



Ethics and governance in the digital age

European View
2021, Vol. 20(2) 175–181
© The Author(s) 2021
DOI: 10.1177/17816858211061793
journals.sagepub.com/home/euv



Jana Mišić

Abstract

This article argues that ethics need not be toothless or side-lined in the technology governance debates. Rather, moral evaluation is necessary, even when legal compliance is already possible. Moral evaluation supplies answers not only to what is legal or illegal, but also to what is good and better for society. The article first defends a pragmatist ethics approach to uncovering the inevitability of values and norms embedded in digital technologies and related to their design and use. It then makes the case for policymakers engaging in the anticipatory ethics of technology. This approach provides a toolbox to tackle moral dilemmas and better understand what trustworthiness and ethics mean in certain contexts. The convergence of ethics and policy is not only worth pursuing but a necessity for good technology governance if we are to achieve a Europe fit for the digital age.

Keywords

Ethics, Governance, Emerging technologies, Anticipatory ethics of technology

Introduction

Controversies related to issues such as fake news and voting manipulation, privacy and surveillance, algorithmic discrimination and bias (to name a few) have prompted increased public scrutiny of digital technologies over the last decade. It has become clear that both end-users and those associated with technology development are eager to part with the ‘move fast and break things’ mentality of Silicon Valley and have found an alternative by shifting the focus to ethics and the ethical design and use of these technologies. The pertinence of advancing the ethical approach was solidified in 2019 when then newly elected President of the European Commission, Ursula von der Leyen, set a 100-day deadline for her Commission to propose policies to shape the future digital

Corresponding author:

J. Mišić, Rathenau Instituut, Anna van Saksenlaan 51, 2593 HW Den Haag, The Netherlands.
Email: j.misic@rathenau.nl



policy of the Union. Furthermore, making ‘Europe fit for the digital age’ was inaugurated as one of the six key areas of her agenda for Europe (European Commission 2021).

Yet, almost as quickly as ethical technology discourse has proliferated across the digital world, it has become apparent that changing the mantra will not suffice. The limitations of making ethics a buzzword have become discernible: ethics is often too abstract and toothless; it can be subsumed into corporate logics and incentives, or it can be kept in the echo-chambers of either the policymakers or the individual engineers who are now expected to build it into their everyday work (Green 2021). As a result, critics have voiced concerns about ‘ethics washing’ and ‘ethics bashing’ (Bietti 2020), that is, the promotion of ethics research to defuse criticism and regulation without committing to ethical behaviour.

While it has become clear that there is a significant gap between the theory of ethics principles for artificial intelligence (AI) or the Internet of Things, and the practical design and governance of these systems, this article argues that their convergence is not only worth pursuing but a necessity for good technology governance if we are to achieve a Europe fit for the digital age. To do so, two preconditions have to be met. The first defends the *ethics approach* to uncovering the inevitability of the values and norms embedded in digital technologies and surrounding their design and use. The second, then, is this approach’s contribution to the anticipatory governance of technology. The key proposal of this article is that ethics does not need to be superficial and depoliticising; rather, a practical, pragmatist ethics for emerging digital technologies would be a useful resource for the better anticipation and future governance of technology.

The importance of ethics in technology debates

It was only a few decades ago that the development of the ethics of technology as a discipline in its own right began (Franssen et al. 2018). This late emergence may seem odd given the importance technology has played in human progress, especially since the industrial revolution. However, without going into the rich history of the field, it is important to note two things. Initially, the dominant stance was one of technological determinism, or the value-neutrality thesis, that attributes consequences to the users of technology, rather than to the technological tools themselves, their designers or the design process.

As Joseph Pitt (2014, 95) famously asked, ‘Where are the values?’ in a specific bridge, smartphone or algorithm. Indeed, the physical manifestation of values, particularly in the age of digital technologies, is a tricky business. Pitt argued that people, their actions and the context contain or express values, *not* artefacts. Depending on how technology is used, the plethora of values it embodies are subject to change—an algorithm can be used for the better synthesis of information, but it can also be used as a tool of discrimination. Intuitively, however, it seems insufficient to claim that the artefact itself is void of values in its design or use. Consider Amazon Rekognition, facial recognition software that matched 28 members of Congress, disproportionately people of colour, with criminal

mugshot images (Snow 2018). If the algorithm discriminated, is it not possible that this was (inadvertently) built into its data training set, the choice of algorithmic models or perhaps its deep learning logic?

