


# Ethics Review of AI research: An approach to reviewing and revising existing governance structures

Research Ethics  
2026, Vol. 22(2) 357–370  
© The Author(s) 2025  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/17470161251406299  
journals.sagepub.com/home/rea  


Angeliki Kerasidou<sup>1</sup> , Karen Melham<sup>1</sup>, Alberto Giubilini<sup>1</sup>,  
Charalampia (Xaroula) Kerasidou<sup>1</sup> , Federica Lucivero<sup>1</sup>,  
Michael Morrison<sup>1</sup>, Antoniya Georgieva<sup>1</sup>, Andrew J. Fletcher<sup>1</sup> ,  
Mirae Harford<sup>1</sup>, Casey Johnson<sup>1</sup>, Jane Kaye<sup>1</sup>, Robert Klassen<sup>1</sup>,  
Paul Leeson<sup>1</sup>, Fatima Ndanusa<sup>1</sup>, Christoffer Nellaker<sup>1</sup>,  
Alex Novak<sup>1</sup>, Daniel O'Connor<sup>1</sup>, John Powell<sup>1</sup>,  
Nabila Puspakesuma<sup>1</sup>, Sara Ratner<sup>1</sup>, Andrzej Rys<sup>2,3</sup>,  
Abram Schonfeldt<sup>1</sup>, Andrew A. S. Soltan<sup>1</sup>,  
Carolyn Ten Holter<sup>1</sup> and Patrick Thomson<sup>1</sup>

## Abstract

The recent advancements in artificial intelligence (AI), and data science more broadly, have led to a proliferation of new methods and tools, such as machine learning (ML), that are used in all kinds of scientific research, from biomedical research through to environmental and education research. Research ethics review bodies are increasingly required to review AI research protocols that cover these different fields of enquiry. Questions have been raised regarding the appropriateness of existing ethics governance principles, practices, and processes to deal with the ethical challenges that AI and data science are introducing to research. Universities and research institutions across the world are trying to understand

<sup>1</sup>University of Oxford, UK

<sup>2</sup>St Antony's College, University of Oxford, UK

<sup>3</sup>Innovative Health Initiative, European Union

## Corresponding author:

Angeliki Kerasidou, Ethox Centre and Wellcome Centre for Ethics and Humanities, Oxford Population Health, Nuffield Department of Population Health, University of Oxford, Big Data Institute | Li Ka Shing Centre for Health Information and Discovery, Oxford OX3 7LF, UK.

Email: angeliki.kerasidou@ethox.ox.ac.uk



how to translate and practically implement broad AI ethical principles into research ethics governance guidelines and processes. In this article, we report on an expert stakeholders' workshop organised at the University of Oxford as part of the process of reviewing its ethics governance for AI research. We describe the workshop and present the reflections and recommendations that emerged from it. The aim of the article is to share the approach taken by the University of Oxford CUREC in reviewing its ethics governance processes, and the insights gained with the broader research community, as a way of contributing to this scarce body of literature, facilitating further dialogue, promoting debate and collaboration on this important issue.

### Keywords

AI research, ethics review, ethics governance, research ethics, Research ethics committee (REC), Institutional Review Board (IRB)

## Introduction

Recent advancements in artificial intelligence (AI),<sup>1</sup> and data science more broadly, have led to a proliferation of new methods and tools, such as machine learning (ML), that are used in scientific research, from biomedical research (eithealth, 2024) to environmental (Zhu et al., 2023) and education research (Ersozlu et al., 2024). AI algorithms help researchers develop new hypotheses, collect, visualise and process vast amounts of data, design and conduct experiments, and evaluate results and outputs (Wang et al., 2023). Despite the potential value of this new technology, existing literature identifies a long list of ethical concerns pertaining to the use of AI in research, including issues relating to privacy and security, accountability and responsibility, energy usage, the use of data to train algorithms, developing trustworthy and human-centred AI, inclusion and bias (Borenstein and Howard, 2020; Goirand et al., 2021; Kaye et al., 2024; McGovern et al., 2022; Morley et al., 2020b; Murphy et al., 2021; Nuffield Council on Bioethics, 2018; Ratti et al., 2025).

At Universities, research ethics committees (RECs) and review boards<sup>2</sup> are tasked with reviewing research projects to ensure that research conforms to recognised ethical standards (Gelling, 1999). Ethical principles, such as respect for persons, beneficence, non-maleficence, justice, and social responsibility typically guide the review process (NIH, 1979). In their assessment, research ethics review bodies are expected to consider and balance risks to participants against potential benefits to future generations and society, and the interests of researchers (Gelling, 1999). This means that they are expected to ensure that new knowledge and insights are produced which are important and useful without unduly and unjustifiably burdening research participants and other stakeholders (e.g. local systems and structures). Being able to assess the ethical aspects of a research proposal depends, amongst other things, on the ability to comprehend methods used, identify and foresee associated risks, and assess social value of the research (Gelling, 1999). As such, one of the challenges RECs face is the need to stay abreast of new methodological and technological developments in research. Although sometimes researchers regard the experience of applying for ethics approval as a cumbersome formality (Chatzidamianos, 2013), research ethics review bodies remain the main ethics governance structure in research.

