A CULTURE OF ETHICAL AI: REPORT

What steps can organizers of AI conferences take to encourage reflection on the societal impacts of AI research?

JULY 2022

Ada Lovelace Institute

CIFAR

PARTNERSHIP ON AI

ABOUT THE PARTNERS

The Ada Lovelace Institute is an independent research institute with a mission to ensure data and AI work for people and society. Through research, policy and practice, it aims to ensure that the transformative power of data and AI is used and harnessed in ways that maximize social wellbeing and put technology at the service of humanity.

CIFAR

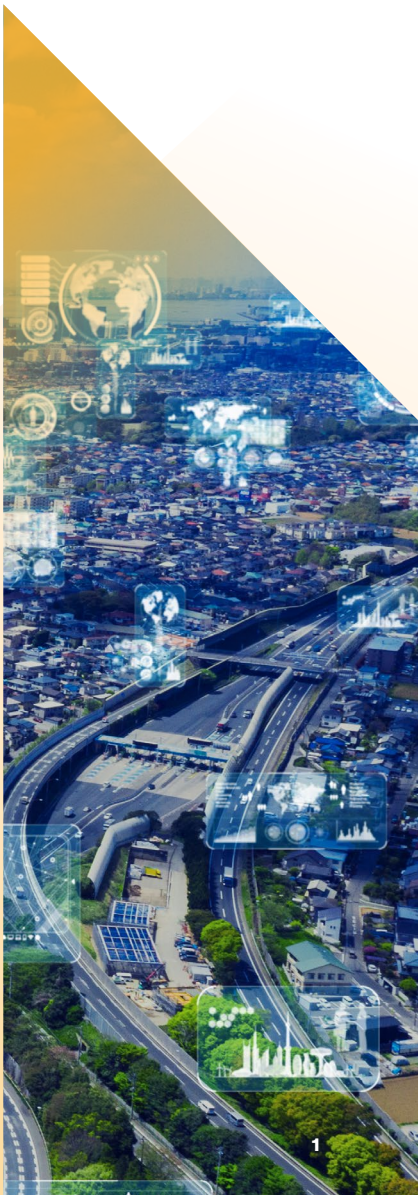
CIFAR is a Canadian-based global research organization which convenes extraordinary minds to address the most important questions facing science and humanity. In 2017, the Government of Canada appointed CIFAR to develop and lead the Pan-Canadian Artificial Intelligence Strategy, the world's first national AI strategy.

The Partnership on AI is a non-profit community of academic, civil society, industry, and media organizations addressing the most important and difficult questions concerning the future of AI.

LAND ACKNOWLEDGMENT

We wish to acknowledge this land on which CIFAR operates. For thousands of years it has been the traditional territory of many nations including the Mississaugas of the Credit, the Anishnabeg, the Chippewa, the Haudenosaunee and the Wendat peoples and is now home to many diverse First Nations, Inuit and Métis peoples. We are grateful to have the opportunity to work on this land. We also acknowledge we are all responsible for reconciliation. CIFAR's AI & Society program seeks to advance our understanding of the societal implications of AI to design a future of responsible AI. A future of responsible AI includes one that centres the concerns of Indigenous communities. CIFAR is committed to prioritizing Indigenous perspectives in the development and design of responsible AI.

TABLE OF CONTENTS



We view this report as a menu of options for future AI and ML conference organizers to choose from, pilot and iterate on at their AI or computing conferences. If you are an organizer who introduces one of the interventions mentioned here, we would love to hear from you at info@cifar.ca.

EXECUTIVE SUMMARY

Against the backdrop of increasing use of artificial intelligence (AI) technologies in everyday life and growing private investment in the area, more researchers are entering the field of AI than ever before. The increasing relevance of AI has come with a wider awareness of its potential harmful real-world impacts, including on the environment, marginalized communities, and society at large.

How can the AI research community better anticipate the downstream consequences of AI research? And how can AI researchers mitigate potential negative impacts of their work such as inappropriate applications, unintended and malicious use, accidents, and societal harms?

In the last few years, some leading AI and machine learning (ML) conferences have begun to take on this challenge.

The NeurIPS 2020 conference introduced a 'broader impact statement' requirement for all submissions to reflect on the potential environmental and societal implications of the research. This was followed by the rollout of the ethics review process at ACL 2021 and the use of ethics checklists at NeurIPS 2021. Organizers of all these conferences recommended that authors discuss the ethical considerations of their research choices and the potential side effects of their work in their submissions.

Many conferences, however, are yet to initiate similar ethics review practices and the successes, challenges and potential downsides of these recent pilots need to be better understood.

Earlier this year, CIFAR, Partnership on AI, and the Ada Lovelace Institute brought together recent ML conference organizers and AI ethics experts, to consider what conference organizers can do to encourage the habit of reflecting on potential downstream impacts of AI research among submitting authors.

KEY TAKEAWAY

Organizers across different AI conferences should continue to collaborate more closely in forums like our workshop and others, to share lessons learnt and discuss community-wide approaches for encouraging more ethical reflection.

The following report synthesizes the insights we gathered from the convening. The report includes five big ideas for how AI and ML conference organizers can address these challenges, along with a wider list of interventions proposed by participants to foster a more responsible research culture in AI.