The stance that technologies embody, advance or hinder certain values countered the value-neutrality thesis rather promptly. Novel views include conceptualisations of technology as a political phenomenon, a social activity, a cultural phenomenon, a professional activity (engineering ethics) or a cognitive activity (Franssen et al. 2018). In this article, we zoom in on the idea that technology can embody specific forms of power and authority. Langdon Winner (1980), the proponent of the political artefacts thesis, claims that certain technologies can indeed show us where values are, in the sense that they demand compatibility with certain social and political relations. For example, nuclear power plants or autonomous warfare drones seem to require a hierarchical, authoritative management structure. This structure is often closed, exclusive and a cause of moral dilemmas among policymakers and the public. Technology as a political artefact can also embody political bias. Winner refers to the Long Island overpasses which were designed in such a way that public buses could not pass underneath. This meant that the main users of public transportation, that is, the poorer black classes, were less likely to access Manhattan's public parks. Thus, while the design fulfilled the task of highway transportation, it also encompassed 'purposes far beyond original use' (Winner 1980, 125). To avoid such a negative political function, what is needed is a certain level of democratisation of technological development, that is, the inclusion of ordinary people in its shaping. This is not always possible, and so we are left with the open-ended question of how to ensure 'democratic' innovation.

Democratic innovation demands awareness from users of how to value the trade-offs (privacy versus autonomy, safety versus sustainability, etc.) that are inherent in governing technologies. Value trade-offs, on the other hand, can only be revealed by engaging with the ethics or normativity of digital technologies. Ethics answers the question of what ought to be done and aids us in delineating the notions such as trustworthiness, justice, fairness or privacy that permeate discussions on policymaking and technological innovation. It does so in at least one important way. The ethicists of technology may make or choose to abstain from prescribing a certain course of action, but their work brings clarity about which values are asserted as a rationale for choices at the forefront of the design or implementation of technology.

As an analytical approach, infusing ethics into the assessment process thus helps to articulate norms and value assumptions. Beyond the hard impacts that we can measure, such as formal procedures (e.g. risk impact assessments) or legal regulation (e.g. the General Data Protection Regulation), ethics draws out the soft impacts as well (Swierstra 2015). Examples of the soft impacts of technology can be the consequences for wellbeing, relationships, quality of life, identity or social order. For example, with the proliferation of cameras and facial recognition, privacy takes on the form of a struggle for intimacy and control over one's identity. Along these lines, some have argued that soft and hard ethics are complimentary—that pure regulation is not enough, and that what is

legal is not necessarily conducive to the idea of a good society or a good life (Floridi 2018).

Navigating the soft and hard ethics of technology

What digital innovation will bring next is an exciting prospect. Innovation in its own right is something that we value, and technological innovation especially so, particularly in modern society. The question is, what comes next? A new essential healthcare app reliant on AI? A quantum computing smartphone in everyone's pocket? Or a smart assistant that goes beyond informing us about the weather or playing a favourite song? In any case, the future seems to be riddled with disruptive digital and emerging technologies. Unlike mature technologies, emerging technologies such as deep learning algorithms are still in their infancy—the research and development phase—and their exact impact is uncertain and difficult to predict. This disruption is both exciting and challenging since it presupposes that innovation leads and everything else follows: expectations, labour conditions, habits, norms, business models and standards of living (Floridi 2018).

To understand the disruptive effects of emerging digital technologies, we need to turn to practice-oriented forms of ethics. These kinds of ethics emphasise that norms and values are a part of our engagement with the world. Engagement is not only theoretical but practical as we continuously have to make value judgements: is autonomy more valuable than giving away our personal choice to AI recommendation systems for the sake of comfort? Pragmatist ethics offers a way to understand the interaction between technology, values and the world as a lived practice. It stresses the process instead of the values as objective entities and implies that values are interactive, dynamic and tied to doing. The argument is as follows: bearing in mind all other factors, but zooming in on technology, the introduction of new technologies creates consequences that require new moral responses (e.g. the value of sustainability has become very pronounced lately), but also creates new opportunities, new moral dilemmas (e.g. predictive genetics) and new experiences (e.g. online dating). Arising from these, new morally problematic situations require new moral solutions, and thus we can speak of technologically induced value change.

Without going into epistemological and ontological questions about values, we can agree that human interactions and experiences lead to a certain conceptualisation of what the right thing to do is and how this might change over time or with context, leading to a chance to review our actions. People might still use the terms 'privacy' or 'autonomy', but the meanings attached to them might change. Before the introduction of algorithms into the decision-making process, the ability to make judgements and decide on 'fairness' was reserved for people. Today, however, some consider that an algorithm might be better suited to making fairer, more bias-free decisions than a human. The value judgements that people make are not necessarily entirely subjective, nor would pragmatists require them to be. But they would stress their function as evaluative devices in specific situations from which we learn. As time goes by, morality then tends to go unnoticed as it has been routinised in regular evaluations of what should or should not be done. However, this is also why emerging technologies cause a reflection on morality:

something that we have never considered before challenges the routine. An example would be the use of autonomous drones for military purposes, which cause a shift in a soldier's moral experience and in the moral responsibility assigned to/felt by soldiers (Boenink and Kudina 2020). This continuous evolution and interaction between morality and technology has been called 'technomoral change' (Swierstra et al. 2009) and is essential to any debate about the ethics and governance of any technology. The question now becomes: how should we imagine future governance in lieu of technomoral change to achieve what the EU has called a human-centric approach?

