



# Ethical AI governance: mapping a research ecosystem

Simon Knight<sup>1,2</sup> · Antonette Shibani<sup>1</sup> · Nicole Vincent<sup>1</sup>

Received: 14 September 2023 / Accepted: 22 December 2023 / Published online: 14 February 2024  
© The Author(s) 2024

## Abstract

How do we assess the positive and negative impacts of research about- or research that employs artificial intelligence (AI), and how adequate are existing research governance frameworks for these ends? That concern has seen significant recent attention, with various calls for change, and a plethora of emerging guideline documents across sectors. However, it is not clear what kinds of issues are expressed in research ethics with or on AI at present, nor how resources are drawn on in this process to support the navigation of ethical issues. Research Ethics Committees (RECs) have a well-established history in ethics governance, but there have been concerns about their capacity to adequately govern AI research. However, no study to date has examined the ways that AI-related projects engage with the ethics ecosystem, or its adequacy for this context. This paper analysed a single institution's ethics applications for research related to AI, applying a socio-material lens to their analysis. Our novel methodology provides an approach to understanding ethics ecosystems across institutions. Our results suggest that existing REC models can effectively support consideration of ethical issues in AI research, we thus propose that any new materials should be embedded in this existing well-established ecosystem.

**Keywords** Research ethics · Research governance · Artificial intelligence · Professional learning · Ethical AI · Sociomaterial

## 1 Introduction

How do we assess the positive and negative impacts of research about- or research that employs artificial intelligence (AI)<sup>1</sup>? This is a pressing question, with ambiguities around the role of researchers, governance bodies, and those who will use or/and be impacted by technologies, across both academic and industry contexts. While significant work

has been undertaken to describe ethical challenges of AI, and develop guidance and principles to guide practice, there remains concern regarding the governance of AI research, the gap between principles and practice, and the participation of stakeholders in deciding how AI may be used about, with, for, and on them.

This paper engages with this challenge through the analysis of existing research governance material, by investigating

✉ Simon Knight  
simon.knight@uts.edu.au

<sup>1</sup> University of Technology Sydney, TD School, PO Box 123, Broadway, NSW 2007, Australia

<sup>2</sup> Centre for Research on Education in a Digital Society (CREDS) School, University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia

<sup>1</sup> In this work we will use the term 'AI' in a broad sense to refer to techniques that include both symbolic reasoning (e.g. expert systems) and statistical reasoning (i.e. the wide range of techniques often collectively referred to as "machine learning" (ML), as well as hybrid techniques that employ both an ensemble of statistical and symbolic reasoning techniques, targeting tasks that would otherwise require human intelligence, following the early Dartmouth workshop definition of AI [57].

the resources that researchers and research ethics committees (RECs) draw upon in articulating and navigating ethical issues arising out of AI-related research. Resources, such as formalised ethical principles and articulated processes, inscribe knowledge. As such, they act as reflections of knowledge and practices, while also shaping that practice through their conceptual and normative (or regulatory) impact on actors. Beyond this, knowing what resources researchers and RECs actually draw upon can provide important insights into existing knowledge and practices, as well as shape future practice. Such knowledge can also help ascertain whether the hype of AI ethics deserves the attention it's getting, by illuminating whether—and, if so, how—the numerous published AI Ethics principles are actually used, which in turn will help to appraise the utility of such publications. In doing this we aim to contribute to understanding the ethical issues that AI-related research gives rise to and how learning about these might be (or could be) taking place. Through our socio-material analysis of materials relating to AI from a single institution's ethics committee process, we address this concern, exploring how these materials provide a lens onto and reflection of the ethical concerns of AI research.

## 2 Literature review

### 2.1 Ethical principles and practices

To foster ethical action in the developing areas around use of AI and data, a wide range of guidance and sets of principles have been developed. A recent review identified 84 sets of AI ethics guidelines globally with 11 themes among them [45], while another review of 36 principles documents identified that consensus could be seen across eight common themes [32]. A third review of only research studies regarding ethical principles identified 27 such studies with 22 principles [49]. Finally, a fourth review of public, private, and non-governmental organization (NGO) documents providing AI guidance identified 112 such documents [75]. Significantly, this last review identified significant differences in the focus of documents produced by different stakeholders, and their production, with NGOs and public organisations covering more topics, and being more likely to have engaged participatory approaches in their development. Nevertheless, across these reviews, identified principles overlap significantly with the classic Belmont principles.

Moreover, there have been various calls to move from a focus on developing AI ethics principles, to instantiating them in practice and organisational structures to support *practical ethics* [50, 60, 72, 89, 91]. These calls emphasise the significance of micro-ethics or 'ethics-in-action', and a shift from procedural to situational ethics [33, 38, 43, 51, 64].

This shift, particularly as expressed by [38] reflects both that when we apply procedural ethics we are engaged in practices, and that this process of translation is not mechanical and requires interpretation. This is a concern of recent AI work, reflecting that ethics is fundamentally imbued with action, and ongoing interactions in design processes, in ways hard to capture in procedural ethics. In addition, recent calls have highlighted the importance of analysis of ethical issues of technologies in terms of both immediate or direct impacts (hard impacts), and long-term or indirect impacts (soft impacts), that may affect people's lives [80].

### 2.2 The role of research ethics committees

Beyond the myriad of AI ethics principles, the role of governance structures in oversight of novel applications of AI has received attention, beyond the governance pages of companies and universities, in popular media coverage (e.g. [13, 42, 52, 53]). These structures—in the form of Research Ethics Committees (RECs) and Institutional Review Boards (IRBs) — play a crucial role in university research internationally, with mounting pressure to create similar bodies in companies, and a recognition of the challenges such bodies face. In this paper, we will use RECs as a general term that includes medical MRECs, human HRECs, and IRBs, except where explicitly stated otherwise.

RECs are typically comprised of multi-disciplinary research experts, alongside non-research members—in some systems, including lay people—who oversee and review research that involves human participants.<sup>2</sup> Researchers who wish to undertake such work typically submit an application that explains what the research will involve, and how it will address key ethical principles including the Belmont principles of respect for persons (or autonomy), beneficence, and justice [66]. The role of RECs is to assure these principles are instantiated in research that is approved, and to provide feedback to researchers [44]. RECs do this by assessing materials submitted by researchers. However, due to the formalised process of this work, tensions have emerged regarding the bureaucratisation of research and control by RECs with perceptions that RECs are particularly suited to work in a bio-medical model [5], although some of these concerns may relate to local—changeable—practices rather than underlying theoretical issues [40].

#### 2.2.1 International context of RECs

Significantly, the requirements and remit of RECs vary internationally. While research ethics systems share much

<sup>2</sup> We limit discussion here to *human* research ethics here, although similar systems exist in the context of the ethics of research involving animals.

common history, a number of publications have investigated similarities and variations across international ethics standards and committees [37, 44, 88] and common emerging themes—including that of data and AI [88].

However, there are also more or less nuanced differences in expression and execution of REC processes internationally. A particularly salient example, given that US experiences are often universalised, is found in the United States IRB guidance, which explicitly directs members as follows:

*§ 46.111 Criteria for IRB approval of research: (2) Risks to subjects are reasonable in relation to anticipated benefits, if any, to subjects, and the importance of the knowledge that may reasonably be expected to result. In evaluating risks and benefits, the IRB should consider only those risks and benefits that may result from the research (as distinguished from risks and benefits of therapies subjects would receive even if not participating in the research). The IRB should not consider possible long-range effects of applying knowledge gained in the research (e.g. the possible effects of the research on public policy) as among those research risks that fall within the purview of its responsibility. [82].*

In considering this quote as exemplifying the direction given in the wider document, and in comparison to international documents, we can note that IRBs receive unclear advice regarding their role in assessing merit and integrity (a key principle in the Australian model [70]), and that they must navigate the directions to (1) balance risks and benefits, with those to (2) “not consider” long-range effects [9]. Moreover, common across RECs is that they provide ‘point-in-time’ governance, but monitor and evaluate largely via periodic self-report of applicants or a complaints-based system for participants [20]. This lack of long-view consideration may create an ethical debt, through which technologies are developed without adequate consideration of long-term impacts [68].

This specific statement has thus been highlighted as a key feature in considering the adequacy of the US IRB model for AI research [25]. Importantly, we should be cautious in universalising models of ethics given cultural and contextual variation in values and practices, and in capacity and procedures for ethics governance such that universal expectations of REC review may exclude researchers from countries where no such review is available [55, 90]. Further, complications stem from disciplinary differences, e.g. [18], and conflicts between REC guidelines and ethical norms of communities with whom research is conducted [21]. Crucially, differences in disciplinary, cultural, and other contextual features must be recognised. In places, this recognition and negotiation is a crucial part of ethical practice, because we should expect values to be contextual. In other contexts,

there may be variation presently that rests on a lack of clear standards or articulation of possible norms to which we might work; expressing this is important.

## 2.2.2 AI and data as challenges to research ethics committees

Even outside the context of AI, concerns have been raised about what we know regarding researchers’ level of understanding of research ethics [10], and correspondingly, regarding expectations around interaction between committee members and disciplinary experts and their respective expertise [26]. However, as Hickey et al., [40] highlight, although there are various criticisms of ethics review boards, disentangling issues of practice in particular institutional committees, from more fundamental concerns with their underpinning principles, is challenging. They suggest that the criticisms of RECs can be addressed through fostering positive learning-oriented ethical review processes, adopting practices of open communication, outreach to research communities, and engagement with disciplinary expertise [40]. These suggestions are echoed by Brown et al.’s [17] specific analysis of education researchers’ views in the UK, finding that although there were concerns regarding understanding of the specific methods and issues in educational research, many respondents had positive interactions with their RECs [17]. Nevertheless, a lingering concern is that RECs act as ‘moral bureaucracies’ via a managerial audit approach to ethics, that is likely to incentivise ‘safe’ practice, and reduce productive rich dialogue regarding ethics, particularly in the context of technology [61].

Indeed, this sentiment is echoed in a discussion of the strengths and limitations of RECs’ specifically focused on AI and data research [29–31]. Here, two key concerns regard the nature of research conducted entirely outside university contexts (in which REC systems are mandatory and established) or in collaboration with such industry partners; and where the research involves secondary uses of publicly available datasets [35].<sup>3</sup> Given these concerns, and particularly the potential for unanticipated and unintended consequences, there have been calls to adjust REC’s understanding of research and data [28]; privacy protection in the context of re-identification (e.g. [15]; and moves in some UK RECs to give greater attention to data and AI [39]. An alternative approach has also been piloted in which researchers wishing to access participating funders’ grants undertake a separate review by an ‘Ethics and Society Review (ESR) board’ [11].

<sup>3</sup> The nature of ‘human subjects’ and personal data is contested in the context of big data research, which often draws on publicly available datasets [58]. For this reason, alternatives to consent have been explored (e.g. [69]. Indeed, disagreements and perceptions of varying practices across researchers, and across academic-industry located research, exist across the Belmont principles with respect to use of online data in computer science [87].

This ESR is constituted of interdisciplinary researchers who review author statements regarding possible impacts and risk mitigation, and provide feedback, with positive initial evaluation of the program. Relatedly, specific guidance has been produced to support industry organisations in establishing ethics bodies [74].