BIG IDEAS

1 AI CONFERENCE ORGANIZERS CAN CONSIDER A MIX OF PRESCRIPTIVE AND REFLEXIVE INTERVENTIONS TO IMPROVE RESEARCHERS' ABILITY TO ASSESS THE ETHICAL IMPACTS OF THEIR WORK

- Participants discussed the advantages and disadvantages of using prescriptive tools like checklists for ethical examination, which can prompt researchers to consider the ethical and societal impacts of their research including the carbon footprint of their work or auditing for bias in datasets.
- One disadvantage of ethics checklists is that they can limit the scope of enquiry — missing out on issues that are harder for researchers to identify. At the same time participants also acknowledged that prompts in the checklist play an important role in kick-starting the initial ethical reflection process for researchers.
- Participants recommended that checklists be used in combination with more open-ended exercises such as impact statements to encourage reflexivity among researchers. Practicing reflexivity can include an examination of a researcher's assumptions, practices, and commitments, including what broader societal impacts they hope their research will produce.

2 CONFERENCE ORGANIZERS SHOULD PRIORITIZE TRAINING MORE RESEARCHERS AND CONFERENCE REVIEWERS ON HOW TO EXAMINE THE POTENTIAL NEGATIVE DOWNSTREAM CONSEQUENCES OF THEIR WORK

- Conference organizers should set up training for (a) those who will review and flag ethical issues in papers submitted to conferences; and (b) members of the overall research community, around common ethical issues and mitigations. Training can only be so effective at upskilling AI researchers on perspectives from the social sciences but it can offer small improvements to the quality and scope of ethics reviews.
- Workshop participants recommended that for training to be more effective and useful to conference participants, they should be run as exercises, not lectures, inviting direct participation from researchers.
- Conference organizers should additionally consider the benefits and downsides of making training either optional or mandatory.
- This training could be supported by shared resources between different conferences, including case studies, a list of subject-matter experts that organizers can call on for specific ethical guidance, and a list of datasets that have been identified as being problematic or requiring additional ethical review.
- All of these interventions will require funding from conference sponsors and funding bodies to succeed. Organizers must consider this challenge in their conference funding strategy.

3 ORGANIZERS SHOULD ENGAGE WITH RESEARCH STAKEHOLDERS INCLUDING IMPACTED COMMUNITIES TO UNDERSTAND HOW CONFERENCES CAN EMPOWER THEM

- Participants noted the importance of creating space at conferences for individuals who are impacted by AI systems.
- Conference organizers could experiment with ways to engage a diverse range of research stakeholders—from representatives of civil society organizations to data enrichment workers to people directly impacted by AI technologies—in a way that recognizes and empowers those who do not otherwise get credit for their contribution.
- This could include inviting these speakers to give talks and speak on panels, prioritizing their travel costs in conference budgeting, and engaging with impacted communities to better understand what kinds of support and compensation they would find helpful.

4 ORGANIZERS COULD SPOTLIGHT EXCEPTIONAL TECHNICAL AND ETHICALLY SOUND SUBMISSIONS

- To encourage more ethical discussion at AI/ML conferences, workshop participants suggested more recognition for research that engages with ethical considerations, as clear examples of best practice for others in the field.
- Conference organizers could publicize and present awards for excellent ethical work: for example, named awards; best paper awards; junior researcher awards; or poster awards. They could create dedicated space within conferences, including keynote addresses, for authors who the organizers believe have done excellent work in this regard. Organizers can consider other incentives in addition to awards such as free registration.

5 CONFERENCE ORGANIZERS COULD INCENTIVISE MORE DELIBERATIVE FORMS OF RESEARCH BY ENACTING POLICIES SUCH AS REVISE-AND-RESUBMIT AND ROLLING SUBMISSIONS

- Participants noted the field of AI research as a whole values speed, novelty, and incrementalism, creating incentives to publish quickly which may mean that it is more difficult to introduce ethical interventions, as these may be perceived as adding unnecessary friction.
- Conference organizers should take into account that the culture of ‘publish or perish’ is a widespread problem in AI research, and could find ways to incentivise slower research by experimenting with rolling submission deadlines, or by introducing a policy to revise-and-resubmit papers.

INTRODUCTION

Once considered science fiction, artificial intelligence (AI) systems now play an increasingly prominent role in daily life: from seemingly mundane tasks like recommending Netflix shows to high-stakes tasks like predicting a patient's healthcare diagnosis.

AI systems refer to a broad range of applications in which some predictive, analytic or decision-making capacity is delegated to computer systems. The increasing prevalence of AI systems creates added responsibility for researchers, developers, and deployers of these systems to ensure they operate safely and in line with social values.

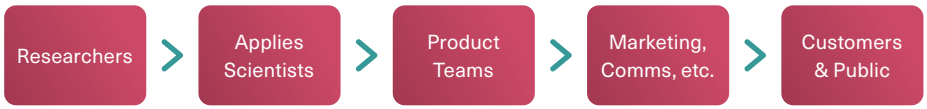
In February 2022, CIFAR, the Partnership on AI and the Ada Lovelace Institute hosted a workshop on ethical review practices for AI/ML conferences. This workshop convened conference organizers past and present, ethical review experts, and AI and machine learning researchers to discuss the important role conference organizers have to play in developing a culture of ethical AI, share the ethical review practices they have implemented or experienced, discuss how they can be used effectively, and ideate what kinds of other interventions might be useful.[1] This report documents this workshop, which we hope will be of use to future AI conference organizers in thinking through their conference submission process, agenda for events, and training opportunities for reviewers and attendees. We thank all the participants for their thoughtful and candid contributions.