Nowadays, research ethics review bodies are increasingly required to review AI research protocols, yet, assessing the ethics of such protocols is neither easy nor straightforward (Ada Lovelace Institute, 2022; Ienca et al., 2018). Recent years have seen a proliferation of published AI governance principles, frameworks and guidelines (see Corrêa et al., 2023) from a variety of sources such

as national and supra-national bodies (e.g. House of Lords UK, Montreal Declaration, EU European Commission's High-Level Expert Group), professional associations (e.g. IEEE), international institutions (i.e. Future of Life Institute, OECD 2019, Alan Turing Institute, 2021) and private companies (i.e. Google, Microsoft), along with various review articles attempting to map, synthesise and identify common themes across these numerous documents (Corrêa et al., 2023; Jobin et al., 2019; Morley et al., 2020a; Schiff and Borenstein, 2019).

However, questions have been raised regarding the appropriateness of existing ethics governance principles, practices, and processes to deal with the ethical challenges that AI is introducing to research (Bouhouita-Guermech et al., 2023; Friesen et al., 2021; Hine, 2021). While these review articles have identified some common ethical themes across these documents (i.e. transparency, justice, non-maleficence, privacy (Jobin et al., 2019), it is not immediately obvious how established ethical principles, such as privacy and respect for persons, ought to apply in AI research, whether consent is still relevant and in what form, or how to evaluate the wider societal impact of AI tools and models to ensure social value (Andreotta et al., 2022; Friesen et al., 2021; Spellecy and Anderson, 2021). Furthermore, the nature of this type of research that brings together computer sciences, life sciences, social sciences, but also law, philosophy, and ethics, requires a multidisciplinary perspective in considering appropriate ways to approach the ethics governance of AI research.

While there is a proliferation of guidelines – some with general and some with a specific focus, such as medicine (Crossnohere et al., 2022), or education (Foltynek et al., 2023) – the majority of them focus on the ethics of the development and use of AI systems, rather than on research ethics (Ada Lovelace Institute, 2022; Wang et al., 2024). According to the Ada Lovelace Institute report, many AI researchers and RECs try to address this gap by drawing on legal guidance relating to data privacy. The introduction of the GDPR in 2018 provided a coherent framework for the protection of research data with guidance on collection, storage and use. However, as the Ada Lovelace Institute report notes, this runs the risk of ‘conflating questions of AI ethics into narrower issues of data governance’ (Ada Lovelace Institute, 2022: 56). The newly introduced EU AI Act (Regulation (EU) 2024/1689) is unique in establishing a comprehensive regulatory regime based on a risk-based approach, however the legal obligations it imposes are also directed mainly to AI developers and users. Although the need to develop tools that could be tested and used in the real world means that the directive of the AI Act will have an impact in research (Resseguier and Ufert, 2024), the Act does not directly cover scientific research. The UK has so far adopted a flexible, principle-based approach that favours voluntary AI guidelines over statutory regulation. Its approach, as set out in the AI Regulation White Paper, introduces a pro-innovation approach to AI regulation that relies on existing laws and regulators (UK DSIT, 2024).

Universities and research institutions across the world are trying to understand how to translate and practically implement broad AI ethical principles into research ethics governance guidelines and processes (Ferretti et al., 2020). A number of publications reflecting on the emerging and changing landscape of research have been reviewing the role and function of ethics committees in dealing with existing and new emerging challenges (Ferretti et al., 2021), and the adequacy of existing frameworks and structures to address these challenges (Morley et al., 2020a). Very few publications however, share first hand experiences of how institutions might go about reviewing and revising their ethics review processes and policies (Bernstein et al., 2021). As a result, there is little opportunity for research institutions to learn from one another.

In this article, we report on an expert stakeholders' workshop organised at the University of Oxford as part of the process of reviewing its ethics governance for AI research. The workshop was set up jointly by the Central University Research Ethics Committee (CUREC) and the Oxford Network for Sustainable and Trustworthy Artificial Intelligence in health and care (OxSTAI). All

the authors of this article participated at the workshop. Here, we describe the workshop and present the reflections and recommendations that emerged from it. The aim of the article is to share the approach taken by the University of Oxford CUREC in reviewing its ethics governance processes, and the insights gained with the broader research community, as a way of contributing to this scarce body of literature, facilitating further dialogue, promoting debate and collaboration on this important issue.

