The prospect of artificial intelligence-supported ethics review

Philip J. Nickel, Eindhoven University of Technology p.j.nickel@tue.nl

Abstract: The burden of research ethics review falls not just on researchers, but on those who serve on research ethics committees (RECs). With the advent of automated text analysis and generative artificial intelligence, it has recently become possible to teach models to support human judgment, for example by highlighting relevant parts of a text and suggesting actionable precedents and explanations. It is time to consider how such tools might be used to support ethics review and oversight. This commentary argues that with a suitable strategy of engagement, artificial intelligence can be used in a variety of ways that can genuinely support RECs and manage their workload and improve the quality of review. It would be wiser to take an active role in the development of AI tools for ethics review, rather than to adopt ad hoc tools after the fact.

In the last 75 years, ethics review of human subjects research has become a norm for most scholarly and medical institutions worldwide. The accompanying institutional apparatus is costly and time consuming. Discussion of the burdens of research ethics review has usually focused on administrative hurdles, paperwork and delays as the process unfolds (Meenaghan et al. 2007), and on tradeoffs between ethics requirements and study design such as the impact of consent bias (Rothstein & Shoben 2013). However, the burden of review also falls on those who serve on research ethics committees (RECs) as practicing physicians and scientists, professional ethicists, patient and administration representatives, legal experts, and lay advisors. These jobs are often carried out as sidelines to the committee members' main work in a hospital, university, or elsewhere. The task of reading and analyzing dozens of dossiers of forms, study protocols, and consent forms, and writing comments for the committee and the researcher, is a kind of invisible muscular activity animating the skeleton of the review process.

With the advent of automated text analysis and generative artificial intelligence, it has recently become possible to teach models to support human judgment, for example by highlighting relevant parts of a text and suggesting actionable precedents and explanations. Generative AI can even be used to compose text. The legal field has begun to adopt these forms of automation, with forward-thinking law firms already exploring ways of using machine learning to support efficient legal analysis (Kluttz & Mulligan 2019) and law schools incorporating AI into the curriculum as an integral part of the future of the profession.

It is time to consider how such tools might also be used to support ethics review and oversight. The potential advantages of using artificial intelligence for ethics review are many.

First, it could reduce the burden on REC members and universities, allowing more focus on learning and best practices and less on the rote aspects of review. Second, it could lead to more consistent and fairer review process across contexts, times, and committee compositions. Third, it could supply information for a dashboard that allows REC chairs and managers to better plan the logistics of the review process, by signaling dossiers likely to be time- or discussion-intensive. The ideal of how AI tools might provide this support is neither to replace human work, nor to make humans the managers of AI tools, but rather a "third way" in which machine learning "amplifies" human capacities (Bankins & Formosa 2023). This third way could be applied to the tasks of human reviewers, amplifying their awareness of the salience of certain points in a dossier and suggesting precedents or principles relevant to these points. From experience, reviewers usually know what points in a dossier are the most salient and have routines about where in the dossier to begin in order to focus on the most salient elements. Machine learning could build on this knowledge in ways that were an extension of the human reviewer's agency and authority.

Pitfalls and limitations of AI-based automation include the possibility that reviewers will become over-reliant on automation — so-called "automation bias" (Goddard et al. 2012) — and fail to focus on important details of a dossier or fail to think critically about its operation. Moreover, members of RECs have specialized roles; each person's job is to look at the study from a particular disciplinary angle. It is not straightforward to use a single decision-support tool to assist people in any and all of these roles. Furthermore, ethics review is a multi-step process, and automated task support is no substitute for a meaningful dialogue between reviewers and the researchers resulting in commitments to improving the study. There are human elements of trust, accountability, and moral judgment that cannot effectively be automated, but are an essential part of institutional ethics review. As Roy-Toole (2011) writes, "Ask a member of a research ethics committee how the quality of research governance decisions could be improved … and you might be told to look to ways to improve the substance of the decision rather than the speed with which it could be delivered." Innovations should be "human-centered" in the sense that they lead to better ethics governance rather than a technological imperative of ever-increasing automation (Schneiderman 2020).

However, the existence of pitfalls and limitations does not imply that we should not shape these technologies to fit our purposes. The lived experience of ethics review is detailed reading and note taking on hundreds of pages of typed materials, followed by a sometimes hurried discussion of multiple protocols within the short time allocated on a review board's agenda. Innovations that allow ethics committees to identify and focus on the most meaningful issues requiring discussion and problem-solving will surely be welcomed if they can prove themselves. The first step is to determine which parts of the review process are most plausibly supported by automation, bringing data scientists and ethics committees together to discuss the possibilities. The second step is to systematize data collection from researchers in ways that enable text analysis, contain meaningful annotations from ethics reviewers, and are sufficiently representative of the "destination" context to provide a basis for decision support at that context. Only after these first two steps would the technical process of training and prototyping a working system begin.

