Evaluating the prospects for university-based ethical governance in artificial intelligence and data-driven innovation

Christine Hine
University of Surrey, UK

Abstract
There has been considerable debate around the ethical issues raised by data-driven technologies such as artificial intelligence. Ethical principles for the field have focused on the need to ensure that such technologies are used for good rather than harm, that they enshrine principles of social justice and fairness, that they protect privacy, respect human autonomy and are open to scrutiny. While development of such principles is well advanced, there is as yet little consensus on the mechanisms appropriate for ethical governance in this field. This paper examines the prospects for the university ethics committee to undertake effective review of research conducted on data-driven technologies in the university context. Challenges identified include: the relatively narrow focus of university-based ethical review on the human subjects research process and lack of capacity to anticipate downstream impacts; the difficulties of accommodating the complex interplay of academic and commercial interests in the field; and the need to ensure appropriate expertise from both specialists and lay voices. Overall, the challenges identified sharpen appreciation of the need to encourage a joined-up and effective system of ethical governance that fosters an ethical culture rather than replacing ethical reflection with bureaucracy.

Corresponding author:
Christine Hine, Department of Sociology, University of Surrey, Guildford, Surrey GU2 7NX, UK.
Email: c.hine@surrey.ac.uk

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Introduction

Recent innovations in computational technologies that supplement or even replace human intelligence are widely acknowledged to raise profound and challenging ethical questions. This field of data-driven technologies encompasses an array of developments including large scale data analytics, intelligent automation, and robotics (Corea, 2019). These technologies employ large-scale data-processing that often entails forms of machine learning or artificial intelligence (AI). Rather than needing to be instructed step-by-step, machine learning technologies are built to discover patterns, to make predictions and to adapt their outcomes on the basis of experience. Designers of such technologies will have an end goal in mind and data resources to draw upon to achieve that goal, but they will not necessarily be able to define in advance the steps by which the goal is to be achieved nor to clearly explain after the event precisely how the result has been arrived at. This array of technologies prompts considerable hopes for enhanced decision-making, informed planning, and responsive services, but these hopes are accompanied by risks such as privacy intrusion, loss of autonomy, bias, and opaque decision-making (Mittelstadt et al., 2016). These risks have become the focus of considerable scrutiny as the data-driven technologies have increasingly become a practical reality (High Level Expert Group on Artificial Intelligence, 2019). In response, ethical principles for the field are emerging, focused around requirements for algorithmic transparency, the need for human oversight, and the need to avoid social biases (Floridi and Cowls, 2019). It is still unclear, however, how these broad ethical principles might be incorporated into effective and authoritative systems of research governance and regulation, whether that be in commercial settings or in academic research contexts.

This paper explores prospects for the university ethics committee, as an established site of research governance, to play a role in the ethical governance of innovations in data-driven technologies. This argument is developed with particular reference to the forms of ethical review enacted in universities in the UK for non-healthcare research. The paper will argue that while university ethics committees are well-established within the academic research community, there are challenges that may prevent them from offering an effective mechanism to ensure that research into data-driven technologies within universities fully conforms to the emerging ethical principles in this field. The first section of the paper outlines the current debate around ethical challenges in data-driven innovation, exploring the
emerging consensus around the nature of the ethical problems encountered in this field and highlighting the current lack of agreement on the arrangements needed for effective governance of ethical review across both commercial and non-commercial settings. The next section then turns to university-based ethical governance and explores three areas where particular challenges emerge in the application of this model to data-driven technologies. The first set of challenges focuses on the tendency of ethical review to concentrate on the human participants in the research process as narrowly defined, rather than subsequent implementation of the research outcomes. Further challenges arise in relation to the remit of university-based ethical review, where ethical governance often becomes entangled with management of institutional reputation and where sometimes unhelpful lines between research and consultancy are drawn for governance purposes. A final set of challenges focuses on the nature of the expertise required for effective ethical scrutiny to take place. In conclusion it is argued that universities do indeed need to rise to the challenge of ensuring that ethical governance reaches all aspects of the research work undertaken under their auspices and that they will need to expand their expertise in this regard to accommodate research involving data-driven technologies. In doing so, however, it is important not simply to shoehorn data-driven technologies into existing systems of ethical governance, since to do so may give an unwarranted illusion of effective scrutiny having taken place. An open debate around the response of universities to ethical governance beyond the humans who are active participants in the research process will be needed. If alternative systems of governance for data-driven technologies beyond the university develop then there will be new complexities to manage in how this new external system interacts with the existing internal research governance that universities offer.