## **Oxford's Central University Research Ethics Committee (CUREC)**

Research at the University of Oxford is governed by the Central University Research Ethics Committee (CUREC). CUREC has overall responsibility for development of the University's Research Ethics Policy, and for the University's ethics review process. It does not review research applications, as this task is delegated to its interdivisional (IDRECS and OxTREC) and departmental (DRECS) subcommittees. In 2023, CUREC initiated a process of reviewing the University's research ethics governance practices to respond to the increasing volume of research that either used AI methodologies or aimed at developing AI tools and models. CUREC set out to assess whether its ethics policies and processes were appropriate for ethically assessing such projects. As part of this process, it reviewed a number of existing documents and guidelines, including the Ada Lovelace Institute reports 'Looking before we leap?' (Ada Lovelace Institute, 2022), and 'Understanding AI research ethics as a collective problem' (Waeiss, 2023), the Alan Turing guidelines on 'Understanding AI ethics and Safety' (Leslie, 2019), and their 'Research Ethics Policy' (The Alan Turing Institute), and the Committee On Publication Ethics report on authorship and AI (COPE Council, 2023). It also invited external speakers to present on issues pertaining to AI research to better understand the role of CUREC and its subcommittees with respect to facilitating ethical AI and data science research, and further explore the associated challenges.

The desk-based research undertaken, followed by engagement with experts, resulted into two recommendations. Firstly, given that AI is still a fast-evolving field, ongoing engagement with the existing and emerging literature AI ethics, including in the form of guidelines and recommendations from other higher education bodies and research institutions, is recommended to help formulate a path forward. Secondly, an engagement with the wider University of Oxford research community should take place with the aim to foster dialogue and obtain feedback from researchers working in this area as a way of informing CUREC's future ethics review processes with regards to the use of AI in research. Here, we report on CUREC's approach to achieve this second recommendation of engaging with and seeking input from the University research community.

## **OxSTAI-CUREC workshop: Dialogue and feedback**

In its effort to facilitate engagement with the University of Oxford wider research community, CUREC enlisted the assistance of the Oxford Network for Sustainable and Trustworthy Research in Health and Care (OxSTAI). OxSTAI is an interdisciplinary network of researchers from across the University and beyond. It includes computer scientists, clinical and biomedical researchers, engineers, statisticians, social scientists, legal, ethics, and philosophy scholars working on different aspects of AI within the context of health and social care. The main purpose of the network is to facilitate 'the exchange of ideas, and development of collaborations with the aim of identifying questions, researching and offering solutions to pertinent issues that relate to the development and deployment of AI in health and care' (OxSTAI, n.d).

OxSTAI organised and ran a cross-disciplinary workshop on the ethics governance of AI research bringing together researchers and members from the CUREC committee and subcommittees to help shape the future ethics governance of AI research at the University of Oxford. The aim of this workshop was to draw from the existing expertise and experience of the university's scientists working in and/or with AI, and who might have already applied for ethics approval for AI research, from the experience of ethics committee members in assessing AI research ethics applications, and also from the theoretical and empirical expertise of Oxford academics considering the ethical and societal implications of AI research. The workshop results were reported back to CUREC in the form of an internal report on insights gained and recommendations proposed.

### *The workshop*

A half-day workshop took place in March 2024, bringing together biomedical and clinical researchers, social scientists, computer scientists, biomedical scientists, statisticians, engineers, philosophers, research governance and legal scholars from the University of Oxford working on AI, as well as representatives from CUREC and its sub-committees. The workshop participants were selected to represent different disciplines (e.g. biomedicine, anthropology, sociology, law, education), and also different kinds of engagement with AI, from researchers conducting foundational AI research to those who use AI methods/tools in their research or develop AI tools. Participants were informed that the workshop was organised in partnership with CUREC, and its overarching aim was to gain insights from researchers on the ground to help CUREC review its existing ethics governance structures and develop recommendations. The workshop was divided into two sections: presentations in the first half of the morning, and a focussed round-table discussion in the second half. The presentations had the role of setting the scene, creating a common starting point and language for the discussion that followed. The first presentation was delivered by the Ethics Lead for the Research Governance, Ethics & Assurance Team, who spoke about the current research governance structures, introduced the 'problem' and governance work undertaken so far. The presentation also highlighted the need for community engagement and a bottom-up approach to identifying problems and solutions in the ethics governance of AI research, both in research developing AI, and in using AI in research. The second presentation on the ethics of AI research was delivered by an applied ethics expert working in the field of AI ethics. It focussed on the main ethical issues that arise in AI research and articulated key ethical requirements of AI research including autonomy, privacy, transparency, fairness, and environmental impact.

The second half of the workshop was a focussed round-table discussion centred around two main questions:

1. *What works well and what does not (what is missed, overlooked, misunderstood) in the current ethics governance process and structure when AI research is concerned?*
2. *Can you suggest any solutions or ways forward?*

From the group of workshop participants, four REC members and four researchers, working with AI in clinical research, neuroscience and bioinformatics who had applied for research ethics approval were asked in advance to reflect on these two questions and offer their views based on their experience, as a way of starting the conversation. The researchers were selected to represent diverse AI research methods and foci, and the REC members to represent the different research ethics subcommittees at Oxford. After the presentations and invited reflections, the floor was

opened for a round-table discussion led by a facilitator. Two note-takers took detailed notes of the presentations and discussions.