While such concerns have been highlighted across a number of studies and media articles, in a seminal study that involved interviewing REC members, some of the interviewed participants rejected the need for yet further guidelines, instead calling for “implementable procedures to assess big data projects effectively” [31], 136). As the authors highlight, differences in requests for specific guidance may lead to differing outcomes from different committees, with potential for negative impact on “researchers’ trust in the oversight system, data sharing practices, and research collaborations” [31], 138). Crucially, a lack of expertise and experience in assessing big data and AI-related projects was also explicitly recognised by the interviewed REC members [31], findings supported by further European and US research [81, 86].<sup>4</sup>

## 2.3 Learning for AI research ethics

Despite apparent gaps in REC systems, as the Future of Privacy Forum report ‘Designing an Artificial Intelligence Research Review Committee’ [46] sets out, in developing models to adequately address AI research, we can learn from the significant work undertaken in human and animal research, and biosafety committees. In similar work proposing developments in RECs, a 2021 collaboration between the UK’s Ada Lovelace Institute, University of Exeter, and Alan Turing Institute, investigated ‘Supporting AI research ethics committees: Exploring solutions to the unique ethical risks that are emerging in association with data science and AI research’ [1], and the associated ‘Looking before we leap’ project [84]. Their report [2] highlighted six challenges for research ethics committees:

1. lack of resources and training
2. mismatch in principles designed for researcher-subject relationships applied to researcher-data subject relationships

3. lack of established norms regarding principles for use specifically in AI and data research
4. cross-institutional (and sector) research which can lead to research being assessed by multiple committees
5. challenges of assessing unexpected impacts
6. transparency with respect to corporate research ethics groups or involvement of corporate entities in research activities

These challenges, and the context described in the preceding sections, led these researchers to recommend some key foci for RECs in considering AI research (which we revisit in Conclusions), synthesised to indicate their strong parallels in Table 1.

### 2.3.1 The need for learning in RECs

These sets of concerns and recommendations are intertwined, each contributing- and being contributed to by others in the list. A lynchpin of these recommendations is learning. This focus involves understanding how RECs, researchers, and stakeholders learn about AI, its impacts, and the systems into which it is deployed and the ethical concerns of those systems. Research is required to understand the processes of this learning, how learning about AI, ethical development and thinking, and systems thinking come together. RECs should have continuous training for staff regarding ethical review processes and their importance, with ongoing development. This applies in university contexts, but notably: “Many corporate RECs we spoke with also place an emphasis on continued skills and training, including providing basic ‘ethical training’ for staff of all levels.” ([2], 37).

In a powerful move highlighting the significance of learning through editorial policy, the Journal of Empirical Research on Human Ethics includes in its manuscript template an Educational Implications section, intended to discuss the ‘key concepts’ from the article to support teaching to different stakeholders, including research and REC communities, as well as students, and external stakeholders such as participants and the general public [47]. In Ferretti et al.’s analysis—published in that journal—this ‘Educational Implications’ section notes the significance of: “knowledge exchange and a more productive engagement among the various factors involved in big data research. These include and are not limited to RECs, researchers, research institutions, private enterprises, citizen science groups, and the public” ([31], 139), highlighting that this responsibility involves developing skills around both the technology (AI), and ethical processes and values. As they also highlight, the range of actors for whom there are implications for learning extends to “informing society about issues related to big data and the use of AI in research. Starting with this democratic engagement, the general public can clarify their expectations

<sup>4</sup> Similar findings were reported in these two further projects. The large EU SIENNA project which surveyed REC members regarding specific technologies including AI and Robotics, with no consistent resources used, some respondents indicating existing guidance sufficed, and others seeking further targeted support [81]. And a survey of US IRB committees with respondents from 63 distinct institutions, which similarly found both mixed responses to what should be required of researchers, and to questions regarding the IRB capability to assess proposals involving data and AI [86]

**Table 1** Recommended areas of foci for RECs in considering AI research, synthesised from Hine [41], Ada Lovelace Institute [2], and Ferretti [29]

Synthesis		Hine*	Lovelace <sup>^</sup>
1	Learning opportunities for ethics committee members	1. Ongoing learning training be provided to ethics committee members and the research community regarding the ethical implications of data and AI research	4. Training and knowledge hubs for ethics committee members
2	Governance regarding industry-based and -collaborative research	2. Governance be reviewed regarding the consultancy-research distinction to ensure research is not being conducted outside of ethics processes under the auspices of consultancy	5. Greater transparency from corporate labs (also row 3)
3	Establish expectations regarding norms of rigour in engagement with stakeholders	3. Requirements for stakeholder engagement regarding the application of data and AI in research be formalised	5. engage stakeholders and those with external expertise (also row 2)
4	Governance regarding potential for ongoing impact of work	4. Ongoing review be adopted to address the potential for unanticipated and unintended consequences	1. Broader impact statements as part of an ethics application 2. Multi-stage ethics reviews e.g. at the point of data collection and the point of publication
5	Governance regarding the constitution of ethics committees with suitable expertise	5. Expectations regarding the input of different types of expertise—technical, stakeholder, and (our addition) disciplinary—be formalized	3. Broader expertise in research committees
6	Establish norms and culture of responsible research embedded across the research ecosystem	If institutional ethics committees cannot undertake this work, make this clear, and explicitly articulate expectations for researchers to undertake other forms of audit <sup>#</sup>	6. Funder consistency regarding application of ethical principles to AI and data work 7. Incentivisation of responsible research culture 8. Increased funding for ethical review of AI research

\* See [41], 476) on the adequacy of ethics committees for AI research. <sup>^</sup> See [2], 8–11). <sup>#</sup> Note this point was made after the 5 recommendations (unnumbered) and is treated as a key recommendation here. These recommendations are also paralleled by [30]



regarding research with big data and thus inform the decisions of other actors involved.” ([31], 140).

### 2.3.2 Beyond principles

The importance of learning regarding the application of ethical concepts to AI research has been highlighted. However, as noted in the introduction to this article, while a significant body of work has engaged in developing guidelines and principles for ethical AI with an aim to disseminate and educate a variety of audiences, the operationalisation of these principles into organisational structures, practices, and professional reflection, has received less attention [50, 60, 72, 89, 91]. As Resseguier et al., put it: “this identified gap in AI ethics finds its root in the very nature of the currently dominant approach to AI ethics, i.e. a view on ethics that considers it as a softer version of the law. [They] point to the need to complement this approach [...and...] call for a shift of attention in AI ethics: away from high-level abstract principles to concrete practice, context and social, political, environmental materialities.” ([72], 3).

Awareness of these principles and guidelines is important, and has positive impact on intention to consider ethical issues, by providing orienting devices for stakeholders to think with. Specifically, [22] surveyed > 1,000 managers in the US, randomising presentation of different four groupings of AI regulations and asking about ethics in AI and their intent to adopt AI. They found “*that information about AI regulation increases manager perception of the importance of safety, privacy, bias/discrimination, and transparency issues related to AI. However, there is a tradeoff; regulation information reduces manager intent to adopt AI technologies.*” ([22], 1). Similar reflections were provided by Miller and Coldicutt [59] who polled UK ‘tech workers’ ( $n = 1010$ ) finding that 81% of those who worked on AI tools ( $n = 155/192$ ), “would like more opportunities to assess the potential impacts” (p.10). Thus, principles can be useful tools insofar as they offer orienting devices to think with. However, learning to engage in ethical practice goes beyond principles in addressing at least four key concerns:

1. *How do we learn to operationalise principles in context*: Principles provide useful anchors, but we must learn how to work with them with particular contexts and people, noting that ethical boundaries may change over time and location. As Resseguier et al. put it: “ethics must entail a sharp attention to specific situations and relations, accounting for the different levels of the personal, the interpersonal, the organisational, up to broader social, political, and environmental configurations” ([72], 10).
2. *How do we learn to navigate tensions between principles*: Classic dilemmas include freedom vs equality, or free speech vs privacy, and there is significant literature on this topic.
3. *How do we learn for a substantive ongoing ethics over procedural ethics*: There are questions around (1) how we probe *why* a tool is being implemented, and (2) whether reinforcement of existing systems closes off opportunity for work that develops futures worth wanting? Focussing on ‘doing things ethically’ can lead to abstracted models of action that fail to interrogate underlying aims in, for example, developing particular tools, including how they intersect with existing power relations [6, 72]. As work on Data Feminism highlights [24], current approaches to AI ethics are inadequate to addressing structural entrenched inequality and the material reality of AI development. As [48] note, a focus on ethics in the technical design of systems misses significant concerns (including in their proposal for an ethical ‘Algorithmic System for Turning the Elderly into High-Nutrient Slurry’).
4. *How do we assess the indirect, long-range, or soft, impacts of our work*: Principles used in research ethics have typically focused on risks to participants, and relatively direct and immediate impacts more broadly (sometimes excluding risks, focussing only on possible benefits). These direct impacts may be relatively predictable, perhaps through a hypothesised pattern of causation and modelling of their likelihood of occurrence. However, many technology systems have broader and long-term impacts. These occur in the ways they may reorganise social relations, and re-shape normative assumptions regarding human value and values.

### 2.3.3 Resources for learning AI ethics for research

Where, then, should we look for resources to support this learning? In their work surveying 54 and interviewing six AI practitioners, Morley et al., highlight that “the AI ethics community is not yet meeting the needs of AI practitioners” ([62], 6), with more practitioners saying further resources would be useful than those who say that what exists is already adequate, across a range of types (from principles, to design guidelines, and ‘best practice examples’) [62]. Where lessons are being drawn from other parts of the community, the historical parallels—for example, the sharing of security flaws in software as a defensive practice—may not carry over into AI [76]. Various resources exist, including worked cases [3, 67], and one helpful example of how one might complete a REC form [23], and emerging reviews of materials such as those in the ‘Responsible AI Pattern Catalogue’ [54]. Understanding how these resources are being, or could

be, mobilised including via the crucial role of RECs is thus important.

Indeed, this is a challenge across fields: understanding how researchers develop and express research ethics, [10]. Based on their review of papers discussing ethical issues in research, Beauchemin et al., [10] highlight a dominance of descriptive ethics, with relatively little use of established definitions or reflection, leading them to call for a greater focus in research outputs on articulating the ethical concepts used [10]. This is particularly salient given RECs may draw directly on literature (or expect applicants to include relevant disciplinary literature) regarding ethical issues. However, if discussion of ethical issues is unusual in academia's primary mode of communications—research outputs—where should researchers and RECs look in seeking to increase “sensitivity to ethical issues [and consider] how empirical data may be relevant to various ethical principles and problems.” ([27], 16).

In their specific analysis of AI in mental health initiatives, Gooding and Kariotis highlight that ethical and legal issues tend not to appear in the peer-reviewed literature, even if they may have been considered in the REC process [36]. Perhaps as importantly, they also flag that most publications in the space report on pilot work, thus obscuring the potential long-range impacts of the research [36]. Even more concerning, one analysis of over 227 publications on health technology (from an initial pool of over 14,000 returned) indicated that approximately half made no reference to ethical principles at all [79]. In a review of software engineering journals, Badampudi et al., [8] report that roughly half discuss one of consent, confidentiality, or anonymity, but only 6 of 95 reviewed discuss all three [8].

Here we see how the roles of advisory bodies, formal RECs, publication processes, and the guidelines and principles come together in an ethics ecosystem [73]. In this ecosystem “individuals (researchers), organisations (research institutions and the various committees within) and external bodies (publishing houses, funding bodies, professional associations and the governance policies they produce)” ([73], 317) participate in developing understanding and evaluating of ethical behaviour, through their roles in the research process.<sup>5</sup> Moreover, we see how different components of a system come together to act on ethical thinking, and provide resources for that thinking. Adopting this view, Chi et al., [19], analysed AI ethics documentation regarding diversity and inclusion, within three large AI infrastructure companies. This expansion of “the range of documents past

high-level corporate principles sheds light on how firms translate principles into action and provides greater clarity about the problems and solutions they hope to address through AI ethics work.” [19]. Through this analysis, they highlight that diversity and inclusion initiatives within these companies is configured to an “engineering logic”, thus while they claim AI ethics expertise, they act as “ethics allocators” pushing decisions regarding impacts of tools downstream to customers [19].