**Ethical challenges in data-driven innovation**

In recent years there has been an explosion in discussion around the ethics of data-driven technologies such as artificial intelligence (Floridi and Cowls, 2019). While concerns about possible ethical challenges raised by artificial intelligence have been live since the emergence of the field itself, the debate has become significantly more active in recent years with the increasing mainstreaming of these technologies. Increases in computing power and the availability of the raw data to fuel machine learning have fostered a flourishing domain of research and commercial application and along with this, a flourishing domain of commentary and concern about potential undesirable consequences. When decision-making is automated, it is feared that ethical responsibilities may be side-stepped and forms of discrimination or ill-treatment that would be unacceptable from one human being to another may be tolerated or difficult to resist when perpetrated by machines. In order to alleviate such prospects, a need to articulate standards for the field and set
boundaries for acceptable practice has been proposed repeatedly, and indeed Floridi and Cowls (2019) describe a “principle proliferation” occurring, as multiple authors and institutions seek to identify the ethical principles that should prevail. Through comparison of six high profile documents Floridi and Cowls (2019) identify five over-arching ethical principles that arise repeatedly or subsume more granular principles: beneficence, non-maleficence, autonomy, justice, and explicability. There is considerable consensus that we want data-driven technologies to work for good and not to do harm, that they should enshrine human autonomy and not machine autonomy, that they operate in the interests of justice rather than discrimination and that all of the principles are achieved through technologies that are open to interrogation because they are intelligible and can be held to account. The authors point out that these principles map closely onto the established principles for biomedical ethics, with the addition of the requirement of explicability or transparency to reflect the notoriously opaque nature of machine-led decision-making.

While there might be some convergence in terms of over-arching principles, there is as yet little progress toward consensus on how the principles are to be enshrined in practice (Morley et al., 2019). Cath (2018) explained that at the time of writing there was considerable international debate about the need for governance of artificial intelligence but no convergence on the mechanisms to achieve this, and it appears that the situation has changed little in the intervening time. Jobin et al. (2019) report on a global landscape across which there is still much disagreement about the priorities and mechanisms for assurance of ethical AI. In particular, a lack of clarity about the nature of any new institutional frameworks is often apparent: the EU guidelines for trustworthy AI (2019), for example, ask for their considerations to be incorporated into existing practices but leave in doubt how this is to be achieved. Nonetheless the impetus is there across many different national contexts and application domains: the October 2020 “State of AI Ethics” report from the Montreal AI Ethics Institute (2020) notes a “rising tide of demands to move to solutions” (p. 11).

Much of the debate about solutions revolves around the governance mechanisms that might be implemented to ensure compliance with ethical standards. As Mittelstadt (2019) explains, there is no existing professional infrastructure for embedding principles of AI ethics and a long way to go before such an infrastructure can be developed and implemented on a sufficiently wide scale. Given the dominance of large companies in this domain there is cause for pessimism about the prospects for effective self-regulation based on ethical values (Black and Murray, 2019). Even though there are developments across the industry focused on in-house solutions for ethical scrutiny, it is often problematic to rely on these without concerns about conflict of interests being raised. Within some sectors and particularly outside commercial contexts there have been some credible targeted
efforts to introduce both guidance and the mechanisms to enforce that guidance: the Alan Turing Institute in the UK, for example, has recently published a guide for the public sector on how to implement AI in responsible fashion (Leslie, 2019). This guide lays out a set of steps focused both on the process of the innovation and the ultimate product, enrolling the views of a broad array of stakeholders and extending consultation beyond the narrow confines of a team of developers.

The broad approach of recourse to a peer community for ethical reviewing in the interests of stakeholders is a familiar one from the arena of research ethics. This move takes the responsibility for decision-making away from the ethical conscience of an individual researcher or developer and attends to what Edwards (2009) describe as the “morally problematic” position of the researcher aiming to maintain ethical practice in the face of potentially competing self-interest. More negatively, Haggerty (2004: 393) describes this form of governance as an “institutionalized distrust” that cannot bear to leave decisions to the individual. How far the shift from the individual researcher’s conscience to the resources of a peer community or institutional bureaucracy is justified and where the balance should be struck is beyond the scope of this paper. Suffice to say that there is a considerable mood within the debate on ethics of data-driven technologies that the individual researcher or team of developers should not have sole responsibility for ethical decision-making and this extends to researchers in an academic context as much as those working in commercial settings. There is indeed some evidence that it would be dangerous to leave researchers to self-regulate: Fell et al. (2021) find some worrying ethical oversights among the researchers active in the field of data-driven technologies that they interview. An ethical review of data-driven technologies that includes a community of peers beyond the individual researcher appears to be a widely favored solution even in the research settings where direct commercial conflicts of interest are not a concern.

