"]("facial recognition" OR "face recognition" OR "voice recognition" OR "iris recognition") OR ["image recognition"]("image recognition" OR "pattern recognition" OR "gesture recognition" OR "augmented reality") OR ["semantic search"]("semantic search" OR "data-mining" OR "full-text search" OR "predictive coding") OR "semantic web" OR "text analytics" OR "virtual assistant" OR "visual search") AND (ethics OR "human rights" OR "human values" OR "responsibility" OR "human control" OR "fairness" OR discrimination OR non-discrimination OR "transparency" OR "explainability" OR "safety and security" OR "accountability" OR "privacy")

News Dataset Data Source

Quid indexes millions of global-source, English-language news articles and blog posts from LexisNexis. The platform has archived news and blogs from August 2013 to the present, updating every 15 minutes. Sources include over 60,000 news sources and over 500,000 blogs.

Visualization in Quid Software

Quid uses Boolean query to search for topics, trends, and keywords within the archived news database, with the ability to filter results by the published date time frame, source regions, source categories, or industry categories. (In this case, we only looked at global news published from January 1, 2020, to December 31, 2020.) Quid then selects the 10,000 most relevant stories using its NLP algorithm and visualizes de-duplicated unique articles.

ETHICS IN AI CONFERENCES

Prepared by Marcelo Prates, Pedro Avelar, and Luis C. Lamb

Source

Prates, Marcelo, Pedro Avelar, Luis C. Lamb. 2018. On Quantifying and Understanding the Role of Ethics in AI Research: A Historical Account of Flagship Conferences and Journals. September 21, 2018.

Methodology

The percent of keywords has a straightforward interpretation: For each category (classical/trending/ethics), the number of papers for which the title (or abstract, in the case of the AAAI and NeurIPS figures) contains at least one keyword match. The percentages do not necessarily add up to 100% (e.g, classical/trending/ethics are not mutually exclusive). One can have a paper with matches on all three categories.

To achieve a measure of how much Ethics in AI is discussed, ethics-related terms are searched for in the titles of papers in flagship AI, machine learning, and robotics conferences and journals.

The ethics keywords used were the following:
Accountability, Accountable, Employment, Ethic, Ethical, Ethics, Fool, Fooled, Fooling, Humane, Humanity, Law, Machine Bias, Moral, Morality, Privacy, Racism, Racist, Responsibility, Rights, Secure, Security, Sentience, Sentient, Society, Sustainability, Unemployment, and Workforce.

The classical and trending keyword sets were compiled from the areas in the most cited book on AI by Russell and Norvig [2012] and from curating terms from the keywords that appeared most frequently in paper titles over time in the venues.

The keywords chosen for the classical keywords category were:
Cognition, Cognitive, Constraint Satisfaction, Game Theoretic, Game Theory, Heuristic Search, Knowledge Representation, Learning, Logic, Logical, Multiagent, Natural Language, Optimization, Perception, Planning, Problem Solving, Reasoning, Robot, Robotics, Robots, Scheduling, Uncertainty, and Vision.

The curated trending keywords were:
Autonomous, Boltzmann Machine, Convolutional Networks, Deep Learning, Deep Networks, Long Short Term Memory, Machine Learning, Mapping, Navigation, Neural, Neural Network, Reinforcement Learning, Representation Learning, Robotics, Self Driving, Self-Driving, Sensing, Slam, Supervised/Unsupervised Learning, and Unmanned.

The terms searched for were based on the issues exposed and identified in papers below, and also on the topics called for discussion in the First AAAI/ACM Conference on AI, Ethics, and Society.

J. Bossmann. Top 9 Ethical Issues in Artificial Intelligence. 2016. World Economic Forum.

Emanuelle Burton, Judy Goldsmith, Sven Koenig, Benjamin Kuipers, Nicholas Mattei, and Toby Walsh. Ethical Considerations in Artificial Intelligence Courses. AI Magazine, 38(2):22–34, 2017.

The Royal Society Working Group, P. Donnelly, R. Browsword, Z. Gharamani, N. Griffiths, D. Hassabis, S. Hauert, H. Hauser, N. Jennings, N. Lawrence, S. Olhede, M. du Sautoy, Y.W. Teh, J. Thornton, C. Craig, N. McCarthy, J. Montgomery, T. Hughes, F. Fourniol, S. Odell, W. Kay, T. McBride, N. Green, B. Gordon, A. Berditchevskaia, A. Dearman, C. Dyer, F. McLaughlin, M. Lynch, G. Richardson, C. Williams, and T. Simpson. Machine Learning: The Power and Promise of Computers That Learn by Example. The Royal Society, 2017.

Conference and Public Venue - Sample

The AI group contains papers from the main artificial intelligence and machine learning conferences such as AAAI, IJCAI, ICML, and NIPS and also from both the *Artificial Intelligence Journal* and the *Journal of Artificial Intelligence Research* (JAIR).

The robotics group contains papers published in the IEEE Transactions on Robotics and Automation (now known as IEEE Transactions on Robotics), ICRA, and IROS.

The CS group contains papers published in the mainstream computer science venues such as the Communications of the ACM, IEEE Computer, ACM Computing Surveys, and the ACM and IEEE Transactions.

Codebase

The code and data are hosted in this GitHub repository.