Importantly for this paper, they highlight a key claim: That on the one hand, the various sorts of documents or material resources organisations produce and draw on are reflections of (or, ‘containers for’) value statements, while (on the other hand) they also shaping this discourse (reflection on) through the resources they provide and the particular kinds of narrative they encourage and recognize as genuinely ethical. In this way “they are a kind of agent, educating clients, the public, and the broader field, articulating and defending values, developing scripts for ethical action that allocate work and responsibility to internal and external actors, and constructing the knowledge and expertise AI ethics work requires.” ([19], 2).

### 3 Method

#### 3.1 The materiality of research ethics

Despite the significant body of work drawing attention to the ethical impacts of AI, alongside corresponding guidelines and principles, relatively little is known regarding the resources drawn on and produced through the workings of actors within the research ethics ecosystem (including researchers, those impacted by the research including participants, and ethics committee members and secretariat).

These diverse components of ethics ecosystems, including the ethics process itself, are forms of knowledge which, as Freeman and Sturdy [34],<sup>67</sup> put it, are inscribed, embodied, and enacted. *Inscribed* in different kinds of artefacts that encode knowledge, including ethical principles and templates, that are made available for use across contexts. *Embodied* within individuals who bring this knowledge to bear in their actions, often in implicit ways. And *enacted*, in the sense that new knowledge emerges from interactions, and is available for use in, particular contexts. An example offered by Freeman and Sturdy is helpful: “When a

<sup>5</sup> The research ethics ecosystem can of course also be connected to other research institution policies, including data and privacy regulation (and committees relating to these), and the broader structures and regulation for responsible AI beyond research contexts and the relevant material resources and their design characteristics [54].

<sup>6</sup> A similar framing is provided by [83] in analysis of research ethics.

<sup>7</sup> As an aside regarding the social nature of research. The lead author attended a workshop run by these authors as part of a large EU project just as they entered postgraduate research (12 years ago); the benefits of academic meetings are often slow, and diffuse, a point which is salient in consideration of immediate and long-range impacts and consideration of knowledge infrastructure.

*committee convenes, embodied and inscribed knowledge is brought into the room in the form of what each of its individual members knows [embodied], whether through education or experience, and in what has been recorded in the minutes of previous meetings and in the documents prescribing the committee's remit and procedural rules. But the committee's knowledge is not limited to what is brought into the meeting. In the course of discussion, the committee may generate new knowledge: new ideas and insights, new aims, and new rules for how to fulfil them [enacted]" ([34], 12). How the learning transfers beyond the members of the committee is an important question to ask.*

Because many resources that inscribe knowledge in an ethics process are used explicitly as objects by multiple agents—researchers, REC members, stakeholders, compliance organisations, to name a few—these resources act as boundary objects [4, 16, 78]. In this way, resources such as formal policies, standardised forms, principles, and learning resources, are materials that *inscribe* knowledge in order to span across context and actors. Simultaneously, they are interpreted and reinterpreted in context, thus their meaning is not only held in the resource itself, but in the way its knowledge is mobilised and negotiated (or, enacted). Resources such as REC application forms and the materials to which they refer thus provide textual lenses reflecting both stances in their own right, and instruments that shape discourse [63].

### 3.2 Mapping research ethics

In Australia, the primary research ethics document—with legal standing in national research governance systems—is the National Statement on Ethical Conduct in Human Research [65] (henceforth “National Statement” or just “Statement”). While research ethics naturally extends beyond this document, and the document is grounded in histories of practice, culture, and artefacts, here we will treat it as the first order ethical document. From this document, we can see second order materials arising through the ethics process, at varying steps removed from the National Statement:

1. Institutions develop ethics forms, intended to support researchers in instantiating responses to the Statement's key principles.
2. Researchers then complete these forms, with reference to the Statement, and other soft (e.g. disciplinary guidance documents) and hard (e.g. privacy legislation) policy.
3. Completed forms are evaluated by RECs, and their evaluations are articulated (using National Statement concepts) with the intent that researchers will respond to them.
4. Researchers then conduct research, within the scope of their REC approval, and ongoing commitment to ethical

practice (which may be further informed by disciplinary and cultural norms).

5. Journal editors and reviewers, funders, and RECs, will then review submissions at different levels of granularity regarding the completed work, for internal or external reporting/dissemination. These should include reference to the REC process (at minimum reporting that one was completed), and any issues arising.

Therefore, to map the material ethics ecosystem we conducted a review of:

1. Resources available institutionally, to be drawn on by the REC process.
2. Where these resources are taken up in research practice, through a systematic search of our internal REC application database, and external publication databases.
3. And within these materials, an analysis of the REC application detail, supplemented by semi-structured interviews of the respective researchers, and further review of published outputs.

While previous research [19] has mapped documents from multiple organisations to analyse expressions of values and ethics, we instead focus on a single organisation. In that prior work documents were coded as representing different functions regarding ethics: (1) pedagogic tools; (2) product documentation; (3) legal/policy; (4) general communications. As our focus is on understanding networks of resources linked to specific projects internally, and how this analysis can help us understand the socio-material context of the work reflected, we develop an approach informed by Chi et al. [19]. Specifically, our analysis maps out the following materials:

1. Pedagogical tools—specifically guidelines, courses, and scholarly outputs such as reviews of ethics strategies, of participant preferences, etc.
2. Process resources—these include materials such as ethics proformas in document or web-form format, for example focussing on data protection, and REC application details
3. Legal and policy instruments—these include statements of principles, the Australian National Statement on Research Ethics [65], and legal instruments such as relevant privacy legislation
4. General communications—these include any available communications from the institution referring to relevant issues, made available through general (rather than ethics targeting) channels
5. Discursive resources including REC consultation, and stakeholder consultation (or other forms of input, such as codesign)



6. Reflection on practice, including any expression regarding previous experience (e.g. provided in REC applications), or experiences of relevance within the project (e.g. in publications, or public reflections).

As described above, analysis of these material resources frames these resources as providing an expression of, or lens onto, the conceptual space that shapes and is shaped by ethical discourse.

### 3.3 Interviews

Interviews were conducted based on invitations to researchers from a higher educational institution who had been identified as submitting relevant applications in our search process (described below).

A semi-structured interview schedule was developed, to understand perspectives of the researcher stakeholders regarding their use (or otherwise) of ethical frameworks in their research on data and Artificial Intelligence (AI). Participants were invited to speak about their organisational contexts—which, for some, crossed university and industry settings—and any practical challenges in use of AI and advanced technologies and approaches to address these ethically. Interview questions (Table 2) were developed to probe the dimensions described above, regarding:

1. Developing approaches to ethical concepts and principles
2. Learning to navigate tensions and challenges
3. Procedural and substantive ethics: Process (and adequacy) of REC in mediation
4. Challenges in AI research and soft impacts

The questions were designed to leave open the discussion of principles used and any ways these were identified and navigated by participants, and to allow for discussion of the range of pedagogic, discursive, reflective, legal, and other resources used alongside the formal REC process and any others followed.

Interviews were scheduled for 30–60 min duration, conducted via online video conferencing. They were conducted following a consent process in which we requested access to key REC materials related to projects of relevance (described below), these thus act as an anchor for the interviews, acting as a preliminary stimulus and material artefact. We also provided reference to other principles in advance via the introduction to the interview, including the National Statement, Australia's Artificial Intelligence Ethics Framework [7], and the Human Rights and Technology Issues Paper:

UTS Submission [85]. As a semi-structured protocol, not all questions were asked of all participants, although all themes were introduced in all interviews. The initial questions often naturally led to further discussion of ethical issues and the role of the REC, and the interview protocol served as a guide to steer these conversations (even where the questions were not explicitly used).

The interviews were conducted by a single researcher, who also implemented the first analysis of the interview and REC material data. The interviews were professionally transcribed, and these transcriptions were selectively coded alongside other research texts (submitted REC applications and files) drawing on approaches to discourse and document analysis [14, 71].

In reporting, the transcription convention used is that [...] indicates words were removed (where these are not relevant to the key issue), and [unclear] indicates words that were inaudible or unclear. A non-verbatim transcription is used, with non-linguistic features (gesture, and fillers such as um, er, etc.) not transcribed.

### 3.4 Ethics

The work was internally funded through a faculty seed grant. REC material may fall under the intellectual property of the institution or be considered internal material for the purposes of evaluation and quality improvement. However, because of the research intent of the work, and the inclusion of semi-structured interviews, a REC application was submitted (ETH216658) and data sharing agreement put in place, building on an earlier application (ETH205567) regarding use of ethical frameworks based on responses to a public consultation on AI ethics. This updated application provided approval for:

1. a search to be conducted on the REC database for keywords across titles and summaries, with results provided to the authors;
2. the authors screening these as described above;
3. the authors contacting lead researchers on relevant projects, to seek their consent to access their full REC materials, and invite them to interview (these were treated as separate consents);
4. the authors liaising with the REC secretariat to provide consents (where given) for sharing the REC material for the stated purposes; and
5. using the REC materials to inform the interview discussions, where those occurred.

Separately, we also sought references to REC approval in published works (as above). The reporting here is not

**Table 2** Semi-structured interview protocol setting out thematic area of probe, specific questions, and the issues we anticipated eliciting

Probe	Questions	Intended to elicit
(1) General approach to ethics	Within your own organisation, practice, or/and research, do you draw on or use a set of AI ethics principles or frameworks? If yes, How did you produce or select those principles? How do they relate to the materials (in your HREC application, the national statement, or other frameworks)?	Reference to principles resources, and ground these in the connection to the REC process and specific project
(2) REC role, resource, and experience	Did the HREC process prompt you to think differently about any aspect of the way that your research or organisation approaches ethics? Are there any approaches/steps that you took for your HREC application that you would not necessarily take in a non-research context or in your organisation? In creating the materials, did you find it difficult to select and apply ethical principles to the research, from the different ethics frameworks available? (With respect to the HREC materials, what issues were raised by the stakeholders—this might be other researchers, the HREC, or external stakeholders such as research participants—and how were they addressed?)	Experiences of the REC, how it was used as a resource, and is perceived in relation to institutional-procedural ethics vs wider application
(3) Drawing out specific challenges	Do you have specific examples from your context where you consider how to ethically use data and AI? Are you able to easily apply the AI principles and frameworks to applications in your own contexts? What ethical dilemmas (if any) do you face when dealing with the application of data and technology in your own contexts? What issues have you faced, or do you think you might face, in applying ethical principles/frameworks in your context? E.g. conflicting with your policies, being unclear how you should apply them (for instance, what is ‘fair’ in a specific context), conflict with other regulation, etc.? Thinking about ‘users’ of AI systems How should we inform users about the use of AI and data collected? The user agreement is something that many do not read! What problems may arise as the result of the lack of explainability and transparency in your area? In terms of accountability and privacy, do you have a mechanism that users can refer to?	Specific challenges and tensions and the role of materials (principles, tools, policies) in the instantiation of these issues and their navigation
(4) Connecting to users		Reflection regarding some specific largely technically oriented connections of ethics to AI system users
(5) Connecting to organisational practice and regulation	How much are you aware of the regulations relevant to AI? Does your company/institute have a specific sector/group to deal with application of ethical issues including AI? How do you share ethical issues within your organisation or sector What do you think about a platform/mechanism to collect examples of issues raised by the application of AI ethical principles? Would concrete examples (1) help you or your stakeholders think about how to navigate ethical issues? and (2) help you to understand how people are prioritising different outcomes? How do you engage with legal aspects of AI in your organisation (e.g. through dedicated staff, training, etc.)?	Further elaboration regarding ethics beyond the research ethics process, and resources drawn on in this

intended to identify specific authors or their work, and we have sought in our aggregation and excerpts of interviews and other material to maintain confidence and reduce risk of re-identification.

