Association of Internet Researchers (AoIR) Roundtable Summary: Artificial Intelligence and the Good Society Workshop Proceedings

Cath, Corinne; Zimmer, Michael; Lomborg, Stine; Zevenbergen, Ben

Published in:
Philosophy & Technology

DOI:
10.1007/s13347-018-0304-8

Publication date:
2018

Document version
Early version, also known as pre-print

preprint version. Commentary published in Philosophy & Technology vol. 31(1): 155-162.

Association of Internet Researchers (AoIR) Roundtable Summary: Artificial Intelligence and the Good Society

Corinne Cath, Michael Zimmer, Stine Lomborg and Ben Zevenbergen

Abstract

This conference proceeding comes out of a roundtable held at the Association of Internet Researchers (AoIR) annual conference in 2017, in Tartu Estonia. The roundtable was organized by the Oxford Internet Institute's (OII) Digital Ethics Lab (DELab). It was entitled "Artificial Intelligence and the Good Society". It brought together four expert researchers to discuss the promises and perils of artificial intelligence (AI), in particular what ethical frameworks are needed to guide its rapid development and increased deployment in societies. The paper is based on the talks given by the three speakers - Michael Zimmer, Stine Lomborg and Ben Zevenbergen, and covers three case studies presented by the speakers. These case studies present a distinct overview the ethical issues raised by the use of AI at different levels of analysis: top-down application of AI, bottom-up use of AI, and how academics and governments have reacted to these new challenges. From the case-studies, four areas emerged representing some of the most topical ethical questions related to AI: (1) its uses, (2) its users, (3) its designers, and (4) the data that fuels it. Each of these provided a specific subset of ethical concerns that need further investigation. Concluding, we formulate three maxims for researchers and regulators to ensure the AI has a positive impact on society.

Introduction

Artificial Intelligence (AI) is everywhere. From driverless trucks to the use of machine learning algorithms to improve national security, financial trading markets, and urban planning. AI and the algorithms that underpin it are fast becoming a staple technological system driving some of the most critical areas of societies – including healthcare, provision of government services, sentencing in the justice system, political participation, and interaction with law enforcement. Yet, there is limited understanding of the substantial effects of using AI in these areas. Often, these technologies unfairly disadvantage minorities, women, and people of color.

The development of these new smart technology often outpaces academic research. There is an excellent, and fast-growing, body of literature on improving the auditibility and transparency of AI systems. But, there remains a lack of shared understanding about the fundamental issues at the heart of the debate on AI, algorithms, and ethics. This tension is leading to a renewed academic focus on questions of AI and the "Good Society", specifically on the ethical impact of AI and algorithms on society.

This summary covers the outcome of a roundtable organized by the Oxford Internet Institute's (OII) Digital Ethics Lab (DELab) at the Association for Internet Researchers (AoIR) conference in October 2017. It was entitled "Artificial Intelligence and the Good Society" and brought together four scholars from various fields to share their perspectives on the ethical implications of AI and algorithms in fostering the "Good Society". We facilitated a cross-disciplinary conversation about the challenges and opportunities of AI and looked at what role academic research should play in shaping the debate on the ethical impact of AI on society.

Our speakers were: Privacy and Internet expert Professor Michael Zimmer, technology law and ethics scholar Ben Zevenbergen, and data ethics expert Professor Stine Lomborg. They each presented a case study outlining some of the ethical challenges and opportunities of AI. The workshop was organized and chaired by Corinne Cath, a doctoral student at the University of

Oxford. In this workshop report, an abbreviated version of the talks will be presented, as well as some general considerations to further AI and ethics research and regulation.

Top-Down: The Case of the "Best" Baltic Liberal Arts College

As education becomes more competitive, the ranking of colleges is increasingly important to attract students. Ben Zevenbergen presented a fictional case, abstracted from a real scenario, about a liberal arts college somewhere in the Baltics and what their use of artificial intelligence for improving its ranking can teach us about the ethics of applying AI to the education sector in a top-down fashion.

Before going into the case details, Ben outlined the general approach of the Center for Information Technology Policy (CITP) at Princeton. Their method is to take a case study, understand the engineering ethics issues that arise, apply political theory to the case, and use that to suggest different sets of policy and regulation measures. This overall process ensures that the right incentives are in place, for designers and educators interested in using AI and algorithms, for improving higher education.