The need for ethical governance of data-driven technologies is thus widely accepted but there is considerable doubt still remaining on the mechanisms through which this is to be achieved. Within the university setting there are some features of this debate that suggest that the university ethics committee might be well placed to act. Many of the principles proposed for ethical data-driven technologies are very familiar from the biomedical ethics that form the core of the current research ethics infrastructure. The mechanisms of peer community review and attention to stakeholder concerns proposed for ethical governance in data-driven technologies are also familiar ones for the university ethical review context. However, there are also some considerable mismatches between the university ethics committee as a regulatory infrastructure and the needs for ethical governance of data-driven technologies. In the following section these issues will be addressed under three headings: the focus on human participants; the remit of
university ethics committees to focus on research; and the nature of the expertise brought to bear on review.

**University-based ethical governance and data-driven innovation**

*Human participants, data subjects, and ethical governance*

The key mechanism for governance of research involving human participants within universities in the UK is the ethics committee (the equivalent body in the US would be the Institutional Review Board or IRB). A full review of the emergence of university-based ethical review and the responses to those developments is outside the scope of this paper (see Hedgecoe (2009)). Suffice to say that the current system has arisen in a somewhat piecemeal fashion and there is not full consensus that the forms of ethical review that we have arrived at meet the needs of all research communities: some question the applicability of practices derived from biomedical ethics to the social sciences (Schrag, 2011) while others are concerned about the bureaucratization of ethical review into a set of procedures enacted by committee (Haggerty, 2004; van den Hoonoord and Hamilton, 2016). Such concerns remain for the most part in the background for this paper, which begins from the current context within which ethical review committees have become embedded in UK universities as a means to meet expectations around ethical conduct and that as such they form an important feature in the landscape of research governance for any university-based researcher whether or not the researcher is warmly disposed to their activities.

As Edwards (2009, 2010a, 2010b, 2010c, 2011) describes in a series of papers on the role and remit of the ethics committee,\(^2\) the primary focus of review is on the human participants in the research process itself. The committee takes a protective view of these human participants, balanced against the future potential benefits of the research. The focus of review homes in most closely on the well-being of the human participants in the research and tends to narrow in on issues of informed consent and the limits of consent (Edwards, 2010c). Whilst the committee is required to balance the risks of the research against the potential gain to society (Edwards, 2010b), in practice understanding of potential gain is a somewhat speculative and uncertain business. Hence, the focus of ethical review can often come to concentrate on the research process and in particular the impact on the immediate participants in that process.

In its focus on the research process, much of the practice of university ethics committees is founded on the basis forged for biomedical ethics. In a biomedical context, while the same principles of beneficence, non-maleficence, autonomy, and justice are held to apply across research and everyday practice (Beauchamp
and Childress, 2019), the monitoring of these principles happens in quite different ways between research and practice. In everyday medical practice the application of ethical principles is largely reliant on the judgment of healthcare professionals who are offering established treatments in the best interests of patients (Gillon, 1994). By contrast, medical research is subject to an additional level of ethical scrutiny through formal review considering whether it is appropriate to expose participants to the potential for additional levels of risk that research entails. Modeled on this scenario, university ethical review might be expected to examine data-driven research from the perspective of participants but not necessarily to have the capacity to look downstream at the products of such research when they find their way into implementation. In medical contexts, ethical review beyond research focuses on the suitability of new treatments, and there are calls for regulators, policymakers, and developers to be mindful of the need for thorough ethical assessment of AI (Morley and Floridi, 2020; Morley et al., 2020). Within healthcare there is a complex patchwork of responsibilities regarding ethical compliance for AI across clinical settings and direct-to-consumer products (Nebeker et al., 2019). There is as yet even less clarity on the processes of ethical review for the products of university research on data-driven technologies outside the healthcare domain as they might move into an implementation phase, whether through spin-off company or other forms of technology transfer.