The reporting here is also not an evaluative reflection of any work noted, at an institutional or individual level. Our analysis is limited to the data available to us, selected through a particular search strategy at a single institution. We have no reason to suppose that this data is particularly unusual, but nor do we make claims about generalisability either at our institution or more broadly. Rather, our interest is in how the process of conducting such analysis may inform understanding of ethics processes, and how our specific study may provide broader insights.

## 4 Data—instantiations of ai ethics resources in use

### 4.1 Mapping ethics resources

In our first step, we sought to map the institutional ethics ecosystem, using the model described above, and drawing on the visual representation in Chi et al., (see, [19], 4). That foundational work analysed multiple technology companies and their expression of ethics and values with respect to diversity and inclusion. A helpful step in their representation was to (1) colour code documents according to department or product space within the organisations, and (2) draw connections to explicitly highlight how documents referred to each other. Neither is appropriate in our case. That is because (1) the documents we are drawing on are all within the research governance space, with the exception of “general communications”, and (2) the documents are highly interconnected in their present form (again, with the exception of general communications). Figure 1 below indicates the set of resources returned through searches of both internal and external sites. In addition to these resources many other materials may be drawn on by individuals, groups to which we do not have access, and from external sources. Our intent here is not to suggest this resource set is exhaustive either of the set of resources within the institutional ecosystem, or—clearly—of the set of relevant resources in the wider ethics ecosystem.

### 4.2 Search strategy and output

A term-based search was conducted on all REC applications, using the centralised system through which all such applications are submitted. This system allows for searching over the text-field submissions, which comprise most of the application, barring attachments which typically consist of items such as: consent forms; participant information

sheets; budgets; organisational approval letters; data collection materials of various sorts, such as survey instruments and interview protocols; elaborated answers to text fields, such as rationales for particular approaches, study design diagrams, etc. At the point of the initial search, the RECs received 6–700 applications per year across the full REC panels and faculty level delegation.

The initial search was conducted in October 2021, for applications dating from 2015 (when the system was launched). A follow-up search was conducted in September 2022. Applications on which any of the co-authors were an investigator, or involved in the research in a non-investigator role (e.g. advisor, participant, student-of), were excluded. In some cases, no researcher was still at the institution, and these applications were excluded. Some researchers had multiple studies identified, in one case two submissions were discussed in interview; in others, the researchers either declined or did not reply to an earlier invitation, and thus any later applications were also excluded. Results of this search and screening are summarised in Fig. 2.

### 4.3 Research outputs

To complement our search of REC applications, we conducted a bibliographic search of the Web of Science (WoS) core collection (Fig. 3), which provides comparable coverage to Scopus as an indexed article collection [12, 77]. This approach was intended to (1) act as a check on further applications that may have been missed in the internal-system search; and (2) provide further insight regarding the expression of ethics by researchers, through analysis of reflections of ethics in their published works.<sup>8</sup>

We also conducted a search for obtained REC numbers (e.g. searching for “REC-15000”) in Google Scholar, to supplement the materials in the REC process, though this did not identify any further material.

<sup>8</sup> While this approach was intended to augment our internal search, it may underreport on relevant material given that (1) WoS provides an incomplete archive of all scholarly works; and (2) WoS search is based on article metadata (including title abstract and keywords), and not full text. However, in contrast to more complete indexes such as Google Scholar [56] WoS provides more advanced search functionality, including full Boolean search and search over metadata fields. This is particularly significant when searching for terms such as “REC” or “Research Ethics Committee” where their discussion may be incidental. A limitation of this approach is that it requires ‘ethics’ to be explicitly mentioned, however in our context this maximises the chances of retrieving publications with a substantive discussion of ethics.

See p.9 discussion which indicates that levels of reporting in publications are low.

<b>Principles</b> <ul style="list-style-type: none"> <li>• <a href="#">National statement on Ethical Conduct in Human Research</a></li> <li>• Institutional Research Impact Principles (human centred; improvement oriented; translational)</li> </ul>	
<b>Pedagogical tools</b> <ul style="list-style-type: none"> <li>• Exemplar REC applications from disciplines (including one using social media data under a consent waiver model)*</li> <li>• Self-enrol <a href="#">online courses</a></li> </ul>	<b>Process Resources</b> <ul style="list-style-type: none"> <li>• REC process guidance, and forms. Including proforma for specific themes including direct quote of social media, parent/guardian consent, photo consent, biobank protocols, and clinical trials*</li> <li>• <a href="#">Guidance</a> re: quality assurance vs research distinctions in REC process</li> </ul>
<b>General communications</b> <ul style="list-style-type: none"> <li>• A search of publicly available material (excluding courses, and publications) indicates (7/3/2023) ~26 documents available in Google, using the query string: ("ai" OR "artificial intelligence") "research ethics" site:uts.edu.au -opus -epress -search.lib -handbook.uts. Relevant documents set out aspects of work at the institution and the institutional framing of AI ethics in these contexts, specifically: <ul style="list-style-type: none"> <li>• <a href="#">In education</a> and <a href="#">deliberative democracy for AI in education</a></li> <li>• <a href="#">In public policy and human rights</a></li> <li>• <a href="#">Targeting facial recognition</a></li> </ul> </li> </ul>	<b>Legal/Policy</b> <p>Many governance documents may be relevant. Targeted pieces include:</p> <ul style="list-style-type: none"> <li>• Overview of Integrity governance</li> <li>• Governance of Research Ethics</li> <li>• Overview of Research Ethics</li> <li>• <a href="#">Australian Code for the Responsible Conduct of Research</a></li> <li>• <a href="#">AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research (the AIATSIS Code)</a></li> <li>• Privacy policies and relevant legislation (including the Australian Privacy Act), compiled for institutional context</li> <li>• Internal REC pages setting out key responsibilities, including Research policy, code of conduct, HRE guidelines, privacy policy, and child protection policy*</li> </ul>
<b>Discursive Resources</b> <ul style="list-style-type: none"> <li>• Faculty level “A Research Integrity Adviser is an experienced researcher with a sound understanding of responsible research practices who can provide advice to anyone who has a concern about research conduct. UTS has appointed seven Research Integrity Advisers across UTS to promote and support research integrity.”</li> <li>• Expectation that e.g. student projects are first discussed with supervisors prior to submission</li> <li>• Drop in ethics clinics for researchers</li> </ul>	<b>Reflection on Practice</b> <p><i>No explicit references</i></p>

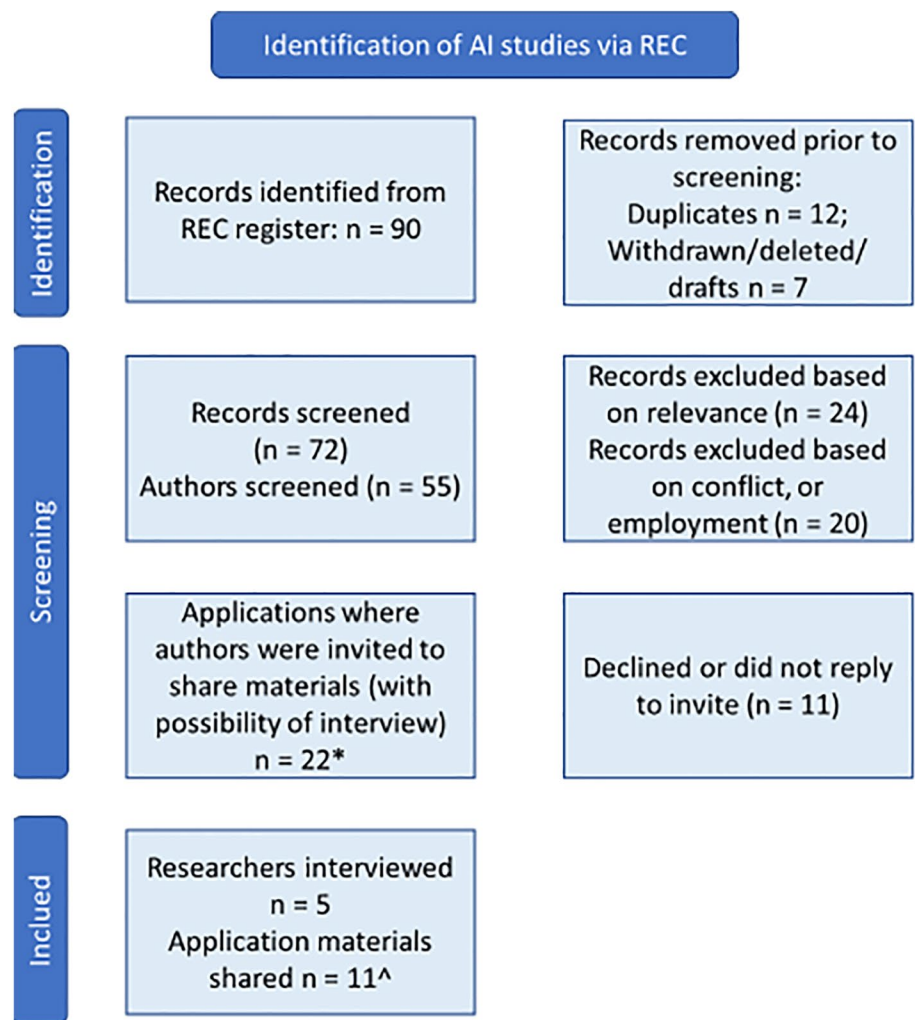
\* Internal only

**Fig. 1** Mapping the institutional ethics ecosystem

Finally, we also conducted preliminary searches of the institutional repositories (an Open Access self-hosted repository, and via the Dimensions database, with which we have an institutional arrangement), using a full-text search for the same terms. These searches were not systematically reviewed due to significant overlap with the WoS search which yielded data saturation.

## 5 Results and analysis of ethics materials and interviews

The materials retrieved were analysed and drawn upon to identify and invite interviewees. From the  $n = 11$  applications shared, the set of resources drawn on explicitly within the application, or via the interview data, were mapped using the framework in 1. As indicated in Fig. 4 there is significant

**Fig. 2** Search strategy for AI Studies via REC

\* Researchers were excluded at this stage for a range of reasons, some new applications identified authors who had previously declined or not responded. In some cases researchers would be away or unavailable.