Previous work from CIFAR, the Partnership on AI, and the Ada Lovelace Institute has examined the different kinds of ethical challenges that arise in stages of the 'AI lifecycle'—from early allocations of funding to the research, development, and deployment of an AI system. In particular, these institutions have produced research that focuses on AI research: an early stage of this AI lifecycle in which certain ethical risks can be baked into a dataset or AI system and made opaque for downstream uses (a concept sometimes referred to as 'ethical debt'). Previous work includes research into responsible publication practices, how to sustain a culture of ethical AI practices, and the challenges that research ethics committees are facing in enabling responsible AI practices.

[1] A full list of workshop participants can be found in Appendix I.

A key concern highlighted by this work is the need for stronger incentives for researchers to consider the ethical implications of AI research. The development of research questions, the nature of research work, and the publication and dissemination of research findings are all potential sites for ethical questions to arise, as well as drivers and influencers of ethical concerns at other stages of the AI lifecycle. However, many researchers face pressure to 'publish or perish' and' engage in 'fast science,' and engagement with the ethical implications of research is often not encouraged or rewarded by conferences, institutions, and journals.

GRAPHICAL OVERVIEW OF THE AI RESEARCH-TO-PRODUCT LIFECYCLE

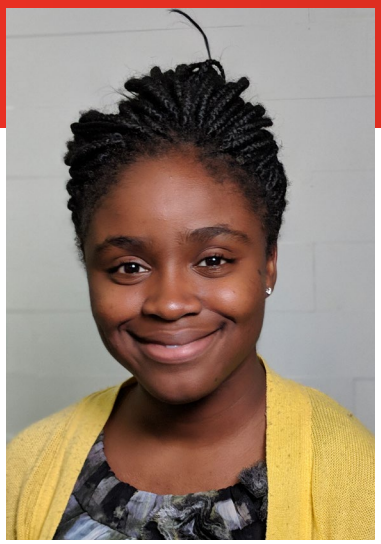


From Hanna Wallach, Navigating the Broader Impacts of Machine Learning Research.

Amongst the main drivers of incentives for the AI research community are organizers of major AI conferences. Conferences in AI research (and in computer science more broadly) play a similar role to academic journals in other disciplines: papers are accepted based on peer review and presented papers are archived for the academic record. As a result, AI conferences determine which work in this field is published and recognized: consequently, the review process by which papers are accepted to these conferences presents an opportunity to create greater incentives for researchers to examine the ethical implications of their work.

FIRST PLENARY SESSIONS 1 & 2:

CURRENT ETHICAL REVIEW PRACTICES AT AI CONFERENCES



Inioluwa Deborah Raji
Mozilla Foundation

SESSION 1

Traditionally, AI conferences have not considered the broader societal impacts of research in determining whether to accept or reject a paper. Inioluwa Deborah Raji (Mozilla Foundation) opened the workshop's first session with a presentation on the development of ethical review practices at Neural Information Processing Systems (NeurIPS), one of the largest AI conferences in the world, for which she co-chaired the Ethics Review process in 2021 with Samy Bengio.

In 2020, NeurIPS conference organizers introduced a requirement for submissions to include a short statement evaluating the potential broader societal and environmental impacts of their work. The organizers introduced this feature out of an awareness that AI and ML research has increasing impact in real world settings, and that researchers should consider not just the beneficial applications of their work but also 'potential nefarious uses and the consequences of failure.'

A SNAPSHOT OF THE AI CONFERENCE PROCESS

The 2020 introduction of the broader societal impact statement at NeurIPS was greeted with confusion, skepticism and even hostility by some researchers; it became clear that some AI researchers did not see the broader societal implications of their research—such as the potential carbon footprint of their system, or the potential ways their research could be used in ways that cause harm to particular demographic groups—as their responsibility. These responses made it clear that many machine learning researchers still adhered to what Deb calls the 'over-the-wall' paradigm of research: a paradigm in which research is tossed over a 'wall' to engineers, who use this research in their development of tools (which are in turn tossed over another 'wall' to users). In this paradigm, engineers are often considered by researchers to be the ones responsible for the uses to which research outputs and findings are put.

As Deb explained, however, this 'over-the-wall' paradigm ignores the fact that an increasing amount of AI research work presented at conferences has been conducted by or in collaboration with industry, including in-house research teams at companies, so in some cases there is less of a clear 'wall' between research and engineering[2]. This paradigm further ignores the ethical problems that can arise at the research stage.

These could include, for example:

- an inappropriate choice of research topic;
- conflict of interests for the research team;
- harmful research design and/or methodology;
- issues of research integrity and misconduct; or
- legal issues including issues related to copyright and terms of use.

[2] According to the 2022 AI Index, AI publications by companies and academic-industry collaborations have steadily increased since 2010. Daniel Zhang, Nestor Maslej, Erik Brynjolfsson, John Etchemendy, Terah Lyons, James Manyika, Helen Ngo, Juan Carlos Niebles, Michael Sellitto, Ellie Sakhaee, Yoav Shoham, Jack Clark, and Raymond Perrault, "The AI Index 2022 Annual Report," AI Index Steering Committee, Stanford Institute for Human-Centered AI, Stanford University, March 2022. https://hai.stanford.edu/research/ai-index-2022

This over-the-wall paradigm further ignores the ethical problems that can arise at the research stage, including an inappropriate choice of research topic; conflict of interests; harmful research design and/or methodology; issues of research integrity and misconduct; or legal issues.

THE NEURIPS CHECKLIST GUIDANCE FOR SUBMISSIONS

The broader societal impact process was revised for the 2021 conference into checklist guidance for submissions. Of the tens of thousands of papers that were submitted to NeurIPS 2021, 300 papers were flagged by reviewers for additional ethical review. Deb and her colleagues wrote a retrospective blog on the process.