A focus on direct participants may itself become challenging in data-driven research: Metcalf and Crawford (2016) note that it is difficult for a university ethics committee comprehensively to consider potential harms to humans from data-driven research, since harms may be diffuse, occur well downstream of the original research context and affect those who were not involved or indeed identifiable at all within the narrowly defined scope of the research project that would receive ethical review. A comprehensive review of data-driven research shifts away from the protection of immediately identifiable humans as participants to include humans who are rendered as data subjects and onto the review of algorithms, training datasets, arrangements for regulations and governance, and implementation of human oversight. There may also be a need to expand the definition of human participants, since the research may involve what appears to be simply secondary data or textual analysis but have downstream consequences for identifiable people.

The focus of ethical review on the immediate research process contrasts with the more recent movement to promote responsible research and innovation or RRI (Von Schomberg, 2013) that builds in a stronger focus on looking downstream to implementation and aims to enhance deliberation at the stage of determining research directions and priorities. RRI expands the perspective beyond the individual research protocol that tends to be the focus of ethical review and builds in an explicit orientation toward the future:
Approaches to responsible innovation extend the governance discussion to encompass questions of uncertainty (in its multiple forms), purposes, motivations, social and political constitutions, trajectories and directions of innovation. (Stilgoe et al., 2013: 1570)

Stilgoe et al. (2013) describe four key aspects of responsible research and innovation: anticipation; inclusion; reflexivity (particularly in relation to the limits of one’s own knowledge); and responsiveness. It would be difficult for a university-based system of ethical review meaningfully to develop such an evaluation on the researchers’ behalf without this becoming purely speculative. It seems more reasonable that ethics committees should expect researchers to come to review having already conducted a broad ranging evaluation of their proposed direction of travel, including the views of funders, user constituencies, and other stakeholders as proponents of RRI suggest and that the outcomes of this process should be presented to the committee. In the field of data-driven technologies it is important that such an evaluation should draw on the emerging principles of ethical AI and keep in touch with current debate on potential harms and benefits.

The university ethics committee currently provides a peer-based scrutiny of proposed research largely from the perspective of its impacts on the immediately identifiable participants in the research. This may indeed be important for research on data-driven technologies, where humans are providers of the data that feeds machine learning or where humans are the recipients of interventions occasioned by as yet untried predictive models derived from machine learning. People impacted in this way by data-driven research would quite reasonably expect that their involvement be underpinned by a degree of ethical scrutiny beyond the individual researchers and in proportion to the degree of risk involved and that they might reasonably be treated as active participants in the research rather than passive data subjects. However, taken alone this falls far short of the kind of process-based governance that Leslie (2019) recommends for ethical AI in the public sector that encompasses an array of requirements for stakeholder consultation, diversity impact, safety assessment, and transparency statement. This process-based approach also allows for issues to arise throughout the course of a project rather than confining attention to those issues that can be anticipated and addressed in advance as is often the case for university ethical review. A similar concept is expressed by Dove (2020) in proposals for ethics committees to see themselves as engaged in a regulatory stewardship that takes a deliberative approach to achieving ethical research with applicants and accepts a need for ongoing engagement throughout the research process rather than one-off review. Such an ongoing approach to ethical assurance may be particularly important given the contingent and emergent nature of outcomes in machine learning. Unless we extend ethical review to include a wider array of issues downstream into the potential impacts of the products of the research, in line with Leslie’s (2019) recommendations for the
public sector approach to AI and with the broader agenda of RRI, it is unlikely that university-based ethical review would be in a position to encompass some of the key principles outlined as key for ethical AI. Ultimate implications for beneficence and maleficence, for example, might well only be reasonably definable in relation to ultimate contexts of use beyond the confines of immediate research scenarios.

The remit of ethical review in universities

It might seem obvious that university ethics committees should confine themselves to matters strictly related to ethics, but behind this straightforward statement considerable controversy plays out concerning the appropriate object of scrutiny. University ethical review does not in practice cover all of the activities carried out by members of university staff. As outlined above, the focus on research ethics often narrows down the remit of scrutiny in practice to the time-bounded process of active research involving human participants. There are also limits that are mandated by the requirement for other institutions to conduct reviews, such as the requirement for review by NHS research ethics committees where NHS patients are involved. Here NHS ethical review takes precedence and obviates the need for university review. In addition to this narrowing of the remit of the university ethics committee, a number of other issues act either to limit or to extend the remit of the ethics committee within a university such that not all research comes under their purview and not all that they consider falls under a strict definition of research ethics. The nature of these limits and extensions poses some challenges for the capacity of university ethics committees to accommodate an effective ethical review of research on data-driven technologies. These challenges focus on three distinct issues: the focus of review on research rather than consultancy; the tendency of review to include issues beyond ethics; and the counter-productive effect on ethical thinking of an intensification of ethical review.