## **Workshop reflections**

A number of themes emerged from the workshop discussion, presented as reflections below:

### *AI exceptionalism and responsibility gaps*

One key question that emerged was that of ‘AI exceptionalism’, that is, whether AI is a special case in research which requires its own new ethical and regulatory paradigm. During the discussion, the participants reflected that the issue of exceptionalism is often raised when new methods and technologies are developed (Shevchenko and Zhavoronkov, 2024). As the question of ‘AI exceptionalism’ was debated, there was general agreement that the main ethical principles guiding RECs, such as beneficence, non-maleficence, respect for autonomy, justice, and social value, remain central in research involving AI, while the processes developed and applied to govern research are still applicable. The reasoning was that while AI can improve research potential, it still demands the same protection of participants’ rights and the same responsibilities of researchers and institutions. However, the prospect of ‘responsibility gaps’ in the use of AI (Santoni de Sio and Mecacci, 2021), and particularly in the case of autonomous AI, has the potential for making this technology ethically disruptive and, therefore, exceptional. If opaque AI systems learn autonomously from data and from previous experience and provide outcomes in ways that researchers themselves cannot understand, generating so-called ‘black box’ problems, there is a question about responsibility for those outcomes or any decision based on them. However, some workshop participants suggested that uncertainty regarding responsibility is something for which researchers could reasonably be expected to be accountable for the moment they decide to use AI, a view also reflected in the literature (Di Nucci, 2020; Kiener, 2022; Sauer et al., 2017; Tigard, 2021), and it is, therefore, reasonable to assume that responsible research practices using AI fall under existing ethical standards for accountability. There was general agreement by the workshop participants that existing ethics governance structures including ethics review, assessment, and approval of research protocols by the relevant bodies, such as RECs provide a good basis to support research using and developing AI.

### *Gaps in existing governance structures*

While existing ethics governance structures were deemed appropriate, there was agreement that some caution and openness to ethical uncertainty with regard to the introduction of AI in research would be appropriate. There were calls for adaptations of existing processes to ensure that new challenges posed by AI are met. For example, it was suggested that greater adaptability and flexibility should be built into the current structures to anticipate new AI developments and unintended consequences. Specifically, concerns were raised that the current system might not be fast or responsive enough to accommodate changes and adaptations of AI tools and methods within research projects. Given the pace of research and development in AI, it was deemed that processes might need to be revised to allow for rapid review of research, and to facilitate rapid advice and decision-making. The establishment of a list of experts or specialist advisory panel comprising individuals with expertise in AI research from different disciplinary backgrounds to which CUREC subcommittees could call up for advice and guidance, could facilitate the timely and appropriate review of protocols. It was also suggested that reconfiguring the ethics review as a dynamic

process, for example in the form of periodic reviews, could help capture and quickly respond to changes in research projects (e.g. in situations where the best AI technique to use in a study might change between the REC approval and the commence of the study, given that AI techniques develop quickly).

### *Risk of overregulation*

Some workshop participants also expressed the view that because of its novelty, AI might lead RECs to be disproportionately strict in their review of research applications. It was noted that the mere presence of AI tools or methodologies in a research proposal should not determine the level of scrutiny applied. For example, currently, RECs categorise research protocols as low- or high-risk based on certain criteria, such as involvement of vulnerable populations, use of deception, risk of harm to participants, etc. Low-risk research is reviewed by a smaller committee often consisting of the REC secretariat and chair (often called expedited review process). High-risk research is reviewed by the full committee, which might also take longer and invites more scrutiny. Workshop participants suggested that while there might be projects where AI is central (e.g. interactive assessment and development of a new AI-driven diagnostic tool), and could be seen as high-risk projects, there will be others where AI only plays a minor part (e.g. use of AI model to help translate research findings into artistic representations), in which case a low-risk review process might be more appropriate. They cautioned against the potential risk of overprotectiveness and disproportionate vigilance that could result in all AI research protocols being classified as high-risk research, inviting in some cases, unnecessary scrutiny and delays.

### *Adversarial learning*

The issue of adversarial learning approaches and how the risk of developing ‘bad’ AI in order to train ‘good’ AI was also discussed (Zhang et al., 2018). Adversarial learning is a method of improving the robustness of machine learning (ML) systems by exploiting their vulnerabilities through malicious input. For example, a recent ethical analysis of adversarial ML and ‘data poisoning’ suggested that such methods might be justified in the case of developing facial recognition tools, but not in the case of medical AI when considering issues regarding privacy and purpose limitation (Adomaitis and Oak, 2023). During the workshop, the case of adversarial learning approaches was presented as further evidence that underlines the need to be able to appropriately evaluate risk depending on the type of AI and the role it plays in a research project, as opposed to applying a blanket risk-related characterisation as high- or indeed low-risk to all AI research.