^ One researcher had two relevant projects (both shared); two researchers shared materials but were not interviewed; one researcher was interviewed but did not share materials

overlap between the resources *available* in the ethics ecosystem (Fig. 1), and those *drawn on* (Fig. 4) in practice, notably:

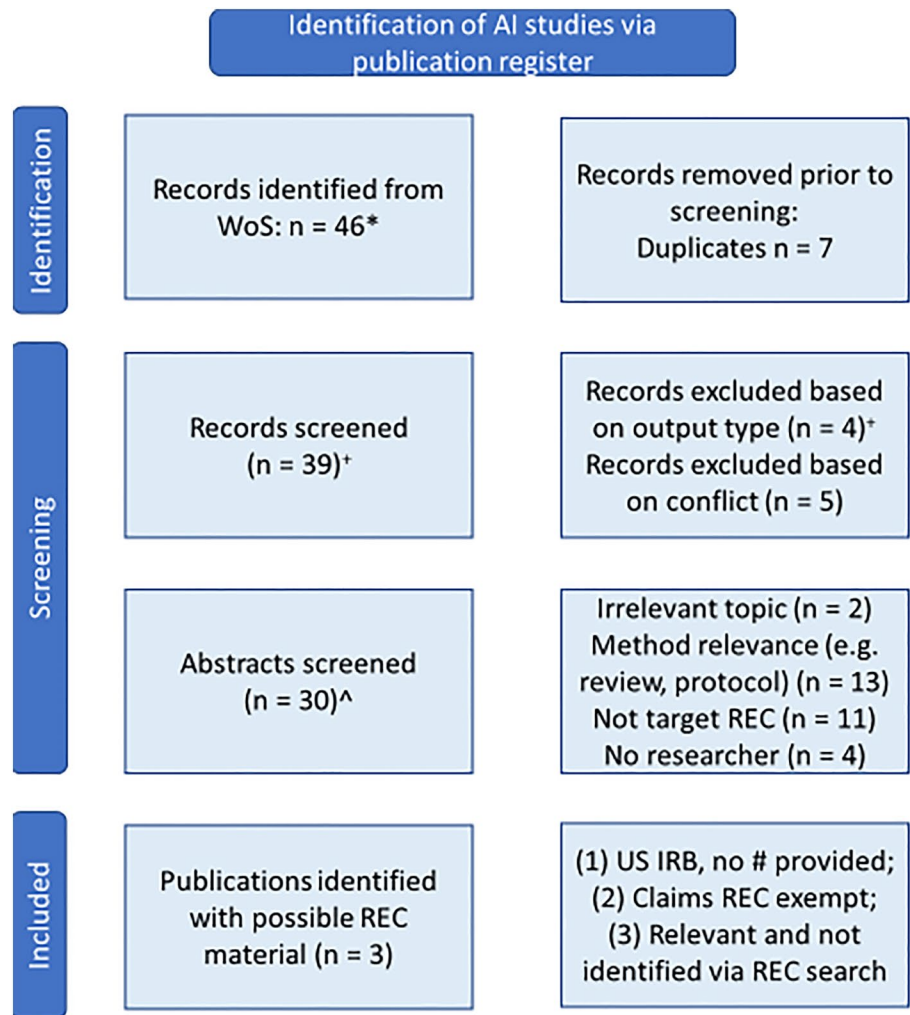
1. The National Statement featured as a central principles document
2. The REC process itself was explicitly noted as drawn on in ethical consideration
3. The Australian Privacy Principles (APPs) and generic 'university policy' provided some policy context

4. Discursive resource via colleagues (peers, supervisors or other senior colleagues), and other stakeholders were mentioned as a key resource

However, as our interview data indicates, the depth of use of these resources is unclear in places. The pedagogical tools and general communications referred to were targeted at the specifics of the projects, and thus differed significantly from those available via the institutional ecosystem. However, although some resources in this internal ecosystem were relevant to AI, with the AI ethics principles being clearly



**Fig. 3** Search strategy for AI Studies via publications



\* Web of Science search (note, WoS searches all metadata including abstract and keywords, but not full text) conducted 18/10/2022: (((ALL=("artificial intelligence" OR "machine learning" OR "data mining" OR "predictive mining" OR "predictive model\*" OR "classification" OR "learning analytics" OR "data science" OR "deep learning")) AND ALL=(ethics OR ethical))) AND OG=(University of Technology Sydney)

+ keynotes, short papers/work in progress.

^ Not mutually exclusive, 27 distinct records of the 30 remaining excluded

highly relevant, the former were not drawn upon at all, and the latter were only mentioned once.

Table 3 sets out key responses from the pool of interviewees, mapped against the four key concerns in learning to engage in ethical practice (see p.8). The five researchers interviewed are identified (R1-5), and the topics of their research projects were:

1. Transcript 1: Understanding (through self-report methods) organisational practices for data projects, using both qualitative and quantitative methodologies (such as path analysis).
2. Transcript 2: Interviewing developers of an AI system to understand how their design practice avoided bias.
3. Transcript 3: Developing and deploying a system at a field site, including secondary analysis of data captured on site (with removal of any data that could de-identify people on site, prior to receipt by the researchers).
4. Transcript 4: Effective delivery of data science initiatives in a specific sector.
5. Transcript 5: How organisations manage and use their AI technologies.

Interviewees 3 and 4 were building AI tools via their research (others may have been in other capacities). This may suggest that in the *process of REC submission*, while information regarding methods is elicited, this elicitation does not capture the range of relevant approaches adopted.

*Developing approaches to ethical concepts and principles:* The participants were invited to consider nationally relevant ethics principles, alongside which they noted national privacy principles, and the European General Data Protection Regulation (GDPR) in passing. Participants referred to self-reflection in contrasting ways, with regard to a trigger for seeking out an ethics framework (R3), and the idea that “My ethical framework is myself, and that’s good enough, I think.” (R1).

*Learning to navigate tensions and challenges:* Participants referred to challenges in operationalisation of principles not only into practice but into other material forms, for example saying “so this is what the documents say, and how are we going to transfer it into our ethics applications” (R3). This went alongside a sense that outside of the university context, ethics is not a consideration in research and development, with one researcher (R3) who contacted external researchers reporting that “they don’t have ethics or they don’t really care about the ethics around this.” (R3).

In discussing the published outputs of their research, R3 notes that a core concern of their methodology was to ensure that the site of their research could not be re-identified from images contained in these outputs. For instance, R3 described their response to a person who, at a conference presentation, asked about their collection process and ethics: “I told them, if you want more details, this is the ethics number. [...] So, you can contact us. Because [they were] very interested. [They] wanted to do something similar. And [they were] interested about the ethics and the data collection process around it.” (R3). Another (R2) notes their research was informed by a well-known case of “AI failure”, with part of the work investigating how designers seek to avoid these kinds of biases: “my idea was, are there processes that we can put in place to prevent that from happening” (2–1). In both cases, we see how knowledge is inscribed in resources made externally available for shared learning.

Tensions between data quality and ethical considerations were a recurring theme, as were reflections on whether standardised processes could help navigate such tensions. For instance, R3, who required images of a physical space, but not the people in it, noted that “one of the main questions that was raised was do we have the consent of those [people] to be appearing in the video. [...] if I had set up the cameras by myself, then I would have [inadvertently] captured the [people]”. To navigate this challenge, secondary data (alternative images) were provided and filtered to ensure people were not visible, but this meant that the images obtained were not captured from a position the researcher would have

chosen. Two researchers (R3 and R4) noted the challenge that high-quality imaging increasingly makes it harder to de-identify subjects by filming from a distance, because bystanders may still be recognisable even if they were not the intended target of analysis. R4 commented: “Then you say, hang on, I got 25 projects trying to do the same. What can we standardise? What’s the guiding principles? What are the governance frameworks?”. However, as R1 observed, a challenge for such standardisation, and broader concerns about consent, is that particular research methods used may require bespoke (i.e. not standardised) approaches: “An interesting experience I made recently is that people don’t understand my analysis, not even academics, and that might make it a bit complicated in terms of, maybe, ethics as well.” (R1).

*Procedural and substantive ethics: Process (and adequacy) of REC in mediation:* Participants reflected on tensions between procedural and wider ethical considerations, including features of the REC process and requirements around such things as data privacy. R5 observed: “it’s stipulated by the university what you need to do and how you need to keep your data [...] So, it’s not really an ethical decision. It’s more like there’s rules to follow. So, I don’t need to make any ethical decision.” Later also noting that “there’s a difference between following the law and an ethical decision” (R5).

Nevertheless, participants recognised that established REC processes supported ethical reflection by encouraging them to “think about things a bit differently” (R1) and “stop and think a moment” (R2), and even suggested: “I think it would be wise if more organisations would have an AI ethics committee to stop and think before they build the AI because there are so many problems around this area. And many organisations don’t stop and think. They just do, and as a result we have a lot of biased AI and a lot of problems. So, I think the concept of having an AI ethics committee can be very, very valuable and we should actually move that from the university to the more corporate world as well.” (R2).

The importance of fine-grained contextual factors, and not merely relying on generic procedural ethics, was also noted. For example, that consent practices must be adapted to specific contexts, moving beyond basic procedural requirements. R3 notes: “We have to establish that dialogue with them. They’re not into reading consent forms, user agreements. So, we have to do the face to face dialogue and to explain to them. And some of them didn’t even realise what machine learning, AI, deep learning means. For them, it’s like they think we’re doing something robotics when we talk about AI. So, it’s very understanding. Different people have different perceptions. [...] So, it goes beyond documented consent forms and user agreements. You have to have these dialogues, verbal communication. I think that’s very, very important in AI research.”.

<b>Principles</b> <ul style="list-style-type: none"> <li>• <a href="#">National statement on Ethical Conduct in Human Research</a></li> <li>• <a href="#">Australian AI Ethics Framework</a></li> <li>• Professional standards for a specific discipline (e.g., for degree accreditation)</li> </ul>	
<b>Pedagogical tools</b> <ul style="list-style-type: none"> <li>• In one case, the literature section of the application was referred to, although none of the sources appeared to explicitly mention ethics (rather focusing on the potential of big data for achieving various ends), knowledge of the literature (and ethics in this) was referred to by one interviewee.</li> <li>• Bostrom, N. and E. Yudkowsky, The ethics of artificial intelligence. The Cambridge Handbook of Artificial Intelligence, 2014: p. 316334</li> <li>• <a href="#">Elements of AI online course</a> (a part of which focuses on ethics)</li> <li>• AI oriented webinar</li> <li>• Podcast by international company with an AI ethicist</li> <li>• Well known example of 'AI failure' used as background to project</li> </ul>	<b>Process Resources</b> <ul style="list-style-type: none"> <li>• Consent form</li> <li>• HREC form itself and the process, as a tool for thinking</li> <li>• In two projects, ethics was a part of the project, reflected in survey instruments: (a) asking about capacity to act ethically (b) a trade-off between privacy and data use, (c) asking about use of explainable AI, and (d) harm avoidance (and specifically bias mitigation)</li> </ul>
<b>General communications</b> <ul style="list-style-type: none"> <li>• Blog of large AI organisation re: future potential of AI</li> </ul>	<b>Legal/Policy</b> <ul style="list-style-type: none"> <li>• Australian Privacy Principles</li> <li>• Privacy statement of tools used</li> <li>• Standards and state/national guidance alluded to</li> <li>• University policies</li> <li>• 'Legislation' (generally referred to)</li> <li>• International legislation (e.g. GDPR)</li> <li>• Reflection on role of regulators in requiring e.g. human readable privacy policies (with a bad practice exemplar from one large social media company).</li> </ul>
<b>Discursive Resources</b> <ul style="list-style-type: none"> <li>• Community representative group suggested for consultation by REC process</li> <li>• Team / supervisor discussion</li> <li>• REC 'drop-in' support</li> <li>• AI ethics framework and specific data privacy guidance suggested by REC process</li> </ul>	<b>Reflection on Practice*</b> <p>Reflection on a previous project and how new awareness of some ethics frameworks (for transparency, etc.) might have helped with project success.</p> <p>Some explicitly referred to conversations and dissemination around AI ethics.</p>

*\*The interviews of course provide a clear indication of reflection on practice. Here we are specifically interested in examples of resources that are designed to promote reflection, or/and instances where materials (including the interview data) refer explicitly to a prior occurrence of reflection, such as learning from previous experience on a similar project.*

**Fig. 4** Mapping Aspects of the institutional ethics ecosystem drawn on in practice. \*The interviews of course provide a clear indication of reflection on practice. Here we are specifically interested in examples of resources that are designed to promote reflection, or/and instances

Perhaps unsurprisingly, a technical framing of ethical concerns was another theme among our participants'

where materials (including the interview data) refer explicitly to a prior occurrence of reflection, such as learning from previous experience on a similar project

responses—for instance, seeking to employ technical approaches to explainability or bias to proceduralise

**Table 3** Summary of Key issues and resources identified in interviews

Area	Resources mentioned	Key concerns	Interview questions
Developing approaches to ethical concepts and principles	Core principles, Self-reflection	Trigger points for identifying need for ethics framework; tension in self-reflection and use of external guidance	Q1, Q1.1 Q2.2: Q5.1–5.4
Learning to navigate tensions and challenges	Principles, Process Resources, Pedagogical Tools	Operationalisation challenge Industry-academia and international variation in standards Shared resources for navigating issues	Q2.3, Q3, Q3.1, Q3.2, Q4.3
Procedural and substantive ethics: Process (and adequacy) of REC in mediation	Process resources Pedagogical Tools	Tensions in navigating e.g. data quality, consent, and deidentification	Q2, Q2.1
	Core principles	Technical concerns particularly highlighted, with concepts sometimes used interchangeably	Q4, Q4.1–4.3
Challenges in AI research and soft impacts	Core principles, legal and policy context	Perception of industry-academia, and international variation in standards	All

ethics—and using terms like “explainable”, “ethical”, and “responsible” interchangeably.