Possibly the most challenging exclusion from university ethical review in the case of data-driven technologies is the focus of review on academic research rather than consultancy. Given the extent of commercial involvement in university research in this field and the tendency for consultancy work to be explicitly excluded from formal processes of ethical review there is potential for significant ethical issues with technologies developed by university researchers to miss out on peer review. A high-profile example concerns the case of Cambridge Analytica, the company accused of making inappropriate use of Facebook users’ personal data in political campaigning. In written evidence to a Parliamentary committee inquiry on disinformation and fake news, Aleksander Kogan (a researcher at the University of Cambridge who played a key role in providing the dataset used by Cambridge Analytica) stated that “as with all of my research, the research that
used the app was first ethics approved by the University” (Digital Culture Media and Sport Committee, 2019). A statement by University of Cambridge (2018) about the case developed a more nuanced position pointing out a distinction between academic research that was subject to review and various commercial associations that a researcher might develop in a personal capacity. The statement also highlighted that an application to the university for ethical approval to use data collected within the commercial context for academic research purposes had been rejected by them. The case makes apparent the complexity attendant on distinctions between academic research and commercial activity and the potential for this to operate against joined-up ethical review. Effective ethical review of research into data-driven technologies will require equivalent standards to be in operation for both academic and commercial research.

Ethics committees do not always confine themselves to ethics. Hedgecoe (2016) aligns the growth in university ethics committees with the extension of a particular form of managerial practice into universities and notes that over time ethical review becomes increasingly a tool for managing risks to institutional reputation rather than focusing on strictly ethical issues. In practice it is difficult for committee members with a commitment to their institutions to exclude from their ethical deliberations considerations on how a particular research topic might impact on the public standing of the university. Committees are thus often accused of being excessively risk averse for reasons beyond the strictly ethical. Such concerns might well pervade research into data-driven technologies where there is a potential to impinge on sensitive or politically volatile areas concerning automation and replacement of human labor or decision-making. Researchers in data-driven technologies might reasonably ask whether there are risks that their research might also come to be constrained by reputational rather than ethical concerns. Submitting for ethical review also has the tendency to expose researchers to further forms of scrutiny beyond ethics and reputational connotations. In recent years ethical review within many British universities has become embedded within a wider framing of research integrity and governance, following on from the publication of the Universities UK (2012) “concordat to support research integrity.” In many cases arrangements for ethical review will now be overseen by research integrity and governance offices that also take responsibility for oversight of compliance with other aspects of research regulations such as data protection, health and safety, and security-sensitivity (Universities UK, 2016). It is indeed important for research into data-driven technologies to submit to the requirements of research integrity and governance but the coupling of these forms of additional scrutiny with ethical review may present both practical and principled difficulties. In practice, the risk assessments and review of data protection raised by data-driven technologies are likely to be highly technical documents requiring specialist scrutiny. In principle, it can be important for ethics to be decoupled from other aspects of
research integrity and governance in order for the ethical issues to be fully appreciated.

Against the tendency for important areas of work to be excluded from review by university ethics committees thus sits an opposing tendency for their remit to extend and for ethical concerns to become dominated by bureaucracy. As a bureaucratic system, university ethical review is prone both to extension into new areas and intensification: Haggerty (2004) identifies a problem of “ethics creep” as the mandate of ethics committees extends to encompass more activities and those activities deemed within scope are regulated more intensely. The attendant problem, Haggerty (2004) argues, is that ever more formalized ethical review may be counter-productive if it shifts researchers’ attention away from actually acting ethically and rule-based compliance takes over from more fundamental attention to doing the right thing. This argument was made in the specific context of ethical review extending into social science, but we might consider whether the arguments have any relevance for the situation of data-driven technologies. Arguably, the social sciences did have a more established culture of ethical reflection before the formalization of ethical review than is currently the case for data-driven technologies, although Jennings (2012) suggests that even in the social sciences there are significant shortfalls in the ability of researchers to self-reflect sufficiently on ethical concerns. In the case of research into data-driven technologies we may have an opportunity to build a culture of self-reflection on ethics and a system of ethical governance simultaneously and hence to take care that the latter does not over-ride the former. Concerns about bureaucratization are not, in themselves, reasons to resist the extension of ethical review by committee into data-driven technologies, but there is a case for thinking very carefully about the nature of the work to be reviewed and the boundaries between research and consultancy, ethics, and reputation that are enacted. There is on the one hand every possibility that the scope of university ethical review could be extended again to include a wider array of data-driven research, whilst on the other hand there is potential for such extension to work either for or against the reinforcement of a culture of ethical practice in the field, depending on how it is implemented.