### *Participant engagement and communication*

The importance of clear and appropriate communication to research participants of how AI tools and techniques are developed and used in research, as well as the risks, harms, and potential benefits was also highlighted. It was noted, for example, that given the speed with which AI tools are developed and released, it is possible that those participating in research are also able to enjoy its benefits, rather than considering research benefits as primarily targeted to future populations (Kerasidou and Binik, 2022). It was suggested that public and patient involvement and engagement (PPIE), as well as research co-production methods, could facilitate the development of appropriate language to communicate complex notions and technical concepts hence increasing awareness and potentially facilitating social acceptability and public trust in AI research.

### *Indirect impacts*

Finally, broader points were raised about the role of RECs in general, and the extent to which they should be assessing research's broader implications, such as environmental but also social, cultural and political impact. Specifically, in terms of its environmental impact, a special note was made that, while not unique to AI, its heavy energy demand is a matter of concern. It was argued that there needs to be a demonstrated (quantified) value to the use of energy proportional to the goals of any given AI research project (Budenny et al., 2022). However, it was acknowledged that this is not easily evaluated, nor is energy input always known. Also, it was suggested that it might be easier to make a resource use case for some domains (e.g. cancer research) than others (e.g. art or education). Following this, questions were raised regarding interdisciplinary fairness, and when, how and by whom questions regarding energy and resource use would be justified across potentially vastly different academic fields and projects. It was acknowledged that there is currently no guidance of how this type of judgements ought to be made, as well as at what stage in the research review process such evaluations ought to take place, for example, whether this should arise at the research ethics approval stage or at the funding stage.

### **Workshop recommendations**

The notes taken during the workshop were reviewed by CUREC's Ethics and Research Governance lead, and one of the workshop co-organisers, who also received input from the other workshop co-organisers representing OxSTAI. The review of the insights gained from the workshop resulted in a number of recommendations listed below. These recommendations were shared with the rest of the workshop participants for refinement and approval, and afterwards were put forward for CUREC's consideration.

1. Given that the level of ethics governance scrutiny applied to research projects depends on evaluation of risk, research needs to take place to delineate types of AI used and developed in research. These different types then should be analysed with the view of developing a set of criteria of what constitutes high- or low-risk AI research. A periodic review of types of AI and risk-criteria should also take place to account for advancements and changes in the field.
2. The establishment of a Specialist Advisory Panel comprising experts in AI research from different disciplinary backgrounds (e.g. computer science, mathematics, social sciences, law, philosophy, and ethics). The panel could provide specialist support to CUREC and its subcommittees in the development of risk levels and subsequent procedures, and advice to subcommittees and research ethics staff.
3. Develop training for REC members on topics related to AI use and development in research.
4. Engage with researchers and existing Patient and Public Involvement and Engagement (PPIE) groups across the University to develop clear explanations of AI for participants.
5. Maintain a register of studies developing or using AI to track range of research review outcomes, areas of concentration, and issues emerging. This would be held by research ethics staff and reviewed with Specialist Advisory Panel members.
6. Trial a 'rolling review' of projects predicted to change significantly during their lifecycle.
7. Evaluate the process of 'rolling review'. This could include consideration of whether a rolling review process can be accommodated through existing processes of amendment and annual report, or whether other processes would be required.

CUREC has considered these recommendations and is in the process of implementing a number of them, including establishing a Specialist Advisory Panel, and setting up a register of AI studies and monitoring AI research applications to ascertain if a method of ‘rolling review’ could be applicable and beneficial. The committee is also considering steps to develop and deliver training on the ethics of AI for REC members. Although OxSTAI is not tasked with developing training for CUREC, it is possible that members of the networks as well as workshop participants might be amongst those who could form the Specialist Advisory Panel (see recommendation #2), or assist CUREC with developing appropriate AI research training for its members (see recommendation #3).

Many of the recommendations resulting from the workshop are also reflected in the literature. For example, the need for relevant training and expertise, and the suggestion of ongoing engagement with the ethics review bodies, or as we call it, a rolling review process, can be found in relevant articles and reports (Ada Lovelace Institute, 2022; Ferretti et al., 2021; Hine, 2021; Ienca et al., 2018; Knight et al., 2025). It is worth noting, however, that whilst some points raised during the workshop were able to be translated into recommendations and action points for the REC, others, such as the broader issues of how to incorporate assessment of the environmental impact of research in an ethics governance structure, or how to assess and evaluate the broader societal implications of AI research, particularly of generative AI (GenAI), were less easily actioned. It was acknowledged, however, that ethics governance of research also takes place outside the function of a research ethics committee that sits within a research or academic institution. Ethics governance committees and data access committees positioned within funding bodies, data banks, and other bodies which evaluate and approve new technologies, could be more appropriate places where some of these broader issues are captured, assessed and evaluated. Furthermore, these bodies and groups, rather than operating independently, could come together to create a more comprehensive research ethics governance structure that is able to follow the research from ideation to end-product, thus ensuring the ethical governance of AI research.