*Challenges in AI research and soft impacts:* Although the REC ethics process was generally seen as rigorous, and our participants viewed their own research as posing relatively low risks, concurrently they also observed that AI more generally might raise more- and different kinds of issues that the REC process is “*not reflective of, how should I say, the ethical implications for artificial intelligence for the whole of society.*” (R1).

Two examples of the relatively-uncontroversial nature of many current uses of AI were that human-in-the-loop systems are often used to mediate AI’s decisions, and that the purpose for which AI is used is often relatively tame. As R5 put it: “*not to be dramatic or controversial. So, I guess that’s an ethics thing that they are tapering their AI. They’re not making the extreme*” (R5). At the same time, though, they also recognised broader ethical concerns regarding responsibility to the conduct of science and the public’s trust in science: “*I guess that people feel aggrieved or unfairly dealt with. So, I guess if that feeling swelled, there would be less people who’d want to take part in my research if there was that feeling that it was unsafe, unsecure. And then I wouldn’t be able to conduct my research. Or if it grew wider, then no one would conduct any research sort of thing if there’s such mistrust there.*” (R5).

Participants also commented on issues in AI research around soft impacts and commercialisation, including that the ethical use of data and ethical use of AI raise different issues (R4), and their sense that there is a gap in ethical research and development outside of universities: “*outside, people are doing whatever they want*” (R1), and “*AI is not necessarily localised and AI is borderless, and organisations would need to apply to all these different regulations when*

*building the AI, which doesn’t really help in the process.*” (R3).

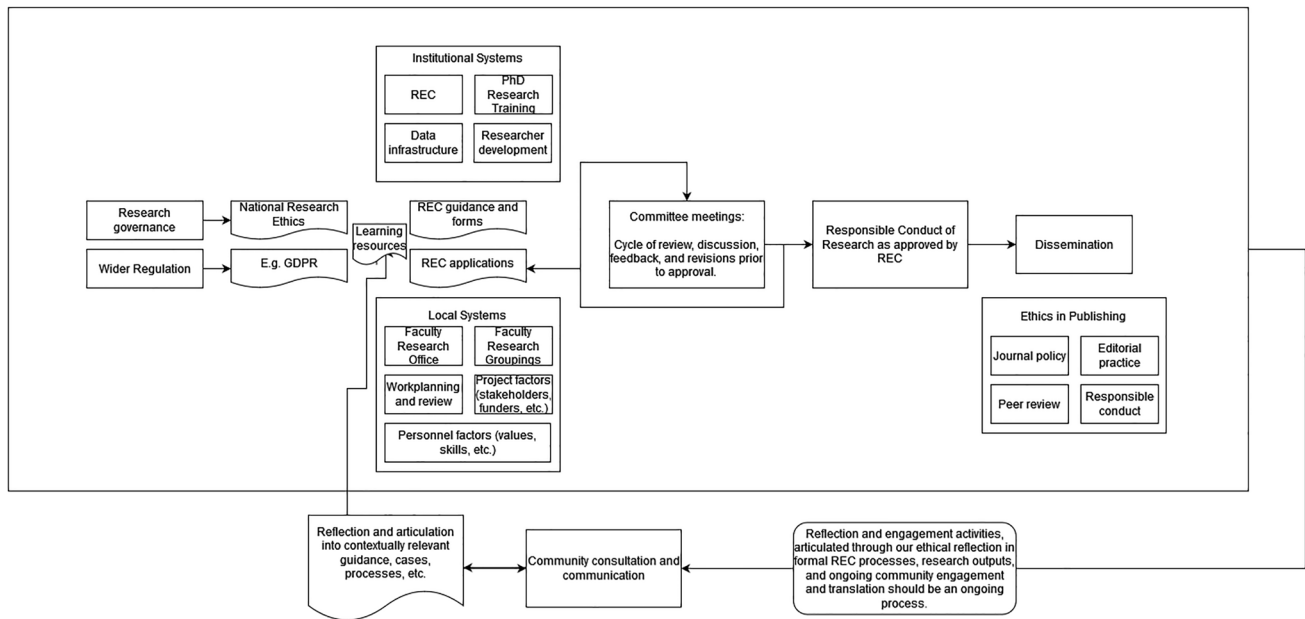
## 6 Discussion

Demands are emerging to put into place governance structures for AI research across sectors, inspired by existing research ethics governance models. In light of the findings of this research, we point to key issues and reflections in Table 4. As the table indicates, findings are largely consistent with prior work. The researchers were generally positive about the REC process as a means to support their reflection and provide oversight, however noting concerns regarding oversight of cross-sector work and long-range impacts. The implication, then, is that in considering the ethics ecosystem (Fig. 5), and how it draws on resources (Fig. 4), attention should be paid to how ethics governance and reflection can be inscribed so as to cross-institutional and temporal boundaries, in order to foster ethical reflection and action across all research (and in this context, all research involving AI) (Table 4).

Grounded in the findings reflected in Table 4, we propose a broader updated ethics ecosystem (Fig. 5) that builds on the governance recommendations reviewed (Table 1), Samuel et al.’s [73] ethics ecosystem model, and the features of it described in Sect. 3.2, highlighting the kinds of resources, and their role in learning, borne out in this research.

## 7 Conclusion

Rising awareness of AI has prompted increasing demands for its ethical governance and a plethora of ethical AI guidelines. RECs have a well-established history in ethics



**Fig. 5** Elaborated research ethics ecosystem

**Table 4** Reflections on recommended areas of foci for RECs in considering AI research

Focus	Reflections
1 Learning opportunities for ethics committee members	Generally, where the REC process was referred to it was referred to positively at this institution, with some concerns regarding inter-sector and international variation in approaches to research ethics
2 Governance regarding industry-based and -collaborative research	Concerns were raised regarding industry-based research, although these were a mixture of indirect anecdote and more direct (but not firsthand involvement) experience; participants indicated they thought the REC process would provide a valuable model for broader adoption in research and development across sectors
3 Establish expectations regarding norms of rigour in engagement with stakeholders	There were specific examples of stakeholder engagement, and the significance of shared resources for issues including consent, and particular forms of data. Gaps in available resources were also highlighted, indicating a desire for such resources to support engagement
4 Governance regarding potential for ongoing impact of work	Concerns for regarding ongoing ethical impact(s), beyond the temporal scope of the REC process, were expressed
5 Governance regarding the constitution of ethics committees with suitable expertise	There was a recognition that non-experts may assess REC applications, however no participant was critical in this particular context, instead noting that this is a standing challenge in all research
6 Establish norms and culture of responsible research embedded across the research ecosystem	The self-selecting interviewees all engaged with reflection on the ethical aspects of their research, with some tensions between ‘personal ethics’, ‘ethics as regulation’, and more ongoing discursive approaches to ethics-in-action. The range of resources drawn on and discovered in the process suggests a need to understand how researchers learn to engage with ethics and indeed how they learn about ethics resources of various forms

governance, but there have been concerns about their capacity to adequately govern AI research. However, no study to date has examined the ways that AI-related projects engage with the ethics ecosystem, or its adequacy for

this context. This project is based on a single institution, of projects identified via the particular search strategy, and notably only of those that undertook a REC application. These contingencies present limitations, although we



have no particular reason to believe that the results are particular to our institution (where AI is a strategic focus). Moreover, the model developed for analysing these applications presents a novel approach to understanding and assessing an ethics ecosystem, a contribution with broad application across both university and industry RECs.

Our results suggest that, despite calls for new structures, existing REC models can effectively support consideration of ethical issues in AI research. REC principles and processes were drawn on and referred to by our participants, and — in the Australian REC context at least — are embedded in a lineage of work on research ethics that is continuing to develop. Thus, where new materials are required, we propose that they should be embedded in this existing well-established ecosystem, rather than creating novel governance mechanisms tailored specifically to AI.

Gaps were identified in the resources drawn on, and by participants in the interviews. Participants expressed uncertainty about some practices, and noted that long-range impacts and issues such as secondary use of data may not be effectively addressed in existing guidance. However, it is not clear these issues are addressed in AI ethics guidelines, and indeed only one participant referred to use of AI ethics principles specifically, with multiple participants raising concerns that outside the research ethics context — a context from which these new guidelines have largely emerged — practices were more varied, and less rigorous.

One upshot of our study's findings is that the development of new AI principles may not be an optimal strategy for addressing ethical issues related to AI. Indeed, it is far from clear that the proliferation of AI-targeted principles has helped in practice. The results indicate that shared artefacts of practice, such as ethics applications and published articles referencing ethics, provide one lens (socio-material) into the practical usage of principles in context. These resources may be used to support learning by individual and organisational stakeholders. In tandem, organisations seeking to engage with ethics and AI should look to the well-established structures of RECs to build on this lineage. RECs themselves may develop further and support uptake in new contexts by evaluating how their communities — REC members, researchers, the public, etc. — learn regarding ethical issues, and where within institutional governance structures the kinds of issues specific to emerging technologies are addressed, and updated in an ongoing way.

**Acknowledgements** The authors would like to acknowledge the support of the UTS Research Ethics secretariat, in particular Racheal Laugery, for their assistance in this project. Our thanks to the researchers who generously shared their materials and time with us in interviews for their contribution. Mark Israel (Australasian Human Research Ethics Consultancy Services, AHRECS), provided helpful input on a number of issues regarding RECs particularly in international comparison. Our thanks too to Linda Przhedetsky for her research assistance.

**Funding** Open Access funding enabled and organized by CAUL and its Member Institutions.

**Data availability** Due to the nature of the research, and the legal and ethical restrictions on sharing of internal materials, supporting data is not available.