Lay and expert input in ethical governance

The final set of issues to consider is what forms of expertise the ethics committee can offer to the ethical review of data-driven technologies. The concept of appropriate expertise is a complex one for ethics committees, since the review process is explicitly designed to include voices beyond the researcher who might be deemed the one with most expertise on the project itself. While ethics committees need training in the fundamentals of ethical review, they are not necessarily either subject specialists in the material that they review, nor are they professional ethicists. Their
expertise is, rather, in the process of evaluation of the researcher’s self-assessment and the collective identification of the potential risks attendant on research directly involving humans (Edwards, 2011). Working within that model of expertise it is common for university ethics committees to include lay members from beyond the research community. The Economic and Social Research Council, for example, recommends inclusion on ethics committees of members who are users of health, education, or social services where these are the focus of the research (Economic and Social Research Council, n.d.). Lay input to ethical review offers the prospect of expanding the committee’s appreciation of risks, benefits, and harms (Edwards, 2009). Drawing on this array of voices aims at informing a balanced, inclusive assessment that does not favor the interests or perspective of the research community above those of wider society. Within this framing, lay members are being invoked as experts in their own condition and their ethical expertise is an everyday competence, albeit one that has to be balanced out by the committee through processes of deliberation (Edwards, 2011). Similarly, one of the key dimensions of responsible research and innovation is the broadening out of decision-making on new technologies, enabling a wider array of sectors of society to have a legitimate say on the direction of research and innovation (Owen et al., 2012).

Against this idealized model, there are, however, frequent concerns expressed about the extent to which ethics committees in practice actually display an appropriate degree of expertise such that they are able to reach sensible or consistent decisions (Schrag, 2011). Such complaints pose a direct challenge to ethics committees’ claims to hold sufficient expertise for adequate decision-making. There is therefore a live question on how far a committee should need to display domain-specific expertise, and how far the committee should handle areas where the research is highly technical and difficult for a non-specialist to understand. This may well be an issue for many areas of data-driven research where explicability is often in question. Whilst holding to the model of lay expertise as relevant and valuable, it may be important for ethics committees to draw on specialists to advise them in such areas. As Rawbone (2010) describes, where committees find themselves lacking in the expertise to identify risks and benefits they can draw on specialist advice and indeed specialist committees may convene to review research, for example at the level of an individual university faculty rather than the university as a whole, in order to balance the requirement for an inclusive community with the specialist expertise to interrogate the details of the research. Any system of ethical review for data-driven technologies in universities will need to consider the question of where appropriate expertise will be drawn from and the extent to which specialist advice or sub-committees may be required. It is important, however, that the recourse to specialists does not neglect the need for independent decision-making or preclude the inclusion of diverse lay voices as experts in their own condition.
Recommendations

There is clearly a need to develop an effective “practical machinery” for ethical review (Hedgecoe, 2009) in relation to the research on data-driven technologies being conducted within universities. Jobin et al. (2019) note that research ethics bodies such as university ethics committees will increasingly be called upon to undertake reviews of projects that involve AI. As outlined above, there is a need to enhance the ability of the university ethics committee to anticipate potential consequences beyond the immediate research process, to focus its remit appropriately to attend to the complex commercial and academic entwining of data-driven research, and to draw effectively on both lay members and specialists in order to build an appropriate set of expertise. A set of practical recommendations to take in order to move toward this set of requirements is as follows:

- undertake a program of ongoing training both for ethics committee members and across the array of disciplines within the university to promote awareness of the potential ethical pitfalls of data-driven research and the need for ethical review in this field
- review the application of research governance arrangements to consultancy projects with a view to ensuring that commercially-sponsored projects receive an appropriate level of ethical scrutiny
- formalize a requirement for researchers to undertake a thorough stakeholder consultation and impact assessment in preparation for ethical review
- consider building an ongoing dialog into ethical governance arrangements to offer a forum for addressing issues unanticipated by pre-emptive review
- formalize expectations around specialist technical input and lay or stakeholder consultation