Among their key recommendations were:

- that machine learning research should learn from the ethical approaches used in human subject research;
- research should acknowledge all human participants—including crowdworkers—not just the researchers;
- researchers should disclose and provide documentation for models and datasets;
- conferences should work more closely together to ensure that governance is unified: when the rules are different for different conferences, it is harder to create accountability for ethical problems. Deb noted that there are positive developments in this area: in particular, conference leaders are talking to each other more regularly.

SESSION 2

Margarita Boyarskaya (New York University) gave the second presentation in this session. Presenting work she co-authored titled, *Overcoming Failures of Imagination in AI Infused System Development and Deployment*[3], Margarita talked about 'failures of imagination' in efforts to 'responsibilize' AI systems.

Margarita Boyarskaya
New York University

Current trends towards 'responsible AI,' Margarita shared, tend to focus on principles, and checklists which codify those principles. These checklists, however, tend to be universal: they neglect to consider differences between different kinds of technology, different applications, and different stakeholders in the systems under consideration. More insidiously, she said, checklists tend to obscure the norms which shaped how they were put together. Checklists also implicitly place boundaries around the harms that we can foresee or even imagine resulting from AI systems, the ways that we look for and measure these harms, and the ways that we mitigate them.

In a context in which regulation and oversight mechanisms struggle to keep pace with technological development, she said, researchers are increasingly asked to step in. However, it is difficult to predict the social consequences of a technology, and many ethical issues are so-called 'wicked problems:' dependent on so many different dynamic factors that they are difficult to formulate and understand, let alone solve.

Margarita quoted Arthur C. Clarke on the two ways in which forecasting fails: failure of nerve, which prevents us from seeing new possibilities; and failures of imagination, in which impoverished visions of the future do not capture the complexity of upcoming reality. Margarita noted that the pipeline between the publication of AI research and its use is becoming shorter, meaning that these failures of imagination are becoming more material.

Margarita's work, with her co-authors, sampled NeurIPS impact statements from 2020[4]. This work found a number of emerging leitmotifs: neglecting to consider all relevant stakeholders; outsourcing of ethical responsibilities to other actors, for example attributing harms of 'biased inputs' to a system, not the system itself; conflating technological advances with positive impact; limiting the scope of ethical enquiry only to the subject being researched; emphasizing 'net impact' by listing benefits in order to 'balance out' the harms; and overconfident statements of no harmful impact.

[3] Overcoming Failures of Imagination in AI Infused System Development and Deployment https://arxiv.org/abs/2011.13416

[4] Similar work from Carolyn Ashurst analyzed nearly one thousand former NeurIPS impact assessments from 2020 and came way with three suggestions for future broader impact work: (i) the importance of creating the right incentives, (ii) the need for clear expectations and guidance, and (iii) the importance of transparency and constructive deliberation. See Ashurst, C. et al. (2021) 'AI Ethics Statements—Analysis and lessons learnt from NeurIPS Broader Impact Statements', arXiv:2111.01705 [cs]. Available at: http://arxiv.org/abs/2111.01705 (Accessed: 28 April 2022).

At the same time, however, this research also found more encouraging trends in the impact statements they examined: recognition of uncertainty, consideration of a wide range of stakeholders who may be impacted by research, and an interrogation of known benefits.

Margarita and her colleagues advocate, she said, for context-specific examination of impact: considering the role, vulnerability and agency of different stakeholders, as well as the system affordances; who can access a system; and what the system could disclose or expose. They urge researchers to think about impact in terms of what the impact could be: for example, threats to dignity, agency, representation, physical or emotional well-being, opportunity, or allocation of resources; how the impact happens i.e. whether it is immediate, frequent, or through 'nudging' of different behaviours; and why the impact happens—as a result of a harmful process, or as a result of outcomes from a process.

They also urged researchers to challenge the default framing of technology as benevolent; to reflect on inherent assumptions behind both prescriptive ('allow list') and restrictive ('block list') approaches to anticipating harms; and to consider who is vulnerable to harms.

PRINCIPLES FOR RESPONSIBLE INNOVATION IN AI

Margarita recommended considering a set of principles for responsible innovation in AI:

- Anticipation, in other words thinking systematically about the context of AI applications, their sociotechnical affordances, and the different (and potentially conflicting) interests of stakeholders;
- Reflexivity, i.e. interrogating the commitments and assumptions of researchers themselves, including about the generality and neutrality of base models, and about the beneficence of technology adoption; and
- Inclusion, in the form of seeking out inputs from domain experts and from people affected by the applications of AI systems.

They also urged researchers to:

- Challenge the default framing of technology as benevolent;
- Reflect on inherent assumptions behind both prescriptive ('allow list') and restrictive ('block list') approaches to anticipating harms; and to
- Consider who is vulnerable to harms.

FIRST PLENARY OPEN DISCUSSION

After these presentations, workshop participants engaged in an open discussion. The need for reviewers with relevant expertise within the conference system was one of the key themes that emerged. Establishing a taxonomy of reviewers with different kinds of expertise —for example, being able to assess legal compliance—has proven useful. However, bringing in this interdisciplinary expertise is still difficult. Participants also raised concerns that even a broad network of ethical experts may not be diverse enough—or have the lived experience—to represent the wide range of stakeholders who might be affected by the deployment of AI systems.