Finally, it was recognised that other institutions around the world are going through a similar exercise with respect to the ethics governance of AI research. The work presented here could form part of a more concentrated national and even global effort in this area. For example, and in recognition of the cross-national and collaborative nature of research that often requires approval from different ethics committees, further consideration could be given to identifying aspects that could be standardised globally and others that would require local governance. It was suggested that CUREC could initiate further conversations with global partners (e.g. academic institutions, research institutions, national and supranational research governance bodies) working on enhancing appropriate ethical oversight of AI research.

This is the first time that the University of Oxford’s central research ethics committee has incorporated a bottom-up approach in the review and revision of its ethics governance policies. Previously, any revisions and reviews had been conducted internally and with the engagement only of selected experts. This time, however, it was deemed appropriate and necessary, particularly given the novelty of the technology and the fast pace by which AI research is developing, to invite views, experiences, and recommendations from the University’s research community. Furthermore, as ethics committees are often accused of introducing unnecessary red-tape that hampers important and valuable research, research ethics governance needs to strike a fine balance between facilitating research and protecting research participants and data subjects. Developing an ethics governance policy with the input from researchers on the ground could help achieve this difficult balance.


## Conclusion

Developing an ethics governance framework for research involving AI requires balancing considerations about the uniqueness of AI with considerations about its similarities with other tools and methodologies in research. While AI does not, at this stage, seem to require its own ethics governance framework, some of its distinguishing aspects require a more flexible, dynamic, and responsive governance structure than the one typically adopted in the case of other research tools and methodologies. The OxSTAI-CUREC workshop was intended as a first step to identifying the relevant questions that RECs need to address in assessing research involving AI, particularly ML and more recently GenAI, and what work needs to be done, and by whom, to supplement existing frameworks with AI-relevant considerations. We hope that this report article prompts further discussions in that direction, and also encourages greater institutional openness, at the organisational, national and global levels, in sharing experience and expertise, and examples of best practice. Finally, we hope that it serves as a useful example of stakeholder-engagement process in developing ethics governance structures and policies.

## Acknowledgements

We would like to thank Caroline Green, Fergus Gleeson and Harshal Thaker for their participation at the OxSTAI-CUREC workshop, and for sharing their invaluable views and insights.

## ORCID iDs

Angeliki Kerasidou  <https://orcid.org/0000-0001-9344-3297>

Charalampia (Xaroula) Kerasidou  <https://orcid.org/0000-0002-9794-8492>

Andrew J. Fletcher  <https://orcid.org/0000-0003-2182-1089>

## Ethical Considerations

Not applicable.

## Consent to participate

Not applicable.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: AK was supported by the Wellcome Trust, [203132/Z/16/Z].

AK and XK have been supported by an NDPH Senior Fellowship.

AG was supported by CAVAA European Commission, EIC 101071178.

XK was supported by the Wellcome Trust, [203132/Z/16/Z].

MH was funded by the National Institute for Health and Care Research (NIHR) Oxford Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

MM was funded by the European Union under the Horizon Europe grant REALM 101095435. The views and opinions expressed are solely those of the author(s) and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

JP receives funding from the National Institute for Health and Care Research (NIHR) Applied Research Collaboration Oxford and Thames Valley at Oxford Health NHS Foundation Trust.

CTH is supported by the Engineering and Physical Sciences Research Council [Responsible AI IA091 Grant Ref: EP/Y009800/1].

## Declaration of Conflicting Interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Dr Alex Novak - Declaration of Conflicts of Interest.

### Personal Appointments

- Standing Member of National Institute of Health and Care Excellence Diagnostics Advisory Committee.
- Member of Cochrane Acute and Critical Care Editorial Board.
- Member of National Institute of Health Research (NIHR) Research for Patient Benefit (RfPB) Grant Panel.
- Member of Royal College of Emergency Medicine Research and Publications Committee.

### Grants Received

- Grants received from:
  - National Institute of Health Research (NIHR)
  - Innovate UK
  - Small Business Research Institute
  - NHSX (Qure.AI)
  - GE Healthcare Ltd
  - Perspectum Diagnostics Ltd
  - Lunit
  - Radiobotics
  - Reporting and Imaging Quality Control
  - Seroxo
  - Abbott

### Honoraria

- Consultancy work previously undertaken for:
  - Abbott (2025)
  - GE (2021)

## Data Availability Statement

Not applicable.

## Notes

1. We use AI as an umbrella term to refer to computational techniques that enable computers and machines to simulate aspects of human intelligence such as learning, comprehension and problem solving. Machine Learning (ML) is a particular, and v prevalent, type of AI which involves computers ‘learning’ without specific instructions, doing so by using algorithms to analyse and draw inferences from data. Generative AI (GenAI) is another prevalent type of AI that is widely used in research and development at the moment. At the time of the workshop GenAI had not as ubiquitously used as it is today.
2. There is a variety of terms used globally to describe such bodies whose aim is to consider the ethical implications of research. In the UK, they are usually referred to as Research Ethics Committees (RECs) while in the US they are more commonly called Institutional Review Boards (IRBs). Other terms include, ethical review boards (ERB), independent ethics committees (IEC), or research ethics boards (REB). In this article, we will use the REC term throughout. See for an introduction and short history of their origins.