If universities are clear that they cannot provide a comprehensive and technically astute ethical review for data-driven research that takes full account of stakeholder views and potential impacts within their existing framework of ethics committees, then this must be made explicit. One alternative would be to require researchers to come to university ethical review with a transparency statement or audit conducted according to an accepted professional toolkit or discipline-specific framework, stating how they will ensure that both their process of innovation and the product that results conform to ethical standards, along the lines explored by Winfield and Jirotka (2018). Such statements could be subject to external review with appropriate technical input. This solution would allow university ethics committees to focus on their existing expertise in review of the research process whilst not allowing researchers to assume that this form of review encompasses their entire ethical responsibility.
Conclusion

This paper has explored some challenges that data-driven research poses for the ethical review systems currently in place in British universities. Some of these challenges arise because the field is relatively fast moving and ethical standards are not yet fully established within governance frameworks (Mittelstadt, 2019). To some extent, then, the challenges will be resolved as these arrangements mature. There will remain, however, some more fundamental challenges specific to data-driven technologies. In particular, the technically complex and often opaque nature of the technologies themselves, the prospects of diffuse and unpredictable harms beyond the research process and the intertwining of academic research with commercial interests render data-driven research particularly challenging for university ethical review as currently conceived. As in other areas of academic research, it is likely that some level of university-based scrutiny will continue to be required alongside any advisory professional or disciplinary codes of conduct that may develop in the field. We need to be alert to the possibility that researchers will consider their research to be ethical if it passes university ethical governance processes, but that those processes as currently constituted may not in fact be fit to fulfill the purpose envisaged by ethical guidelines in AI.

Funding

All articles in Research Ethics are published as open access. There are no submission charges and no Article Processing Charges as these are fully funded by institutions through Knowledge Unlatched, resulting in no direct charge to authors. For more information about Knowledge Unlatched please see here: http://www.knowledgeunlatched.org

Preparation of this article was funded in part by APEX award APX\R1\201173 awarded by the British Academy, the Royal Academy of Engineering and the Royal Society and with generous support from the Leverhulme Trust.

ORCID iD

Christine Hine https://orcid.org/0000-0002-1172-0252

Notes

1. Within UK universities ethical review is conducted by university-based committees apart from those areas where review by another body is mandated. This is so in the case of most health-related research where review by an NHS research ethics committee is required. The current paper therefore excludes from consideration most health-related research involving AI conducted by university-based researchers.

2. These descriptions largely focus on health care scenarios but have considerable relevance in the university setting where ethics committees inherit much of their practice from biomedical contexts.

3. I am grateful to an anonymous reviewer for drawing attention to this comparison.
References

Beauchamp TL and Childress JF (2019) *Principles of Biomedical Ethics*. New York: Oxford University Press.

Black J and Murray AD (2019) Regulating AI and machine learning: setting the regulatory agenda. *European journal of law and technology* 10(3).

Cath C (2018) Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376: 20180080.

Corea F (2019) *An Introduction to Data. Everything You Need to Know about AI, Big Data and Data Science*. Cham, Switzerland: Springer.

Digital Culture Media and Sport Committee (2019) Disinformation and ‘Fake News’ Inquiry. FKN0077 Written evidence submitted by Aleksandr Kogan. Available at: https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/179102.htm (accessed 24 November 2020).

Dove ES (2020) *Regulatory Stewardship of Health Research: Navigating Participant Protection and Research Promotion*. Cheltenham: Edward Elgar Publishing.

Economic and Social Research Council (n.d.) Research ethics. Available at: https://esrc.ukri.org/funding/guidance-for-applicants/research-ethics/ (accessed 27 April 2021).

Edwards SJ (2009) The role, remit and function of the research ethics committee—1. The rationale for ethics review of research by committee. *Research Ethics* 5: 147–150.

Edwards SJ (2010a) The role, remit and function of the research ethics committee—2. Science and society: The scope of ethics review. *Research Ethics* 6: 58–61.

Edwards SJ (2010b) The role, remit and function of the research ethics committee—3. Balancing potential social benefits against risks to subjects. *Research Ethics* 6: 96–100.

Edwards SJ (2010c) The role, remit and function of the research ethics committee—4. Limits to Consent? *Research Ethics* 6: 159–163.