To ensure that papers are reviewed by appropriate experts, it has proven necessary to map ethical issues that could arise in the research papers: participants noted that this is a lengthy and complex process. There is recognition that as understanding of AI ethics evolves, there is more understanding of how concerns around broader societal impacts of research differ from research ethics concerns that focus on the methodology of the research, such as whether data used to train an algorithm was obtained with the appropriate consent of the data subjects.

As well as workload, concerns were also raised about a perception that technical review of the performance and reproducibility of the work and ethical review of its method and potential broader impacts are separate processes—and that technical merits outweigh ethical concerns. Within AI research, ethical flaws are not yet widely recognized as being as important as technical flaws. Participants also raised concerns that in a competitive academic environment, the pressure to publish may override ethical concerns.

The review process for papers submitted to major conferences is already a complicated and lengthy process, and concerns were also raised about the need to train reviewers to identify ethical issues so that papers of concern could be flagged to experts: reviewers already have a heavy workload. Participants reported that in some previous conferences, where reviewers had been asked to review for ethical issues, some papers were erroneously flagged as requiring ethical review, while other papers of ethical concern were not picked up in review.

Participants were clear that the goal of ethical review practices at conferences is not to turn every AI researcher into an ethicist or a philosopher: it is about building a culture which supports researchers examining the broader social and environmental impact of their work. There was some discussion about ‘ethical’ versus ‘responsible’ research: participants noted that an ethical framework imposes a moral judgment about the work being done, while a ‘responsible research’ framework could involve examining different ways in which the work could be conducted and used. Researchers publish code and datasets as well as academic papers: it was suggested that these could be accompanied by information about the limitations of their usefulness, or even restrictions that they were made available only for certain uses.

AI conferences are international, but different countries have different laws, customs and contexts which may create differentiation in research practices. Participation in AI conferences is also dominated by participants from (or based in) Global North countries. Some participants noted that, for reviewers, it can be hard to know the ethical principles they should review research against. One suggestion which emerged from the discussion was whether it is possible to use existing shared international norms to guide ethical decision-making in conferences, such as the UN Universal Declaration of Human Rights, or the UNESCO Recommendation on the Ethics of AI. The point was also made that human rights can be a good starting point: research doesn't need to be limited to only complying with what is in the human rights framework, but might also go beyond compliance, to help address human rights issues, global disparities, and global problems.

Participants discussed the merits of a prescriptive approach compared to an encouraging one. Within the field of AI research, some researchers are already enthusiastic to engage with ethical considerations; others are willing but feel ill-equipped to do so; while a further group does not see this as part of their job at all. Education and training initiatives, which equip researchers and conference reviewers with both the understanding and the vocabulary to examine the downstream impact of their research, were seen as key: participants were supportive of a process-oriented approach to help build competency to bring an ethical approach to different parts of the research process. Recognizing the different levels of interest and experience, participants were mindful that education efforts should be constructive rather than critical (or worse, patronizing).

The participants also discussed the merits and disadvantages of checklists as tools for ethical examination. On the one hand, checklists were seen as prescriptive and potentially limiting the scope of enquiry and the questions being asked: they raised concerns that researchers might focus on the specific, concrete items included in checklists, at the expense of more open-ended questions that could prompt reflection and which might surface harder-to-identify issues. Participants were also worried that the implementation of checklists could be seen as sufficient for ethical review and that they might replace ethical impact statements and other tools that create space for a more open-ended form of reflection. On the other hand, participants also noted that with the right questions, checklists can act to foster a more deliberative process, by providing prompts to help researchers surface and discuss different issues at different stages of the research process.

The goal of ethical review practices at conferences is not to turn every AI researcher into an ethicist or a philosopher: it is about building a culture which supports researchers examining the broader social and environmental impact of their work.

SECOND PLENARY:

EXPLORING INTERVENTION OPPORTUNITIES FOR ETHICAL AI

Kristy Milland
Singleton Urquhart Reynolds Vogel LLP

SECOND PLENARY

Kristy Milland (Singleton Urquhart Reynolds Vogel LLP) opened the second plenary session with a presentation on the role of crowdworkers in AI research, drawing on her own experience as a crowdworker, a labour activist, and a labour lawyer.

Kristy spoke about the lack of focus in AI research on the human workers who, via crowdworking platforms like Amazon's Mechanical Turk, create the datasets that are crucial for AI research, and who participate in studies, but who are to a large extent hidden behind academic papers.

Kristy urged academics to consider treating crowdworker research as human subject research. She recommended considering employment relationships, terms of service and privacy for crowdworkers, and that research that uses crowdworkers publicizes not only the instructions that crowdworkers are given but also the terms under which they work and how they are paid.

She urged academics whose research relies on crowdwork to remember that this is work done by human workers and to engage with their communities. When it comes to conferences, she said, there could be a richer understanding if these conferences involved the actual lived experience of the people doing this work, and how they are treated by academics as well as by the platforms through which they work. Concretely, she recommended finding ways to involve crowdworkers which recognizes not only that they may be giving up paid work to participate, but also gives them agency in how they are involved and what they would like academics to hear.

Concretely, she recommended finding ways to involve crowdworkers which recognizes not only that they may be giving up paid work to participate, but also gives them agency in how they are involved and what they would like academics to hear.

SECOND PLENARY OPEN DISCUSSION

One key concern emerging from the discussion that followed Kristy's presentation was that unethical practices can leave crowdworkers vulnerable: participants suggested that disclosing crowdwork research practices can help with accountability, and also provide examples of good practice for using crowdwork in different ways and across different platforms.