In winning the trust of ethics review boards for such a project, it is crucial to explore the possibilities of automation in an inclusive and open manner, remaining aware of possible hidden agendas of pure efficiency or regimentation that could compromise quality, and taking steps to ensure that the discretion and authority of ethics committees is supported rather than undercut (Nickel 2022). Because university and hospital research ethics review processes take place in a public, not-for-profit setting in the interests of science, it is reasonable to expect an empirical, transparent attitude toward the process of developing

support tools. The alternative to doing so in a self-directed way might be to eventually purchase commercial software with similar functionalities down the line, which might be less transparent and participatory in how it is developed and introduced.

Some may question how we could arrive at a form of research ethics governance so paperwork-intensive and time-consuming that powerful assistive tools would be considered for assisting with it. This is a discussion worth having, but not directly relevant to the present purpose. It takes years to reform ethics review requirements and procedures, and the number of studies for ethics review is not likely to decrease during that period, nor will the hours of focused attention available to REC staff. Ethics review already relies heavily on information and computer tools. Many reviewers already use search and translation software that incorporates machine learning tools, and these are not rejected on the grounds that we should not have put ourselves in a position to want to use them.

An additional feature of machine learning for ethics review is that it could be applied to a dossier before it is submitted, allowing researchers to preview the parts of their application that are likely to draw the attention of a committee. This can be seen as a risk or an opportunity, depending on the underlying culture of ethics in the hosting institution(s) and the way AI is embedded in review (Akaah 1993). The worry is that it might be used "adversarially" to game the ethics review process, cleverly avoiding language that triggers a flag of some kind. Many machine learning tools raise this possible "dual use" concern (Brundage et al. 2018). On the other hand, it could be seen as beneficial to have a tool that novices to ethics review can use to learn more about what aspects of a study are salient — even when their research idea is purely hypothetical or speculative, and even when early-stage students of empirical science or design science are considering potential impacts on human subjects in a classroom setting. Perhaps this is the best place to include such assistive tools: before the REC ever sees a dossier. Again, collaboration with diverse stakeholders in an early phase of thinking about these matters, and experimentation with different institutional embeddings, will be useful in drawing conclusions about the best implementations. And again, if ethics review committees do not lead the way in developing such tools, there is a risk that the tools will be developed without their input and used in ways that they do not foresee or influence. It would be surprising if that were not already happening with existing tools such as Chat-GPT.

A recent policy brief of the Association for Computing Machinery predicts that eighty per cent of U.S. workers will have at least ten per cent of their tasks affected by generative AI (Leslie & Rossi 2023). The kind of work most liable to be affected is text-intensive work. University research ethics committees are already actively considering how best to provide governance for the development of artificial intelligence (Hine 2021). It is incumbent on leaders within research ethics community to shape the adoption of machine learning tools to their benefit. Otherwise, adoption of some sort may happen without their leadership.

References:

Akaah, I.P. (1993). Organizational culture and ethical research behavior. *Journal of the Academy of Marketing Science* 21, 1: 59-63.

Bankins, S. & Formosa, P. (2023). The ethical implications of artificial intelligence (AI) for meaningful work. *Journal of Business Ethics* 185: 725-740. DOI: 10.1007/s10551-023-05339-7

Brundage, M., et al. (2018). *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. ArXiv Preprint. https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf Accessed 10 October 2023.

Goddard, K., Roudsari, A., & Wyatt, J.C. (2012). Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association* 19: 121-127. DOI: 10.1136/amiajnl-2011-000089

Hine, C. (2021). Evaluating the prospects for university-based ethical governance in artificial intelligence and data-driven innovation. *Research Ethics* 17: 464-479.

Kluttz, D.N., & Mulligan, D.K. (2019). Automated decision support technologies and the legal profession. *Berkeley Technology Law Journal* 34: 853-890. DOI: 10.15779/Z38154DP7K

Leslie Rossi (2023). *ACM TechBrief: Generative Artificial Intelligence*. ACM Technology Policy Council Issue 8.

Meenaghan, A., O'Herlihy, A., Durand, M.A., Farr, H., Tulloch, S. & Lelliott, P. (2007). A 55 kg. paper mountain: the impact of new research governance and ethics processes on mental health services research in England. *Journal of Mental Health* 16, 1: 149-155.

Nickel, P.J. (2022). Trust in medical artificial intelligence: a discretionary account. *Ethics and Information Technology* 24. DOI: 10.1007/s10676-022-09630-5

Rothstein, M.A., & Shoben, A.B. (2013). Does consent bias research? *American Journal of Bioethics* 13, 4: 27-37.

Roy-Toole, C. (2011). Passing the buck: how the Academy of Medical Sciences's 'New pathway for the regulation and governance of health research' shifts the regulatory burden but fails to improve the quality of research governance. *Research Ethics* 7, 3: 82-90.

Schneiderman, B. (2020). Human-centered artificial intelligence: reliable, safe, and trustworthy. *International Journal of Human-Computer Interaction* 36, 6: 495-504. DOI: 10.1080/10447318.1741118