The college in the case-study faced a high drop-out rate. What caused it was unclear. The local council suggested that the school should use the different data sets they collected about their students and to consider applying machine learning algorithms to these data sets to better understand whether correlations about the drop-out rates could be found. The college administration did not see any ethical issues with that suggestion, and the IT department proceeded to collect and combine a wide number of data points about the students. This included: information from their ID cards, grades, classes attended, food purchased, family situation, criminal records, and personal device usage (based on the WiFi connection). This data was used to train a machine learning algorithm and find the various variables that correlated with dropping out.

In the first analysis phase, a handful of key indicators was found. It was at this point in the process that the school decided to opt for data minimization, to protect the privacy of the students. The first ethical considerations of this entire process were thus made long after the project was initiated. The prediction rate of the algorithm was remarkably high at 93% accuracy. It was also found that if students received individual support, their chances of dropping out of school were significantly reduced. Through the use of this algorithm, the university rose several places in the overall education ranking.

While successful from the perspective of the school, this case study presents three specific ethical concerns. First, the scope of the data collection. It is very broad. Second, some of the key predictors of success (like the private family situation) were clearly outside of the realm of the university system (i.e. a university cannot change a family situation). Third, the algorithm – as a by-product – was able to predict the likelihood that a student would get a criminal record while in the school. This indicator was clearly outside of the scope of the initial purpose of the top-down data-driven approach and raises a whole separate set of questions surrounding the responsibility of the school vis-à-vis students at risk of getting involved in criminal activities.

Two related issues arose. When the students and staff found out that the school was using their data in this particular way, without their explicit consent, it negatively impacted their trust in the management of the school. Furthermore, teachers were not incentivized to help C students become B students. The system was aimed at ensuring that B students became A students. As such, time spent coaching failing students was not rewarded by this particular system.

These crucial issues notwithstanding, the university took up a second offer from the IT Sales department to use the AI tool to further increase the school's rating. The administration did not ask any critical questions regarding the ethics of the current program and encouraged the IT department to further develop it. This had some detrimental results for the students.

In the second stage of the program, the algorithm was used to determine additional proxies that represent A-grade behavior. This led to a series of administrative decisions including physically removing facilities that do not correlate with A-grade behavior (including eating at the snack bar or attending the campus bar). It also further incentivized supervisors to nudge students towards proxies for A-grade behavior. There were also rewards built into the system for teachers that managed students to get A-grades (while not providing similar rewards for improving D grade to C+ outcomes).

Overall, the ends of the university became to rise up in the university ranking. It did so by creating a controlled environment that treated students and teachers as agents to be optimized. It did not allow for the integration of additional ends often associated with attending higher education, like personal development. The means for rising in the ranking turned the university into a data harvester, which led to distrust amongst students and faculty.

This case study raises several important considerations regarding the ethical AI questions the scholarly community should address:

1. Who designs what for whom?
2. Whose assumptions about the ends are achieved?
3. Whose knowledge of the means drives by which these ends are achieved?
4. Based on which legitimacy?
5. Who has been consulted to understand the impact of these technical decisions?
6. Who oversees changes to the system?

After the students became aware of the data collection and use by the school, the policy changed. The school started a stakeholder process involving staff and students about the program. They also increased their transparency practices, by publishing an open letter that included details about the data collection and processing. This case study is interesting for the specifics of its

progress, from limited data collection and AI-based observations to unilaterally scaling the process with the school's interests in mind, to becoming more sensitive to the specific ethical and privacy considerations of using student data. However, it also resonates with an overarching ethical issue related to AI: it encourages individuals to put a lot more faith and trust in technology than can be justified considering that the issues of uncertainty and bias in statistics apply equally in the case of algorithms and AI-based decision making.

Bottom-up: The Case of Danish AI Understanding

Stine Lomborg presented her ongoing work on how ordinary people make sense of and problematize algorithms and perceive their own agency with a specific focus on everyday uses of digital media, e.g. social media, news websites, search and wayfinding applications such as Google Maps. Research has repeatedly suggested how citizens from all parts of the world are deeply concerned about the possible infringements to their privacy and the risks of social sorting that follow from the use of digital media. At the same time, media usage practices amply demonstrate that people are reluctant to take the frequently recommended steps to reduce the risk of infringement: using safe browsers (Tor), using ad-blockers, deleting cookies and social media profiles etc. She thus highlighted the importance of empirically studying people's awareness and feelings about digital tracking for explaining the gap between what people say and how they address their concerns through their online behavior.

Her talk contributed by presenting preliminary findings from an ongoing empirical study on how ordinary people make sense of and problematize algorithms and perceive their own agency. These insights can be used to reflect on research ethics and ethical obligations of research towards civil society.

