



1st AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society: A Retrospective

Benjamin Kuipers (University of Michigan; kuipers@umich.edu)

Nicholas Mattei (IBM Research; n.mattei@ibm.com)

DOI: [10.1145/3203247.3203250](https://doi.org/10.1145/3203247.3203250)

Introduction

The 1st AAAI/ACM Conference on AI, Ethics, and Society (AIES-18) was held February 1–3, 2018 at the Hilton New Orleans Riverside, in New Orleans, Louisiana. The event was held just before AAAI in order to highlight the overlap of the two conference and logistical support for the conference was provided by AAAI. By attendance measures the conference was a resounding success with a sold out registration of over 300 people. The conference brought together program chairs from the four major focal areas: AI and jobs: Jason Furman (Harvard University); AI and law: Gary Marchant (Arizona State University); AI and philosophy: Huw Price (Cambridge University); and AI: Francesca Rossi (IBM and University of Padova). All the paper from the conference are available for download at the conference website: <http://www.aies-conference.com/>

Student Program

ACM:SIGAI had a large part in funding and organizing the student program for the conference. With funding from all the conference sponsors each student received a \$1,000 travel grant and complimentary registration. The student program was highly competitive with over 70 applicants competing for just 20 spots. The accepted students ran the gamut of conference areas with students from computer science, law, and philosophy represented. Each of the students participated in a special student lunch with all the invited speakers, had a poster during the student poster session, and had the opportunity to publish an abstract of their thesis work in the conference program.

Thursday February 1

AIES began with an evening reception and panel at Tulane University. The panel title was What will Artificial Intelligence bring? Discussing the advent and consequences of superhuman intelligence, and the panelists were Paula Boddington (Oxford), Wendell Wallach (Yale), Jason Furman (Harvard), and Peter Stone (UT Austin). The panel debated in a filled to capacity auditorium at Tulane touching on all the major topics of the conference including AI and Society and the future of AI.

Friday, February 2

The conference opened at the Hilton with a welcome from Francesca Rossi and the rest of the Program Chairs, and the announcement of the two Best Paper Awards.

The first invited speakers were Iyad Rahwan and Edmond Awad from MIT, describing The Moral Machine Experiment: 40 Million Decisions and the Path to Universal Machine Ethics. This very well-known crowd-sourced experiment asks volunteers on the Web to answer questions about a series of scenarios where a speeding self-driving car with failed brakes must choose which of two sets of people will be killed. The sets of people vary over many dimensions (number, gender, age, innocence, passenger vs pedestrian, etc.), and the respondents characteristics (age, gender, nationality, etc.) are also recorded. Many conclusions can be drawn from the collected data. The experimenters emphasize that the purpose of the experiment is descriptive, rather than prescriptive, providing insights into peoples attitudes, rather than determining the right answers to moral questions. Nonetheless, readers persist in interpreting the results prescriptively, and critics raise concerns about the extreme and unrealistic abstraction of the scenarios, and whether participants responses about hypothetical scenarios have

meaningful interpretations. The presentation was very stimulating and led to vigorous discussion, a theme that persisted throughout the conference.

Each hour-long session for oral paper presentations gave each of four presenters 10 minutes both to present the paper and respond to a few direct questions, followed by 20 minutes for general questions or comments from the audience directed at any or all of the papers in the session. The first paper session focused on social norm learning and value alignment (including papers by the two authors of this report). The second morning session focused on bias and fairness, especially in machine learning/big data applications. The two afternoon sessions were on the topic of AI and Law, the first focused on Responsibility, and the second focused on Governance.

Then Carol Rose, from the ACLU of Massachusetts, gave a very compelling talk on the current impacts of AI/ML technology on the criminal justice system, and the need for technologists with interest and awareness of ethics, not to mention traditional American values of Liberty and Justice for All, to get directly involved in campaigning, consulting, and advising legislators on how to use these technologies wisely and appropriately.

The final invited panel of the first day was on the important role played by standards, and standard-making bodies, in shepherding social decisions about technology policy. The panel was chaired by Simson Garfinkel (USACM) for organizer John Havens (IEEE) who was unable to attend, and also included Takashi Egawa (NEC), Dan Palmer (British Standard Institute), and Annette Reilly (IEEE).

Each of these events prompted vigorous discussion, but there was plenty of discussion energy left over for the Conference Reception, sponsored by Deep Mind Ethics & Society.

Saturday, February 3

The second day began with an invited talk by Richard Freeman, an economist at Harvard, who reminded us that concerns about AI, robots, and the elimination of jobs are not new, illustrated with a quote by Herbert Simon from 1966. Previous scary predictions have not come true, but of course things are dif-

ferent now, in terms of the comparative advantage of automation over human workers across a wider range of tasks. Furthermore, it is worth observing that economic inequality has been increasing before AI and Robots have had significant economic impacts. At least, whatever happens, its not entirely our fault!

The first morning paper session included a paper from Georgia Tech describing their experience with a “Virtual TA” answering student questions for a Knowledge-Based AI class. One question is whether it is ethically required to tell students which TA is the virtual one, since most interaction is via web pages. Another paper, from McGill, described the potential for bias in data-driven dialog systems, and ways that bias and hate speech can be detected and avoided. In the second morning session, one paper described methods for generating explanations from deep neural networks, and another described “Purple Feed”, an approach to select high-consensus items for a news feed that cuts across traditional political “silos”. The two afternoon paper sessions included papers on regulating autonomous vehicles, the rights of service robots, trust in healthcare AI, non-intuition-based machine ethics, and a survey of when people want AI systems to make decisions (primarily, when the person asked has previous exposure to machines making decisions).

There were many more interesting papers than can be mentioned here. The papers are available on the conference website <http://www.aies-conference.com/>, so see for yourself.

The day, and the conference, concluded with two keynote presentations. The first was by Patrick Lin, philosopher from Cal Poly State University, who gave an overview of robot ethics. The second was by Tenzin Priyadarshi, the Buddhist chaplain at MIT. He described the “Because we can!” attitude that is often seen in developers of new technologies. He encouraged us to take moral responsibility for the design of intelligent systems, taking the well-being of humans very much into account.

Final Reflection

This conference reminded Ben in some ways of his first IJCAI in 1973. Both were very inspiring about the importance, the breadth, and the promise of the problems and methods being presented. In both cases, the papers were all over the place, presenting interesting results from many perspectives on many different problems, often with little connection to each other. Since then, AI has grown into a much larger intellectual and industrial enterprise, with great impact on society. That very impact suggests that the focus of this conference on AI, ethics, and society will also become increasingly important.



Benjamin Kuipers is a Professor of Computer Science and Engineering at the University of Michigan. His research focuses the representation, learning, and use of foundational domains of knowledge, including knowledge of space, dynamical change in physical systems, objects, ac-

tions, and now, ethics.



Nicholas Mattei is a Research Staff Member in the IBM Research AI group at the IBM TJ Watson Research Laboratory. His research focuses on the theory and practice of AI, developing systems and algorithms to support decision making.