Participants also favoured bringing crowdworkers themselves into the AI research ecosystem: this could include participatory research and/or co-creation of guidance for research that uses crowdworkers. Participants also noted the need to consider how crowdworkers want to be included, including recognizing that participation in conferences, for example, may not be valued as highly by crowdworkers as by academic researchers.

While the AI research community is increasingly recognizing the ethical issues that can arise when using crowdwork, participants also noted that other parts of the research ecosystem actively incentivize the use of crowdwork. In particular, some Institutional Review Boards treat crowdwork research as separate from human subject research and are quicker to approve the former. In a fast-moving field like AI research, this can encourage researchers to use crowdwork: participants highlighted the need to resist this, as well as more generally the need to resist incentives to move fast in research.

Conferences, as key sites for disseminating AI research, could play a role in addressing the ethical challenges of using crowdwork. Participants highlighted that the Association for Computational Linguists (ACL) has championed making visible the role of crowdsourced annotations in research. One emerging recommendation from the discussion is to require IRB ethical review for research using crowdworkers in order to submit this research to conferences.

FINAL WORKSHOP:
BIG IDEAS

In the final workshop session, participants split into three breakout groups, to examine in more detail specific interventions that conference organizers could put in place to address ethics while also considering under-represented and marginalized communities. The full list of ideated interventions can be found in Appendix II, but we draw particular attention to five interventions that the workshop participants worked through in more detail.

1 AI CONFERENCE ORGANIZERS CAN CONSIDER A MIX OF PRESCRIPTIVE AND REFLEXIVE INTERVENTIONS TO IMPROVE RESEARCHERS' ABILITY TO ASSESS THE ETHICAL IMPACTS OF THEIR WORK

Participants agreed that organizers should experiment with a mixture of different prescriptive tools, such as ethics checklists that list out the kinds of issues researchers should address, and open-ended reflexive interventions like impact statements. Participants were broadly in agreement that 'carrots are better than sticks:' positive incentives for ethical work will work better than punishment for ethical violations. Participants discussed the need to link different interventions together: for example, training at conferences to help authors write better impact statements.

Ethics checklists can limit the scope of enquiry, missing out on issues that are harder for researchers to identify. Nonetheless, prompts in the checklist play an important role in kick-starting the ethical reflection process.

Participants recommended that open-ended exercises meant to encourage reflexivity among researchers including an examination of their assumptions, practices, and commitments—must also be encouraged to provide a stronger incentive for reflective consideration of impacts.

2 CONFERENCE ORGANIZERS SHOULD PRIORITIZE TRAINING MORE RESEARCHERS AND CONFERENCE REVIEWERS ON HOW TO EXAMINE THE POTENTIAL NEGATIVE DOWNSTREAM CONSEQUENCES OF THEIR WORK

This would involve conference organizers setting up training for (a) ethical reviewers of papers submitted to the conference, and (b) members of the research community attending the conference around common ethical issues and mitigations.

Workshop participants noted that developing this training would require an understanding of what conference participants know, want to know, and should know: it was the opinion of participants that this training would likely need to start at a basic level until there is a widespread norm of reflexive ethical considerations. Participants recommended that to be most effective and useful to conference participants, they should be run as exercises, not lectures, and advised conference organizers to consider the risks and benefits of making training either optional or mandatory, warning that if trainees feel hectored, they are less likely to engage with the ethical examination of their work.

This training would require experienced trainers and teaching materials such as case studies. They could be supported by shared resources between different conferences, including a list of subject-matter experts that organizers can call on for specific ethical guidance, and a list of datasets that have been identified as being problematic or requiring additional ethical review. To be most effective, this training should be connected to other interventions such as impact statements.

3 ORGANIZERS SHOULD ENGAGE WITH RESEARCH STAKEHOLDERS INCLUDING IMPACTED COMMUNITIES TO UNDERSTAND HOW CONFERENCES CAN EMPOWER THEM

Participants also suggested organizers should experiment with engaging different communities impacted by AI systems in AI/ML conferences. This could take the form of invited talks and panels to workshops that explore how research can lead to unintended outcomes or potential harms.

Examples of stakeholders could include representatives from civil society organizations, activists, or members of the public. This should also include crowdworkers and data enrichment workers, whose contributions to AI research are often not recognized.

These kinds of interventions can help foster more discussion of the kinds of risks and harms that AI systems can produce, and incentivize more knowledge transfer between researchers and individuals with lived experience.

Participants noted that organizers must take great care to avoid extractive forms of knowledge transfer. Conference organizers should compensate these individuals and cover travel costs and should ask these individuals what kinds of compensation or reciprocity they would like for their time.

4 ORGANIZERS COULD SPOTLIGHT EXCEPTIONAL TECHNICAL AND ETHICALLY SOUND SUBMISSIONS

To raise the standard of ethical discussion at AI/ML conferences, workshop participants suggested more rewarding and recognition of research that engages reflexively with ethical considerations, as clear examples of best practice for others in the field.

This could involve a special category of an award on par with 'Best Paper' which rewards careful ethical reflection and documentation of that thinking. Conference organizers could advertise and present awards for excellent ethical work—for example, named awards, best paper awards, junior researcher awards, and poster awards—that could create dedicated space within the conference (such as a keynote) for authors who the organizers believe have done excellent work in this regard, and could accompany these rewards with other incentives such as free registration. Identifying candidates for these awards would require the availability of appropriate reviewers with ethical expertise.

Conversely, organizers could exclude papers from winning other conference awards if they do not have a strong ethical awareness: this could form part of supporting community norms that prioritize ethical expertise in hiring or career advancement.