## References

- Ada Lovelace Institute (2022) Looking before we leap: Ethical review processes for AI and data science research. Available at: <https://www.adalovelaceinstitute.org/report/looking-before-we-leap/> (accessed 13 December 2022).

- Adomaitis L and Oak R (2023) Ethics of Adversarial Machine Learning and data poisoning. *Digital Society* 2(1): 8.
- Andreotta AJ, Kirkham N and Rizzi M (2022) AI, big data, and the future of consent. *AI & Society* 37(4): 1715–1728.
- Bernstein MS, Levi M, Magnus D, et al. (2021) Ethics and society review: Ethics reflection as a precondition to research funding. *Proceedings of the National Academy of Sciences* 118(52): e2117261118.
- Borenstein J and Howard A (2020) Emerging challenges in AI and the need for AI ethics education. *AI and Ethics* 1(1): 61–65.
- Bouhouita-Guermech S, Gogognon P and Bélisle-Pipon JC (2023) Specific challenges posed by artificial intelligence in research ethics. *Frontiers in artificial intelligence* 6: 1149082.
- Budenny SA, Lazarev VD, Zakharenko NN, et al. (2022) Eco2AI: Carbon emissions tracking of machine learning models as the first step towards sustainable AI. *Doklady Mathematics* 106(S1): S118–S128.
- Chatzidamianos G (2013) Ethics stop dodgy experiments but there's too much red tape. *The Conversation*, 8 November.
- COPE Council (2023) COPE position - Authorship and AI. DOI: 10.24318/cCVRZBms.
- Corrêa NK, Galvão C, Santos JW, et al. (2023) Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns* 4(10): 100857.
- Crossnohere NL, Elsaid M, Paskett J, et al. (2022) Guidelines for Artificial Intelligence in Medicine: Literature Review and Content Analysis of Frameworks. *Journal of Medical Internet Research* 24(8): e36823.
- Di Nucci E (2020) *The Control Paradox: From AI to Populism*. Rowman & Littlefield International.
- eihealth (2024) Machine learning in healthcare: Uses, benefits and pioneers in the field. Available at: <https://eihealth.eu/news-article/machine-learning-in-healthcare-uses-benefits-and-pioneers-in-the-field/> (accessed 15 March 2024).
- Ersozlu Z, Taheri S and Koch I (2024) A review of machine learning methods used for educational data. *Education and Information Technologies* 29(16): 22125–22145.
- Ferretti A, Ienca M, Hurst S, et al. (2020) Big Data, Biomedical Research, and Ethics Review: New Challenges for IRBs. *Ethics & Human Research* 42(5): 17–28.
- Ferretti A, Ienca M, Sheehan M, et al. (2021) Ethics review of big data research: What should stay and what should be reformed? *BMC Medical Ethics* 22(1): 51.
- Foltynek T, Bjelobaba S, Glendinning I, et al. (2023) ENAI recommendations on the ethical use of Artificial Intelligence in Education. *International Journal for Educational Integrity* 19(1): 12.
- Friesen P, Douglas-Jones R, Marks M, et al. (2021) Governing AI-Driven Health Research: Are IRBs up to the task? *Ethics & Human Research* 43(2): 35–42.
- Gelling L (1999) Role of the research ethics committee. *Nurse Education Today* 19(7): 564–569.
- Goirand M, Austin E and Clay-Williams R (2021) Implementing ethics in Healthcare AI-based applications: A scoping review. *Science and Engineering Ethics* 27(5): 61.
- Hine C (2021) Evaluating the prospects for university-based ethical governance in artificial intelligence and data-driven innovation. *Research Ethics* 17(4): 464–479.
- Ienca M, Ferretti A, Hurst S, et al. (2018) Considerations for ethics review of big data health research: A scoping review. *PLoS One* 13(10): e0204937.
- Jobin A, Ienca M and Vayena E (2019) The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1(9): 389–399.
- Kaye J, Shah N, Kogetsu A, et al. (2024) Moving beyond technical issues to stakeholder involvement: Key areas for consideration in the development of human-centred and trusted AI in Healthcare. *Asian Bioethics Review* 16(3): 501–511.
- Kerasidou A and Binik A (2022) *Research Ethics*. Routledge Encyclopedia.
- Kiener M (2022) Can we Bridge AI's responsibility gap at will? *Ethical Theory and Moral Practice* 25(4): 575–593.
- Knight S, Shibani A and Vincent N (2025) Ethical AI governance: Mapping a research ecosystem. *AI Ethics* 5: 841–862.