## Declarations

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Ada Lovelace Institute. 2021. 'Supporting AI Research Ethics Committees'. 2021. <https://www.adalovelaceinstitute.org/project/ai-research-ethics-committees/>.
2. Ada Lovelace Institute. 2022a. 'Looking before We Leap Expanding Ethical Review Processes for AI and Data Science Research'. Ada Lovelace Institute. <https://www.adalovelaceinstitute.org/wp-content/uploads/2022/12/Ada-Lovelace-Institute-Looking-before-we-leap-Dec-2022.pdf>.
3. Ada Lovelace Institute. 2022b. 'Looking before We Leap Expanding Ethical Review Processes for AI and Data Science Research Case Studies'. Ada Lovelace Institute. <https://www.adalovelaceinstitute.org/wp-content/uploads/2022/12/Ada-Lovelace-Institute-Looking-before-we-leap-Case-studies-Dec-2022.pdf>.
4. Akkerman, S., Bakker, A.: Boundary crossing and boundary objects. *Rev. Educ. Res.* **81**(2), 132–169 (2011). <https://doi.org/10.3102/0034654311404435>
5. Allen, G.: Getting beyond form filling: the role of institutional governance in human research ethics. *J. Acad. Ethics* **6**(2), 105–116 (2008). <https://doi.org/10.1007/s10805-008-9057-9>
6. Attard-Frost, B., Andrés De los, R., and Deneille R. Walters. 2022. 'The Ethics of AI Business Practices: A Review of 47 AI Ethics Guidelines'. *AI and Ethics*, April. <https://doi.org/10.1007/s43681-022-00156-6>
7. Australian Government: Department of Industry, Science, Energy and Resources. 2019. 'Artificial Intelligence: Australia's Ethics Framework and Consultation'. 2019. <https://webarchive.nla.gov.au/awa/20200921003335/https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework/>.
8. Badampudi, Deepika, Farnaz Fotrousi, Bruno Cartaxo, and Muhammad Usman. 2022. 'Reporting Consent, Anonymity and Confidentiality Procedures Adopted in Empirical Studies Using Human Participants'. *E-Informatica Softw. Eng. J.* **16** (1): 220109. <https://doi.org/10.37190/e-Inf220109>.
9. Barke, R.: Balancing uncertain risks and benefits in human subjects research. *Sci. Technol. Hum. Values* **34**(3), 337–364 (2009). <https://doi.org/10.1177/0162243908328760>

10. Beauchemin, É., Côté, L.P., Drolet, M.-J., Williams-Jones, B.: Conceptualising ethical issues in the conduct of research: results from a critical and systematic literature review. *J. Acad. Ethics* **20**(3), 335–358 (2022). <https://doi.org/10.1007/s10805-021-09411-7>
11. Bernstein, M.S., Levi, M., Magnus, D., Rajala, B.A., Satz, D., Waeiss, C.: Ethics and society review: ethics reflection as a precondition to research funding. *Proc. Natl. Acad. Sci. U.S.A.* **118**(52), e2117261118 (2021). <https://doi.org/10.1073/pnas.2117261118>
12. Birkle, C., Pendlebury, D.A., Schnell, J., Adams, J.: Web of science as a data source for research on scientific and scholarly activity. *Quantitative Sci. Stud.* **1**(1), 363–376 (2020). [https://doi.org/10.1162/qss\\_a\\_00018](https://doi.org/10.1162/qss_a_00018)
13. Blackman, R. 2022. 'Why You Need an AI Ethics Committee'. *Harvard Business Review*, 1 July 2022. <https://hbr.org/2022/07/why-you-need-an-ai-ethics-committee>.
14. Bondarouk, T., and Huub, R. 2004. 'Discourse Analysis: Making Complex Methodology Simple'. In *ECIS 2004 Proceedings*. <https://ris.utwente.nl/ws/portafiles/porta/5405415/ECIS2004-1.pdf>.
15. Bosch, N., Say Chan, A., Davis, J. L., Gutiérrez, R., He, J., Karahalios, K., Koyejo, S. et al. 2022. 'Artificial Intelligence and Social Responsibility: The Roles of the University'. A white paper by University of Illinois Urbana-Champaign. <https://cra.org/cra/wp-content/uploads/sites/2/2022/11/Symposium-on-Artificial-Intelligence-and-Social-Responsibility-.pdf>.
16. Bowker, G.C., Star, L.S.: *Sorting things out: classification and its consequences*. MIT Press, Cambridge, MA (1999)
17. Brown, C., Spiro, J., Quinton, S.: The role of research ethics committees: friend or foe in educational research? an exploratory study. *Br. Edu. Res. J.* **46**(4), 747–769 (2020). <https://doi.org/10.1002/berj.3654>
18. Carniel, J., Hickey, A., Southey, K., Brömdal, A., Crowley-Cyr, L., Eacersall, D., Farmer, W., Gehrmann, R., Machin, T., Pillay Y. 2022. 'The ethics review and the humanities and social sciences: disciplinary distinctions in ethics review processes'. *Research Ethics*, December, 17470161221147202. <https://doi.org/10.1177/17470161221147202>.
19. Chi, Nicole, Emma Lurie, and Deirdre K. Mulligan. 2021. 'Reconfiguring Diversity and Inclusion for AI Ethics'. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 447–57. Virtual Event USA: ACM. <https://doi.org/10.1145/3461702.3462622>.
20. Coleman, C.H., Bouësseau, M.-C.: How do we know that research ethics committees are really working? The neglected role of outcomes assessment in research ethics review. *BMC Med. Ethics* **9**(1), 1–7 (2008). <https://doi.org/10.1186/1472-6939-9-6>
21. Cross, J.E., Pickering, K., Hickey, M.: Community-based participatory research, ethics, and institutional review boards: untying a gordian knot. *Crit. Sociol.* **41**(7–8), 1007–1026 (2015). <https://doi.org/10.1177/0896920513512696>
22. Cuellar, M.-F., Larsen, B., Lee, Y.S., Webb, M.: Does information about AI regulation change manager evaluation of ethical concerns and intent to adopt AI? *Journal of Law Economics & Organization*, April. (2022). <https://doi.org/10.1093/jleo/ewac004>
23. Delft University. 2022. 'Delft University of Technology HUMAN RESEARCH ETHICS COMPLETING THE HREC CHECKLIST (Version January 2022)'. 2022. [https://d2k0ddhflgrkl1.cloudfront.net/TUdelft/Over\\_TU\\_Delft/Strategie/Integriteitsbeleid/Research%20ethics/2\\_CHC-completing%20the%20HREC%20checklist\\_2022.pdf](https://d2k0ddhflgrkl1.cloudfront.net/TUdelft/Over_TU_Delft/Strategie/Integriteitsbeleid/Research%20ethics/2_CHC-completing%20the%20HREC%20checklist_2022.pdf).
24. D'ignazio, C., and L. F. Klein. 2020. *Data Feminism*. MIT press.
25. Doerr, M., Meeder, S.: Big health data research and group harm: the scope of IRB review. *Ethics Hum. Res.* **44**(4), 34–38 (2022). <https://doi.org/10.1002/eahr.500130>
26. Drolet, M.-J., Rose-Derouin, E., Leblanc, J.-C., Ruest, M., Williams-Jones, B.: Ethical issues in research: perceptions of researchers, research ethics board members and research ethics experts. *J. Acad. Ethics* August. (2022). <https://doi.org/10.1007/s10805-022-09455-3>
27. DuBois, J.M., Volpe, R.L., Rangel, E.K.: Hidden empirical research ethics: a review of three health journals from 2005 through 2006. *J. Empir. Res. Hum. Res. Ethics* **3**(3), 7–18 (2008). <https://doi.org/10.1525/jer.2008.3.3.7>
28. Eto, T. 2022. 'Conducting an Effective IRB Review of Artificial Intelligence Human Subjects Research (AI HSR)'. *Technology In Human Subjects Research*. <https://techinhsr.com/wp-content/uploads/2022/08/AI-HSR-WHITE-PAPER-TechInHSR-08.2022-1.pdf>.
29. Ferretti, A. 2021. 'Ethics and Governance of Big Data in Health Research and Digital Health Applications'. Doctoral Thesis, ETH Zurich. <https://doi.org/10.3929/ethz-b-000489154>.
30. Ferretti, A., Ienca, M., Sheehan, M., Blasimme, A., Dove, E.S., Farsides, B., Friesen, P., Kahn, J., Karlen, W., Kleist, P.: Ethics review of big data research: What should stay and what should be reformed? *BMC Med. Ethics* **22**(1), 1–13 (2021). <https://doi.org/10.1186/s12910-021-00616-4>
31. Ferretti, A., Ienca, M., Velarde, M.R., Hurst, S., Vayena, E.: The challenges of big data for research ethics committees: a qualitative swiss study. *J. Empir. Res. Hum. Res. Ethics* **17**(1–2), 129–143 (2022). <https://doi.org/10.1177/15562646211053538>
32. Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., Srikumar, M., 2020. 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI'. SSRN Scholarly Paper. Rochester, NY. <https://doi.org/10.2139/ssrn.3518482>.
33. Frauenberger, C., Rauhala, M., Fitzpatrick, G.: In-action ethics. *Interact. Comput.* **29**(2), 220–236 (2017). <https://doi.org/10.1093/iwc/iiw024>
34. Freeman, R., and Steve Sturdy. 2014. 'Introduction: Knowledge in Policy—Embodied, Inscribed, Enacted'. In *Knowledge in Policy*.
35. Friesen, P., Douglas-Jones, R., Marks, M., Pierce, R., Fletcher, K., Mishra, A., Lorimer, J., Véliz, C., Hallowell, N., Graham, M.: Governing AI-driven health research: are IRBs up to the task? *Ethics Hum. Res.* **43**(2), 35–42 (2021). <https://doi.org/10.1002/eahr.500085>
36. Gooding, P., Kariotis, T.: Ethics and law in research on algorithmic and data-driven technology in mental health care: scoping review. *Jmir Mental Health* **8**(6), e24668 (2021). <https://doi.org/10.2196/24668>
37. Goodyear-Smith, F., Lobb, B., Davies, G., Nachson, I., Seelau, S.M.: International variation in ethics committee requirements: comparisons across five westernised nations. *BMC Med. Ethics* **3**(1), 1–8 (2002). <https://doi.org/10.1186/1472-6939-3-2>
38. Guillemin, M., Gillam, L.: Ethics, reflexivity, and "ethically important moments" in research. *Qual. Inq.* **10**(2), 261–280 (2004). <https://doi.org/10.1177/1077800403262360>
39. Health Research Authority. 2022. 'Improving Our Review of Research Using AI and Data-Driven Technologies'. Health Research Authority. 2022. <https://www.hra.nhs.uk/planning-and-improving-research/research-planning/how-were-supporting-data-driven-technology/sddr/improving-our-review-research-using-ai-and-data-driven-technologies/>.
40. Hickey, A., Davis, S., Farmer, W., Dawidowicz, J., Moloney, C., Lamont-Mills, A., Carniel, J., et al.: Beyond criticism of ethics review boards: strategies for engaging research communities and enhancing ethical review processes. *J. Acad. Ethics* **20**(4), 549–567 (2022). <https://doi.org/10.1007/s10805-021-09430-4>
41. Hine, C.: Evaluating the prospects for university-based ethical governance in artificial intelligence and data-driven innovation. *Res. Ethics* **17**(4), 464–479 (2021). <https://doi.org/10.1177/17470161211022790>

