How to be responsible in AI publication

A white paper from Partnership on AI provides timely advice on tackling the urgent challenge of navigating risks of AI research and responsible publication.

AI research has quickly developed in the past two decades from a niche topic to one that has transformed whole areas of technology, society and scientific research. Remarkably, much of the work has taken place without much ethical oversight. For biomedical researchers, in contrast, ethical considerations are an essential part of the research lifecycle. The field of bioethics began in the 1970s and led to the development of the Belmont Report by the US National Commission for the Protection of Human Subjects of Biomedical and Behavioural Research[1]. The report is currently still an essential reference for institutional review boards (IRBs) and ethical committees in the process of ethical approval of biomedical research projects. Today, bioethics is a complex topic. For instance, informed consent of study participants is an important element in the Belmont Report, but in the current era of big data, human data can be collected at scale while obtaining consent is often impractical. A series of other technological advances have enabled transformative capabilities in the life sciences, including genetic manipulation, cloning and stem cell research. With each new capability that is developed, new questions arise about what are acceptable research directions, and biomedical research communities and societies duly get together to develop bespoke rules, engaging in outreach and consultation[2].

AI research has leapfrogged into a similar situation of ethical complexity. AI algorithms and data are frequently shown to harbour biases and amplify society's injustices, and tools have become available that can have harmful impact on individuals and groups. Examples are facial and other biometric recognition in surveillance applications, AI algorithms deployed in high-stakes decisions directly affecting individuals' lives and deepfake videos and text that can spread misinformation at a large scale. Overall ethical guiding principles like the Montreal Declaration are useful, but the growing risk of harmful impact caused by AI applications indicates a need for the AI community to reflect in a systematic way on downstream consequences of the research they are undertaking. Partnership on AI (PAI), an organization that brings together research institutions, companies, civil society organizations and others to discuss best practices in AI research and to foster public dialogue, has published a white paper that focusses on the specific challenge of ethical reflection in the dissemination of AI research to the wider public[3]. In the AI community, 'dissemination of research' includes several things: journal or conference papers, preprints, blogposts, code repositories and more. PAI's "Managing the risks of AI research: Six recommendations for responsible publication" is written with involvement from various advisors, and *Nature Machine Intelligence* has also contributed.

The white paper provides recommendations for individual researchers, research leadership and for journals and conferences, with a focus on considering downstream consequences of AI research, including unintended applications and malicious use. The paper highlights the importance for authors to provide a transparent discussion of motivation and contribution, and recommends including a statement on downstream consequences that should be as detailed as needed, relative to the level of advance provided by the work. A list of resources is provided as a starting point. There should also be a clear statement of the amount of computational resources used, which is important for reproducibility, for disclosing an important part of research methodology and for downstream consequences in particular regarding environmental impact. Journals and conferences should aim to make engagement with downstream consequences part of existing peer review processes, and establish separate evaluation processes for papers where serious risks are identified. Research leadership should ensure that a reflection on downstream consequences is integrated early in the research pipeline and provide an environment for open discussion where researchers are commended for identifying ethical and societal implications, including negative ones.

The purpose of the white paper is to provide ideas and resources as well as to initiate further discussion. Partnership on AI welcomes feedback and plans on future iterations. The paper acknowledges that a wide range of views exist and that community consensus is lacking. On the other hand, the pace of research is such that the time for action from researchers, leadership, conferences and publishers is now. We agree and have begun asking for an ethical and societal impact statement in papers that involve identification or detection of humans or groups of humans, including behavioural and socio-economic data.

Lots of work is to be done. A challenging outstanding task mentioned in the white paper is deciding what factors contribute to low or high risk in AI research, and when are measures such as additional review to assess the risk of dual use and redaction of certain information warranted. Furthermore, it is noteworthy that the white paper disentangles the topic of research integrity from reflection on downstream consequences, while adding that in practice the areas overlap. We believe these are in fact strongly connected: from the first idea of a research project, the formulation of motivations and goals and the decisions on the research methods, a responsible approach is required. More transparency is required around the use of datasets, with closer attention to issues of consent, choices made in collecting data for training, testing and validation and potential sources of bias. A good start will be to consider the adoption of 'datasheets for datasets' by Timnit Gebru and colleagues[4].

Ethics of research is now everybody's business, remarks the author of the essay in ref. [2], referring to researchers, journal editors, funders, and others in biomedical research. It is time that the AI community makes ethics their business too. ❐

Published online: 19 May 2021
https://doi.org/10.1038/s42256-021-00355-6

References

1. *The Belmont Report* (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979).
2. Franklin, S. *Nature* **574**, 627–630 (2019).
3. *Managing the Risks of AI Research* (Partnership on AI, 2021); https://www.partnershiponai.org/responsible-publication-recommendations/
4. Gebru, T. et al. Preprint at https://arxiv.org/abs/1803.09010 (2018).