Supporting authors to compete for these awards could include providing assistance for excellent technical papers to improve their ethical stance, and allowing time for authors to revise ethical statements. It could also include encouraging the research community to vote for the best papers. Workshop participants noted that additional support might be needed for fundamental research papers, for which it may be harder to write a strong ethical statement.

5 CONFERENCE ORGANIZERS COULD INCENTIVIZE MORE DELIBERATIVE FORMS OF RESEARCH BY ENACTING POLICIES SUCH AS REVISE-AND-RESUBMIT AND ROLLING SUBMISSIONS

The pace of publishing in AI research creates space for dynamic conversations, but the field as a whole values speed, novelty, and incrementalism, creating incentives to publish quickly which may mean that it is more difficult to introduce ethical interventions, as these may be perceived as adding unnecessary frictions.

Workshop participants noted that in this respect, the AI research field could learn from the benefits and the challenges faced by research in other fields, where the pace of publication can be much slower. Participants noted that 'fast research' is a broad problem, and that there is a need for more senior researchers to role-model slower research for their peers and junior colleagues.

At a conference level, slower and more deliberative research could be facilitated by moving to rolling submissions deadlines (as has already been done at Conference On Computer-Supported Cooperative Work And Social Computing), or by (re)introducing the possibility to revise-and-resubmit papers, to allow authors to reflect on and incorporate feedback. Noting that authors frequently submit rejected papers to other conferences, participants also suggested considering requirements for authors to include reviews from other venues as part of their submission, to ensure that feedback is taken into account.

APPENDIX I:
LIST OF PARTICIPANTS

—

We'd like to thank the following people for contributing to our workshop that informed the production of this report:

Grace Abuhamad, Lead, Trust & Governance Lab, Service Now, Canada

Solon Barocas, Principal Researcher, Microsoft Research, United States

Samy Bengio, Senior Director of Machine Learning Research, Apple, United States

Margarita Boyarskaya, Ph.D. Candidate, New York University, United States

Alexandra Chouldechova, Principal Researcher, Microsoft Research, United States

Danish Contractor, Senior Research Scientist & Manager, IBM Research, India

Kate Crawford, Senior Principal Researcher, Microsoft Research, United States

Ravit Dotan, Post-doctoral Student, University of Pittsburgh, United States

Heather Douglas, Associate Professor, Michigan State University, United States

Rebecca Finlay, Chief Executive Officer, Partnership on AI, Canada

Alona Fyshe, Canada CIFAR AI Chair, Assistant Professor, University of Alberta, Canada

Brent Hecht, Director of Applied Science, Microsoft, United States

Rose Landry, AI Ethics Lead, Montreal Institute of Artificial Intelligence, Canada

Sasha Luccioni, Research Scientist, Hugging Face, Canada

Miguel Luengo-Oroz, Senior Advisor, United Nations Global Pulse, Spain

Kristy Milland, Articling Student, Singleton Urquhart Reynolds Vogel LLP, Canada

Jason Millar, Canada Research Chair in Ethical Engineering of Robotics & AI, University of Ottawa, Canada

Elissa Strome, Executive Director, Pan-Canadian AI Strategy, CIFAR, Canada

Joelle Pineau, Managing Director, Meta AI Research (FAIR); Associate Professor, McGill University, Canada, Canada CIFAR AI Chair

Benjamin Prud'homme, Executive Director, AI for Humanity, Montreal Institute for Artificial Intelligence, Canada

Inioluwa Deborah Raji, Fellow, Mozilla Foundation, United States

Marc'Aurelio Ranzato, Research Scientist, DeepMind, United Kingdom

Sarah Rispin Sedlak, Lecturing Fellow, Duke University Initiative for Science & Society, United States

Francesca Rossi, AI Ethics Global Leader, IBM, United States

Gabrielle Samuel, Research Fellow, King's College London, United Kingdom

Alexandra Schofield, Assistant Professor, Harvey Mudd College, United States

Toby Shevlane, Researcher/Ph.D. Student, University of Oxford, United Kingdom

Graham Taylor, Professor, Canada CIFAR AI Chair, University of Guelph / Vector Institute, Canada

Joel Zylberberg, Canada Research Chair in Computational Neuroscience, York University, Canada, Associate Fellow, CIFAR

The indicated affiliations are accurate at the time of the workshop (February 2022). While individuals representing many organizations participated in the workshop, the report should not be read as representing the views of any specific organization. Contributions from individuals do not necessarily reflect the views of their employers.

WORKSHOP ORGANIZERS

Andrew Strait, Associate Director, Ada Lovelace Institute, United Kingdom

Fiona Cunningham, Director of Research, CIFAR, Canada

Johnny Kung, Senior Officer, Knowledge Mobilization & Publications, CIFAR, Canada

Gagan Gill, Program Manager, AI & Society, CIFAR, Canada

Madhulika Srikumar, Program Lead, Partnership on AI, US

EVENT LOGISTICS

Jacqui Sullivan, Director of Meeting & Events, CIFAR, Canada

Joshua Pikus, Communications & Events Coordinator, Partnership on AI, US

LEAD RAPPORTEUR

Laura Carter, University of Essex, United Kingdom

APPENDIX II:
PROPOSED INTERVENTIONS

In addition to the five ideas worked through in detail and discussed in the 'Big Ideas' section above, workshop participants brainstormed potential interventions that conference organizers could put in place to encourage more ethical reflection about practices undertaken during research and potential downstream impacts after the research is conducted. These ideas were not fully discussed and analyzed in the workshop but are included here as they offer additional interventions for conference organizers to consider and experiment with.