42. Hutson, M. 2021. 'Who Should Stop Unethical A.I.?' *The New Yorker*, 15 February 2021. <https://www.newyorker.com/tech/annals-of-technology/who-should-stop-unethical-ai>.
43. 'IEEE Ethics In Action in Autonomous and Intelligent Systems | IEEE SA'. n.d. Ethics In Action | Ethically Aligned Design. Accessed 10 January 2023. <https://ethicsinaction.ieee.org/>.
44. Israel, M. 2015. 'Regulating Ethics'. In *Research Ethics and Integrity for Social Scientists: Beyond Regulatory Compliance*. 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd. <https://doi.org/10.4135/9781473910096>.
45. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. *Nat Mach Intell* **1**(9), 389–399 (2019). <https://doi.org/10.1038/s42256-019-0088-2>
46. Jordan, S. R. 2019. 'Designing an artificial intelligence research review committee'. *Future of Privacy Forum*.
47. Journal of Empirical Research on Human Ethics. n.d. 'Journal of Empirical Research on Human Ethics: Manuscript Preparation'. Accessed 13 January 2023. <https://journals.sagepub.com/pb-assets/cmscontent/JRE/JERPrep.pdf>.
48. Keyes, O., Hutson, J., Durbin, M. 2019. 'A Mulching proposal: analysing and improving an algorithmic system for turning the elderly into high-nutrient slurry'. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–11. Glasgow Scotland Uk: ACM. <https://doi.org/10.1145/3290607.3310433>.
49. Khan, A.A., Badshah, S., Liang, P., Waseem, M., Khan, B., Ahmad, A., Fahmideh, M., Niazi, M., Azeem Akbar, M. 2022. 'Ethics of AI: A Systematic Literature Review of Principles and Challenges'. In *The International Conference on Evaluation and Assessment in Software Engineering 2022*, 383–92. EASE '22. Gothenburg Sweden: ACM. <https://doi.org/10.1145/3530019.3531329>.
50. Kitto, K., Knight, S.: Practical ethics for building learning analytics. *Br. J. Edu. Technol.* **50**(6), 2855–2870 (2019). <https://doi.org/10.1111/bjet.12868>
51. Knight, S., Shibani, A., Shum, S.B.: A reflective design case of practical ethics in learning analytics. *Br. J. Edu. Technol.* (2023). <https://doi.org/10.1111/bjet.13323>
52. Leetaru, K. 2016. 'Are Research Ethics Obsolete In The Era Of Big Data?' *Forbes*. 2016. <https://www.forbes.com/sites/kalevleetaru/2016/06/17/are-research-ethics-obsolete-in-the-era-of-big-data/>.
53. Leetaru, K. 2017. 'AI "Gaydar" And How The Future Of AI Will Be Exempt From Ethical Review'. *Forbes*. 2017. <https://www.forbes.com/sites/kalevleetaru/2017/09/16/ai-gaydar-and-how-the-future-of-ai-will-be-exempt-from-ethical-review/>.
54. Lu, Q., Zhu, L., Xu, X., Whittle, J., Zowghi, D., Jacquet, A. 2022. 'Responsible AI Pattern Catalogue: A Multivocal Literature Review'. *arXiv*. <https://doi.org/10.48550/arXiv.2209.04963>.
55. Macdonald, H. 2014. 'Transnational Excursions: The Ethics of Northern Anthropological Investigations Going South'. *Ethical Quandaries in Social Research*, December. [https://www.academia.edu/11524903/Transnational\\_excursions\\_The\\_ethics\\_of\\_northern\\_anthropological\\_investigations\\_going\\_south](https://www.academia.edu/11524903/Transnational_excursions_The_ethics_of_northern_anthropological_investigations_going_south).
56. Martín-Martín, A., Thelwall, M., Orduna-Malea, E., López-Cózar, E.D.: Google scholar, microsoft academic, scopus, dimensions, web of science, and open citations' COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics* **126**(1), 871–906 (2021). <https://doi.org/10.1007/s11192-020-03690-4>
57. McCarthy, J., Minsky, M.L., Rochester, N., Shannon, C.E.: A proposal for the dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Mag.* **27**(4), 12–12 (1955)
58. Metcalf, J., Crawford, K.: Where are human subjects in big data research? The emerging ethics divide. *Big Data Soc.* **3**(1), 2053951716650211 (2016). <https://doi.org/10.1177/2053951716650211>
59. Miller, C., Coldicutt, R. 2019. 'People, Power and Technology: The Tech Workers' View'. *doteveryone*. <https://doteveryone.org.uk/report/workersview/>.
60. Mittelstadt, B.: Principles alone cannot guarantee ethical AI. *Nat. Mach. Intell.* **1**(11), 501–507 (2019). <https://doi.org/10.1038/s42256-019-0114-4>
61. Molina, J.L., Borgatti, S.P.: Moral Bureaucracies and Social Network Research. *Soc. Netw.* **67**, 13–19 (2021). <https://doi.org/10.1016/j.socnet.2019.11.001>
62. Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., Floridi, L.: Operationalising AI Ethics: Barriers, Enablers and next Steps. *AI & Soc.* November. (2021). <https://doi.org/10.1007/s00146-021-01308-8>
63. Morton, J.: "Text-Work" in research ethics review: the significance of documents in and beyond committee meetings. *Account. Res.* **25**(7–8), 387–403 (2018). <https://doi.org/10.1080/08989621.2018.1537790>
64. Munteanu, C., Molyneaux, H., Moncur, W., Romero, M., O'Donnell, S., Vines, J. 2015. 'Situational Ethics: Re-Thinking Approaches to Formal Ethics Requirements for Human-Computer Interaction'. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 105–14. CHI '15. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/2702123.2702481>.
65. National Statement. 2018. 'National Statement on Ethical Conduct in Human Research'. National Health and Medical Research Council, the Australian Research Council and Universities Australia. <https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018>.
66. Office for Human Research Protections (OHRP). 1978. 'Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research'. <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html>.
67. Pater, J., Fiesler, C., Zimmer, M. 2022. No Humans Here: Ethical Speculation on Public Data, Unintended Consequences, and the Limits of Institutional Review. *Proceedings of the ACM on Human-Computer Interaction* **6** (GROUP): 38:1–38:13. <https://doi.org/10.1145/3492857>.
68. Petrozzino, C.: Who pays for ethical debt in AI? *AI and Ethics* **1**(3), 205–208 (2021). <https://doi.org/10.1007/s43681-020-00030-3>
69. Pickering, B.: Trust, but verify: informed consent, AI technologies, and public health emergencies. *Future Internet* **13**(5), 132 (2021). <https://doi.org/10.3390/fi13050132>
70. Pieper, I., Thomson, C.J.H.: Contextualising merit and integrity within human research. *Monash Bioeth. Rev.* **29**(4), 39–48 (2011). <https://doi.org/10.1007/BF03351329>
71. Rapley, T. 2007. *Doing Conversation, Discourse and Document Analysis*. 1 Oliver's Yard, 55 City Road, London England EC1Y 1SP United Kingdom: SAGE Publications Ltd. <https://doi.org/10.4135/9781849208901>.
72. Resseguier, A., Rodrigues, R., Santiago, N. 2021. Ethics as Attention to Context: Recommendations for AI Ethics Annex to D5.4: Multi-Stakeholder Strategy and Tools for Ethical AI and Robotics. *Sienna Project*. [https://www.sienna-project.eu/digitalAssets/915/c\\_915542-1\\_1-k\\_ethics-as-attention\\_sienna\\_jan-2021.pdf](https://www.sienna-project.eu/digitalAssets/915/c_915542-1_1-k_ethics-as-attention_sienna_jan-2021.pdf).
73. Samuel, G., Derrick, G., van Leeuwen, T.: The ethics ecosystem: personal ethics, network governance and regulating actors governing the use of social media research data. *Minerva* **57**(3), 317–343 (2019). <https://doi.org/10.1007/s11024-019-09368-3>
74. Sandler, R., and Basl, J. 2019. Building data and ai ethics committees. *Accenture and Ethics Institute at Northeastern University*. [https://www.accenture.com/\\_acnmedia/pdf-107/accenture-ai-and-data-ethics-committee-report-11.pdf](https://www.accenture.com/_acnmedia/pdf-107/accenture-ai-and-data-ethics-committee-report-11.pdf).

75. Schiff, D., Borenstein, J., Biddle, J., Laas, K.: AI ethics in the public, private, and NGO sectors: a review of a global document collection. *IEEE Trans. Technol. Soc.* **2**(1), 31–42 (2021). <https://doi.org/10.1109/TTS.2021.3052127>
76. Shevlane, T., and Dafoe, A. 2020. The Offense-Defense Balance of Scientific Knowledge: Does Publishing AI Research Reduce Misuse? In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 173–79. AIES '20. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3375627.3375815>.
77. Singh, V.K., Singh, P., Karmakar, M., Leta, J., Mayr, P.: The journal coverage of web of science, scopus and dimensions: a comparative analysis. *Scientometrics* **126**(6), 5113–5142 (2021). <https://doi.org/10.1007/s11192-021-03948-5>
78. Star, S.L., Griesemer, J.R.: Institutional ecology, translations' and boundary objects: amateurs and professionals in Berkeley's museum of vertebrate zoology, 1907–39. *Soc. Stud. Sci.* **19**(3), 387–420 (1989). <https://doi.org/10.1177/030631289019003001>
79. Steerling, E., Houston, R., Gietzen, L.J., Ogilvie, S.J., de Ruiter, H.-P., Nygren, J.M.: Examining how ethics in relation to health technology is described in the research literature: scoping review. *Interactive Journal of Medical Research* **11**(2), e38745 (2022). <https://doi.org/10.2196/38745>
80. Swierstra, T.: Identifying the normative challenges posed by technology's "soft" impacts. *Etikk i Praksis - Nordic Journal of Applied Ethics*, **1**(May), 5–20 (2015). <https://doi.org/10.5324/EIP.V9I1.1838>
81. Tambornino, L., Lanzerath, D., Rodrigues, R., Wright, D. 2019. SIENNA D4.3: Survey of REC Approaches and codes for artificial intelligence & robotics, August. <https://doi.org/10.5281/zenodo.4067990>.
82. The Department of Health and Human Services. n.d. *45 CFR Part 46 (2018–07–19)—Protection of Human Subjects. Protection of Human Subjects*. Vol. Title 45. Accessed 10 January 2023. <https://www.ecfr.gov/on/2018-07-19/title-45/subtitle-A/subchapter-A/part-46>.
83. Tummons, J. 2022. The Many Worlds of Ethics: Proposing a Latourian Investigation of the Work of Research Ethics in Ethnographies of Education. In *Ethics, Ethnography and Education*, edited by Lisa Russell, Ruth Barley, and Jonathan Tummons, 19:11–28. Studies in Educational Ethnography. Emerald Publishing Limited. <https://doi.org/10.1108/S1529-210X20220000019002>.
84. UKRI. 2022. 'Embedding Ethics in Artificial Intelligence Research'. 2022. <https://www.ukri.org/about-us/how-we-are-doing/research-outcomes-and-impact/ahrc/embedding-ethics-in-artificial-intelligence-research/>.
85. UTS.: AHRC Human Rights and Technology Issues Paper: UTS Response and Submission. University of Technology Sydney (2018). [https://www.uts.edu.au/sites/default/files/2018-12/Human%20Rights%20%26%20Technology%20Issues%20Paper\\_UTS%20submission.pdf](https://www.uts.edu.au/sites/default/files/2018-12/Human%20Rights%20%26%20Technology%20Issues%20Paper_UTS%20submission.pdf)
86. Vitak, J., Proferes, N., Shilton, K., Ashktorab, Z.: Ethics regulation in social computing research: examining the role of institutional review boards. *J. Empir. Res. Hum. Res. Ethics* **12**(5), 372–382 (2017). <https://doi.org/10.1177/1556264617725200>
87. Vitak, J., Shilton, K., Ashktorab, Z. 2016. Beyond the Belmont Principles: Ethical Challenges, Practices, and Beliefs in the Online Data Research Community. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 941–53. CSCW '16. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/2818048.2820078>.
88. Weinbaum, C., Landree, E., Blumenthal, M., Piquado, T., Gutierrez, C.: Ethics in scientific research: an examination of ethical principles and emerging topics. RAND Corporation (2019). <https://doi.org/10.7249/RR2912>
89. Whittlestone, J., Nyrup, R., Alexandrova, A., Cave, S. 2019. The role and limits of principles in ai ethics: towards a focus on tensions. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 195–200. AIES '19. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3306618.3314289>.
90. Zhang, J.J.: Research ethics and ethical research: some observations from the global south. *J. Geogr. High. Educ.* **41**(1), 147–154 (2017). <https://doi.org/10.1080/03098265.2016.1241985>
91. Zhou, J., and Chen, F. 2022. 'AI Ethics: From Principles to Practice'. *AI & SOCIETY*, 1–11.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.