The empirical study, based on in-depth qualitative interviews, is an explorative study of what people think algorithms are, what they perceive them to do, and what tactics they use to act upon algorithms in everyday life. It is informed by work done in critical data studies and builds on a

development of Stuart Hall's concept of decoding to sensitize us to the interpretive efforts used when making sense of technology. For the study, 20 Danes from all walks of life were interviewed, with specific sampling emphasis on variations by gender (male/female/non-binary), age, level of education and occupation. The study has a comparative element as there are mirror studies being done in Finland. These comparisons can help clarify some of the contextual and cultural differences in the data. The aim of the study is to probe people's critical awareness of the ways in which algorithms serve as taken-for-granted infrastructures in everyday life. Such critical awareness is crucial for taking part in democratic debates about and shaping of the role of Artificial Intelligence in fostering a Good Society.

Stine presented the preliminary analysis of the interview data. At the baseline, the respondents seemed to have some understanding of what algorithms are, and how they impact their lives. Most individuals brought up the following insights about algorithms:

1. Algorithms are the interfaces between 'me' and the choices I have, they are mathematical formulas for classification, pattern detecting software, and segmentation systems.
2. Algorithms are ubiquitous; they underpin everything from banking, shopping, pathfinding, social media feeds, transport, and personalized healthcare.
3. Algorithms were often problematized through their proxies: people did not express concern about the algorithms as such, but rather about fake news, the attention economy, filter bubbles, and echo-chambers.

The respondents could entertain a conversation about algorithms, but struggled to recognize and talk about their technical functioning, how they 'meet' them when using digital media and with what implications for their personal lives. They drew heavily upon the prevalent public conversation as it is being had in the mainstream media, which to a large extent reflects worries of how algorithmic filtering can negatively impact society. For instance, respondents would bring up the role of social media platforms in spreading disinformation and thereby breaking the traditional social contract between media and users to serve fact-based information and news or

curating information too uniformly, thereby hampering individuals to make informed decisions about what type of content they would like to consume, and take political stance on the role of digital media in the development of society.

Most respondents, however, did not recognize the critical influence of algorithms as the infrastructure underpinning their individual lives. While most people understood how algorithms were problematic to the functioning of certain democratic institutes in society they did not draw that understanding out to include their own lived realities, beyond expressing discontent of being segmented based on - for example - age and therefore receiving ads for anti-age cream. One female respondent noted about personally targeted ads on Facebook: *"they are simply an unpleasant mirror"*.

When discussing some of their day-to-day digital practices, the interviewees focused on the dangers surrounding communication activities. They indicated being cautious and restrictive in what they posted, liked, favored, and commented on, especially in spaces they understood to be public or semi-public, thus reflecting an understanding that algorithmic logics work more strongly in such settings. However, issues surrounding meta-data, sensor data, and cross-references between data-sets were mostly absent from the conversation. This indicates that most people interviewed for this study had a narrow understanding of what kind of data is tracked, and what kinds of personal information can be gleaned from it. Particularly surprising was that the initial rounds of analysis indicated that the level of unawareness was fairly constant across demographics. Overall, most interviewees agreed that the dangers posed by algorithms were relatively benign.

Three general understandings kept resurfacing:

1. Algorithms are not bringing about existential threats.
2. We shouldn't 'blame the algorithm', as it also facilitates many aspects of life and makes them easier.

3. Many respondents understood and accept the proposition that if the service is free, they are the product and were comfortable with that.

There are several preliminary explanations for these statements. For one, the respondents lack the sort of discriminatory experience at the hands of automated systems that have raised awareness in other contexts (for instance surrounding predictive policing in the US). Related to this, respondents also suggested that it was not their individual responsibility to change the current political-economy of algorithms and data collection: as long as they do not experience the kinds of discrimination or social sorting described elsewhere, they were not convinced they should actually bother. .This lack of concern might also be the result of the Danish political context of the welfare state and the high level of trust in societal institutions.

The fact that few interviewees problematized the social impact of algorithms, or even argued that normative criticisms of algorithmic impact are unbalanced, indicates the prevalence in the Danish context of the algorithm as a service to smoothen the surface of our digitally wired societies.

The ease with which some of the respondents accepted their role as a product did have some parameters. Most of them reported being comfortable with algorithmic categorization, as long as it fit the context. When financial data is used to offer services to improve banking, that was considered acceptable. But if an insurance company used Internet activity to tailor their offers, that was not considered acceptable. Contextual integrity is key.