Edwards SJ (2011) The role, remit and function of the research ethics committee—5. Collective decision-making and research ethics committees. *Research Ethics* 7: 19–23.

Fell J, Shmueli G, Greene T et al. (2021) Seeing humans in the data: Ethical blind spots of Taiwan academic researchers in the era of behavioral big data. In: *Proceedings of the 54th Hawaii international conference on system sciences*, 5 January, 6599.

Floridi L and Cowls J (2019) A unified framework of five principles for AI in society. *Harvard Data Science Review* 1(1).

Gillon R (1994) Medical ethics: Four principles plus attention to scope. *BMJ* 309: 184.

Haggerty KD (2004) Ethics creep: Governing social science research in the name of ethics. *Qualitative Sociology* 27: 391-414.

Hedgecoe A (2009) “A form of practical machinery”: The origins of research ethics committees in the UK, 1967–1972. *Medical History* 53: 331–350.

Hedgecoe A (2016) Reputational risk, academic freedom and research ethics review. *Sociology* 50: 486–501.

High Level Expert Group on Artificial Intelligence (2019) *Ethics Guidelines for Trustworthy AI*. Brussels: European Commission. Available at: AIHLEG_EthicsGuidelinesforTrustworthyAI-ENpdf.pdf (accessed 27 April 2021).

Jennings S (2012) Response to Schrag: What are ethics committees for anyway? A defence of social science research ethics review. *Research Ethics* 8: 87–96.

Jobin A, Lenca M and Vayena E (2019) The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1: 389–399.
Leslie D (2019) Understanding artificial intelligence ethics and safety. Alan Turing Institute. 
*arXiv preprint arXiv:1906.05684.*

Metcalf J and Crawford K (2016) Where are human subjects in big data research? The emerging ethics divide. *Big Data & Society* 3: 1–14.

Mittelstadt B (2019) Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence* 1(11): 501–507.

Mittelstadt BD, Allo P, Taddeo M et al. (2016) The ethics of algorithms: Mapping the debate. *Big Data & Society* 3: 1–21.

Montreal AI Ethics Institute (2020) The State of AI Ethics. Montreal: MAIE. Available at: https://montrealethics.ai/wp-content/uploads/2020/10/State-of-AI-Ethics-Oct-2020.pdf (accessed 27 April 2021).

Morley J and Floridi L (2020) An ethically mindful approach to AI for health care. *Lancet (London, England)* 395: 113712.

Morley J, Floridi L, Kinsey L et al. (2019) From what to how. An overview of AI ethics tools, methods and research to translate principles into practices. *arXiv preprint arXiv:1905.06876.*

Morley J, Machado CC, Burr C et al. (2020) The ethics of AI in health care: A mapping review. *Social Science & Medicine* 260: 113172.

Nebeker C, Torous J and Bartlett Ellis RJ (2019) Building the case for actionable ethics in digital health research supported by artificial intelligence. *BMC Medicine* 17: 137.

Owen R, Macnaghten P and Stilgoe J (2012) Responsible research and innovation: From science in society to science for society, with society. *Science and Public Policy* 39: 751–760.

Rawbone R (2010) Inequality amongst RECs. *Research Ethics* 6: 1–2.

Schrag ZM (2011) The case against ethics review in the social sciences. *Research Ethics* 7: 120–131.

Stilgoe J, Owen R and Macnaghten P (2013) Developing a framework for responsible innovation. *Research Policy* 42: 1568–1580.

Universities UK (2012) The concordat to support research integrity. Available at: https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2012/the-concordat-to-support-research-integrity-2012.pdf

Universities UK (2016) The concordat to support research integrity: A progress report. Available at: https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2016/concordat-research-integrity-progress-report.pdf (accessed 27 April 2021).

University of Cambridge (2018) Statement from the University of Cambridge about Dr Aleksandr Kogan. Available at: https://www.cam.ac.uk/notices/news/statement-from-the-university-of-cambridge-about-dr-aleksandr-kogan (accessed 11 April 2018).

van den Hoonaard WC and Hamilton A (2016) *The Ethics Rupture: Exploring Alternatives to Formal Research-Ethics Review*. Toronto: University of Toronto Press.

Von Schomberg R (2013) A vision of responsible research and innovation. In: Owen R, Heintz M and Bessant J (eds) *Responsible Innovation*. London: John Wiley, 51–74.

Winfield AF and Jirotka M (2018) Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376: 20180085.