There was recognition that there is considerable work to be done in improving ethical practices in AI conferences, but at the same time, a sense that major steps forward have already been taken across much of the field, creating a strong base on which to build.

1 Conference ethics track leads could experiment with coordinating with organizers of other conferences to create consistent, community-wide guidance, practices and enforcement mechanisms for ethical review of research.

2 Conference committees responsible for invited talks and panels could enable the participation of a diverse range of research stakeholders in a way that recognizes and empowers those who do not otherwise get credit for their contribution. This could include inviting these speakers to give talks and participate in panels, prioritizing their travel costs in the conference budget, and engaging with those communities to ask what kinds of support and compensation they would find helpful. Examples of these speakers include data enrichment workers, representatives from civil society and advocacy organizations, and people who are affected by research applications

3 Awards committees could experiment with recognizing papers and researchers that achieve a high degree of ethical excellence as a way to incentivize researchers to engage in these practices. This could take the form of papers that demonstrate a high degree of ethical reflexivity in a broader societal impacts section.

4 Conference organizers could explore steps to incentivize more deliberative research including offering revise-and-resubmit options.

5 Conference organizers could create guidance for reviewers to help them identify ethical as well as technical problems with papers, including a shared database of known 'problematic' datasets that should trigger additional review.

6 Conference organizers could recruit, develop and support a network of expert ethics reviewers with a range of expertise to assess ethical problems identified in review.

7 Conference organizers could work with experts in AI and ML ethics to organize ethics training workshops and initiatives for researchers. This could include training sessions for reviewers and for conference attendees to meet the needs of different researchers with different levels of expertise, experience and willingness to engage with the ethics of their research.

8 Conference organizers could experiment with ways to engage with Institutional Review Boards to enforce effective ethical standards on AI research, including recognizing that crowdworker research is human subject research and considering it accordingly.

—

Other possible interventions raised by the participants for future conference organizers to consider include:

RECOGNIZING EXAMPLES OF BEST PRACTICES ADOPTED BY AI/ML RESEARCHERS

- Spotlighting exceptional ethical (as well as technical) work
- Recognizing interdisciplinary co-authorship
- Showcasing testimonies and reports from successful participants in co-created work
- Awards for best impact statements

SUPPORTING ETHICAL REFLECTION BY SUBMITTING AUTHORS

- Office hours to help researchers develop ethical statements
- Incentivising inclusion of diverse research groups
- Ethics training sessions during conferences
- Development of code of conduct for online data collection
- Supporting IRBs to strengthen their review processes for crowdwork research
- Include carbon footprint assessments in ethics assessments

CONFERENCE GOVERNANCE

- Collaborating between the organizers of different conferences, to eventually lay the foundation for creating a community-level approach to ethics review practices at ML conferences
- Including ethical principles in codes of conduct
- Participation of research stakeholders in the conference design process
- Encouraging interdisciplinary tracks/submissions
- Invited talks, panels and discussion sessions on social science and ethics
- Consequences and enforcement for violations of codes of ethics
- Broader support for attendees to enable more diverse participation e.g. childcare, appropriate lodging
- Developing ethics guidelines in consultation with diverse communities

REVIEW PROCESS

- Recruiting reviewers with diverse ethical expertise
- Requiring discussion of societal impacts in papers
- Guidelines for reviewers on evaluating impacts including assessing claimed benefits
- Requiring disclosure of terms for crowdworker research including informed consent, pay rates, rejection rates
- Providing extra 48-72 hours after the submission deadline for authors to submit ethics review statements
- Requiring authors to take a module on ethics in order to submit papers
- Requiring declarations of interests
- Offering optional ethical review before conference deadlines
- Including representation from underrepresented and marginalized communities among reviewers

RESISTING FAST RESEARCH OR 'PUBLISH OR PERISH' CULTURE

- Incentivizing slow, thoughtful research through 'revise and resubmit' and rolling submissions
- Learning from other fields which allow submissions to be revised and resubmitted (instead of solely accept/reject)

EDUCATION

- Highlighting examples of known ethical concerns to a wider audience
- Distinguishing between the actions of surfacing ethical concerns and actually solving them and making it clear where the emphasis should be placed in different use cases
- Providing materials to alleviate concerns about perceived trade-offs between duties to the research community and society in general

APPENDIX III:
RESOURCES SHARED BY ATTENDEES

- Amazon Mechanical Turk: Gold Mine or Coal Mine?
- Arizona State University Socio-Technical Integration Research
- Behavioral Use Licensing for Responsible AI
- Boundaries Between Research Ethics and Ethical Research Use in Artificial Intelligence Health Research
- Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing
- Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI
- Guidelines for Academic Requesters (written by workers and academics as a team)
- It's Time to Do Something: Mitigating the Negative Impacts of Computing Through a Change to the Peer Review Process
- Johns Hopkins Berman Institute - Research Ethics Consultation Service
- NeurIPS 2020 Invited Talk: A Future of Work for the Invisible Workers in A.I.
- NeurIPs Code of Ethics
- Responsible AI Licenses
- Responsible Sourcing of Data Enrichment Services
- A Retrospective on the NeurIPS 2021 Ethics Review Process
- Turk-Life in India
- Turkopticon: From Software to Organizing
- UMass Amherst MTurk Guidance
- The Values Encoded in Machine Learning Research
- We Are Dynamo: Overcoming Stalling and Friction in Collective Action for Crowd Workers
- Worker Demographics and Earnings on Amazon Mechanical Turk: An Exploratory Analysis
- World Economic Forum Young Scientists — Code of Ethics