- Leslie D (2019) Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. Reportno. Report Number[, Date. Place Published; Institution].
- McGovern A, Ebert-Uphoff I, Gagne DJ, et al. (2022) Why we need to focus on developing ethical, responsible, and trustworthy artificial intelligence approaches for environmental science. *Environmental Data Science* 1: e6.
- Morley J, Floridi L, Kinsey L, et al. (2020a) From what to how: An Initial Review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics* 26(4): 2141–2168.
- Morley J, Machado CCV, Burr C, et al. (2020b) The ethics of AI in health care: A mapping review. *Social Science & Medicine* 260: 113172.
- Murphy K, Di Ruggiero E, Upshur R, et al. (2021) Artificial intelligence for good health: A scoping review of the ethics literature. *BMC Medical Ethics* 22(1): 14.
- National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979) *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. U.S. Department of Health and Human Services. Available at: <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html>
- Nuffield Council on Bioethics (2018) Artificial Intelligence (AI) in healthcare and research. Available at: <https://www.nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research-1.pdf> (accessed 5 March).
- OECD (2019) *Recommendation of the Council on Artificial Intelligence*. OECD/LEGAL/0449. OECD Legal Instruments. Available at: [https://www.oecd.org/content/dam/oecd/en/publications/reports/2019/06/artificial-intelligence-in-society\\_c0054fa1/eedfee77-en.pdf](https://www.oecd.org/content/dam/oecd/en/publications/reports/2019/06/artificial-intelligence-in-society_c0054fa1/eedfee77-en.pdf)
- OxSTAI (n.d). *Oxford Network for Sustainable and Trustworthy AI in Health and Care (OxSTAI)*. University of Oxford. Available at: <https://www.ethox.ox.ac.uk/Our-research/research-projects/network-ai-health-and-care-oxstai>
- Ratti E, Morrison M and Jakab G (2025) *Ethical and Social Considerations of Applying Artificial Intelligence in Healthcare; A Two-Pronged Scoping Review*. BMC Medical Ethics.
- Resseguier A and Ufert F (2024) AI research ethics is in its infancy: The EU’s AI Act can make it a grown-up. *Research Ethics* 20(2): 143–155.
- Santoni de Sio F and Mecacci G (2021) Four responsibility gaps with Artificial Intelligence: Why they matter and how to address them. *Philosophy & Technology* 34(4): 1057–1084.
- Sauer HC, Köhler S, Roughley N, et al. (2017) Technologically blurred accountability? Technology, responsibility gaps and the robustness of our everyday conceptual scheme. In: Ulbert C, Finkenbusch P and Sondermann E (eds) *Moral Agency and the Politics of Responsibility*. Routledge, pp. 36–50.
- Schiff D and Borenstein J (2019) How should clinicians communicate with patients about the roles of artificially intelligent team members? *AMA Journal of Ethics* 21(2): E138–E145.
- Shevchenko S and Zhavoronkov A (2024) The role of exceptionalism in the evolution of Bioethical Regulation. *Cambridge Quarterly of Healthcare Ethics* 33(2): 185–197.
- Spellecy R and Anderson EE (2021) Respect for persons and Artificial Intelligence (AI) in the age of Big Data. In: Jotterand F and Ienca M (eds) *Artificial Intelligence in Brain and Mental Health: Philosophical, Ethical & Policy Issues*. Springer International Publishing, pp.73–85.
- The Alan Turing Institute (2021) Turing Research Ethics (TReX) Policy. Available at: <https://www.turing.ac.uk/turing-research-ethics-policy-trex-0> (accessed 5 March).
- Tigard DW (2021) There is no techno-responsibility gap. *Philosophy & Technology* 34(3): 589–607.
- UK DSIT (2024) Implementing the UK’s AI Regulatory Principles: Initial guidance for regulators. Available at: [https://assets.publishing.service.gov.uk/media/65c0b6bd63a23d0013c821a0/implementing\\_the\\_uk\\_ai\\_regulatory\\_principles\\_guidance\\_for\\_regulators.pdf](https://assets.publishing.service.gov.uk/media/65c0b6bd63a23d0013c821a0/implementing_the_uk_ai_regulatory_principles_guidance_for_regulators.pdf) (accessed 29 October 2025).
- Waeiss Q (2023) Understanding AI research ethics as a collective problem. Available at: <https://www.adal-ovelaceinstitute.org/blog/ai-research-ethics-collective-problem/> (accessed 13 January 2023).
- Wang H, Fu T, Du Y, et al. (2023) Scientific discovery in the age of artificial intelligence. *Nature* 620(7972): 47–60.

- Wang J, Huo Y, Mahe J, et al. (2024) Developing an ethical regulatory framework for Artificial Intelligence: Integrating systematic review, thematic analysis, and multidisciplinary theories. *IEEE Access* 12: 179383–179395.
- Zhang BH, Lemoine B and Mitchell M (2018) Mitigating unwanted biases with adversarial learning. In: *Proceedings of the 2018 AAAI/ACM conference on AI, ethics, and society*, New Orleans, LA, USA, pp.335–340. Association for Computing Machinery.
- Zhu J-J, Yang M and Ren ZJ (2023) Machine learning in environmental research: Common pitfalls and best practices. *Environmental Science & Technology* 57(46): 17671–17689.