This research also raises the fundamental question what the ethical responsibilities of the researcher are vis-à-vis the respondents. If a respondent fundamentally misunderstands or underestimates what an algorithm does, it is the duty of the researcher to inform them? And what responsibility do researchers have to educate the public-at-large about algorithms? Or even push for regulation and stronger policy engagement in the area? While no single answer was formulated for these questions, it was suggested that researchers could do more to draw from

insights of the ethics of care to operationalize their research results towards achieving the algorithmic Good society. Instead of 'doing no harm', we might begin pursuing impact through the ethical stance of 'doing good'.

The Academic and Government Response: The Case of the "AI Now Initiative"

Michael Zimmer talked about the AI Now Initiative, which started as series of workshops organized by New York University and Obama administration White House. The purpose of the workshops was to focus on the opportunities and challenges presented by AI, and what solutions exist from a technical and legal perspective. The 2016 symposia focused on four themes: Social inequality, labor, healthcare, and ethics. The 2017 symposium expanded that focus to include: labor and automation, bias and inclusion, rights and liberties, and ethics and governance. The growth in the number of topics considered to be critical represents the increasing application of AI to a wide array of societal spheres, as well as the diversification of the researchers working on these issues.

The AI Now Initiative has recently evolved into a full-blown interdisciplinary research institute based at NYU, which is focused on examining the social implications of artificial intelligence. The institute's research is dedicated to four specific domains: rights and liberties, labor and automation, bias and inclusion, and safety and critical infrastructure. In October 2017, the AI Now Institute released its latest report. The report puts forth ten key recommendations, of which the first is that public agencies should not use black box systems. It is also crucial for AI systems to be subject to rigorous pre-release trials. Even everyday items like gum are subject such trials. Considering the critical areas in which AI is employed, ensuring that trials are run to ensure AI systems will not amplify biases or mistakes stemming from the training data, algorithms, and general system design elements is crucial.

The AI Now approach to the question of the social impact of algorithms is holistic. It is based on a value-by-designs process that ensures that AI systems are continuously monitored across different contexts. There is a clear need for more research in some of the specific focus areas

identified by this initiative, most specifically in the use of AI systems in workplace management and monitoring. The problem with using these types of systems is that they run the danger of replicating existing human biases, like for instance being favoring men over women in hiring processes. Or excluding people of color or other minorities.

Related to this, is the need for more inclusive communities of practice. The homogeneity of much of the current AI community is leading to serious ethical issues, whether we are talking about the inability of a soap dispenser to recognize black and brown hands, to more serious issues like biases in AI systems used for policing, credit scoring, and judicial sentencing. Ensuring that a diversity of perspectives and epistemic communities are part of developing AI systems will address some of the blind spots in the current community.

For any debate on ethical codes of conduct regarding AI to be effective, it is important to ensure that they have teeth. Much of the current work being done in this area, while commendable, lacks strong enforcement mechanisms. Making it hard for these codes of conduct to have a discernible impact on practitioners.

There is, however, an overarching theme beyond the scope of the AI Now report. Even if all the recommendations are implemented, there is still the question of pervasive data collection needed to drive the AI fueled society. To maximize the affordances of AI, it is necessary to have data. To gain that data, a certain level of surveillance – whether self-imposed through use of social media and fit-bits or through corporate and state-sponsored means – is necessary. In order to fully address the ethical questions raised by AI, it is imperative to devote considerable efforts to understand the ethical issues inherent in pervasive data collection.

Considerations

Overall, there were four areas emerged representing some of the most topical ethical questions related to AI: (1) its uses, (2) its users, (3) its designers, and (4) the data that fuels it. Each of

these provided a specific subset of ethical concerns that need further investigation. When discussing what an ideal future research agenda for AI and ethics looks like, the following three research questions where defined:

1. Who designs what for whom, and why?
2. How do we empower the users of AI systems?
3. How do we go beyond focusing on technological issues for societal problems?

When clarifying their specific approaches to AI and ethics research, the experts emphasized the following three maxims, for both researchers and regulators:

1. Look at the broader impact of AI, open up your academic worries to other epistemic communities and make your concerns accessible for a non-academic audience, for instance through increasing data literacy of the public.
2. Be critical and always ask "what is the problem this technology is trying to solve?" And as a corollary, be prepared to ask if we should be building a certain technology at all.
3. Work with engineers, be critical of AI systems that are aimed at shareholder maximizing models only, work on sector-specific guidelines, and talk to policy-makers to ensure regulation and policy is evidence-based.

The speakers agreed that the positive impact of AI can be manifold. Not only does it have the potential to facilitate decision making, improve the delivery of government services, as well as raising some important debates about discrimination and bias in society it also encourages further critical and much-needed discussion about the interaction between AI technology and society.