Anticipatory ethics and governance

The key point of this article is that ethics need not be superficial and depoliticising; rather, it can be conducive to good technology governance. Let us consider the validity of this claim in the context of the governance of AI, a key emerging digital technology today. A great many public and private actors have issued statements on how AI will change society, culminating in the advent of ethical guidelines (Jobin et al. 2019). However, as noted earlier, a significant gap exists between the theory of principles and practical design and governance. High-level principles are vague and difficult to operationalise in the 'what' and 'how' of AI governance. That being the case, we then turn to tools such as ethical impact assessments or privacy assessments as ethics checklists, which are often too strict and unresponsive to specific contexts. Privacy within the family and at home has a different value to us compared to privacy in public, for example.

As argued above, digital technologies are morally disruptive, and since they are currently in their nascent phase, it is difficult to evaluate just what societal impact they will have. It is therefore important to seek out future concerns and the likely technomoral value change. This is where ethics comes in. This article proposes that not only engineers but policymakers should be involved in the anticipatory ethics of technology (AET) (Brey 2012). The AET approach combines ethical analysis and foresight methods with public and stakeholder engagement. As such, it provides enough interaction between future policy and ethics, as well as addressing the three levels of technology in society: the type of technology, the features of the specific artefact and its application level (Brey 2017, 187). In simpler terms, the process should analyse (1) the risks of AI as an emerging technology, (2) the algorithms and their inherent properties, and (3) their application in a specific context. The analysis starts with foresight, identification and the weighing of ethical issues. In the final stage, after the evaluation, ethicists engage with other stakeholders, such as policymakers. In the governance stage, recommendations are made specifically for each context, avoiding the vagueness of ethics principles and the strictness of generic impact assessments. What is more, the toolbox of foresight methods is rich: stakeholders can engage in horizon scanning, scenario-based methods, expert consultations, relevance trees, the Delphi method, trend analysis, road mapping and participatory deliberation (Brey 2017, 185–6).

While policymakers do consider the future in their evaluation, the dominant narrative is that moral evaluation is not necessary when legal compliance is already available. The

debates surrounding the General Data Protection Regulation exemplified this. However, legal compliance is not enough if the policy is to steer society in the right direction. Moral evaluation then supplies the answer to not only what is legal or illegal, but also what is good and better for society. Such an endeavour should be the preferred state for both ethics (values and judgements) and governance (the management of technology).

Conclusion

What does it mean for an emerging technology to be ‘fair’ or ‘non-discriminatory’ in terms that can be operationalised? Providing a rigorous understanding of these terms has long been a preoccupation of moral and political philosophers, and their work can help illuminate policymaking. It can also help to guide societal discussions on how we want to proceed with innovation and the use of emerging technologies. This article has advocated that ethics is not toothless, but can be constructive and pragmatic in the best sense possible. A forward-looking analysis such as AET can better inform both the design and policy negotiations that are necessary for the acceptable introduction of digital technologies in society.

By engaging in AET, a broad cross section of stakeholders—policymakers, designers, lawyers and citizens—can work towards more robust and ethically sound governance in at least three ways. First, by revealing the values and value judgements that unavoidably arise with the introduction of new technologies. Second, by articulating not only the hard but also the soft impacts of technologies, and weighing their desirability. And third, by bringing the future closer through foresight and future studies. Future technology policy debates should take all three into account. The anticipatory governance of emerging technologies provides the toolbox to tackle moral dilemmas and reveal what trustworthy and ethical technology means. Understanding these notions helps us to better understand ourselves and our relationships with technological artefacts, and as such should form the foundation of any desirable governance framework.

References

- Bietti, E. (2020). From ethics washing to ethics bashing: A view on tech ethics from within moral philosophy. In Association for Computing Machinery, *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)* (pp. 210–19). New York: ACM.
- Boenink, M., & Kudina, O. (2020). Values in responsible research and innovation: From entities to practices. *Journal of Responsible Innovation*, 7(3), 450–70.
- Brey, P. (2012). Anticipatory ethics for emerging technologies. *NanoEthics*, 6(1), 1–13.
- Brey, P. (2017). Ethics of emerging technologies. In S. O. Hansson (ed.), *Methods for the Ethics of Technology* (pp. 175–192). London: Rowman and Littlefield International.
- European Commission. (2021). *The European Commission's priorities: 6 Commission priorities for 2019–24*. https://ec.europa.eu/info/strategy/priorities-2019-2024_en. Accessed 16 September 2021.
- Floridi, L. (2018). Soft ethics and the governance of the digital. *Philosophy & Technology*, 31(1), 1–8.

- Franssen, M., Lokhorst, G. J., & van de Poel, I. (2018). Philosophy of technology. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/fall2018/entries/technology>. Accessed 16 September 2021.
- Green, B. (2021). The contestation of tech ethics: A sociotechnical approach to ethics and technology in action. *SSRN*. doi:10.2139.3859358.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–99.
- Pitt, J. C. (2014). ‘Guns don’t kill, people kill’; Values in and/or around technologies. In P. Kroes & P. P. Verbeek (eds.), *The moral status of technical artefacts* (pp. 89–102). Philosophy of engineering and technology. Dordrecht: Springer.
- Snow, J. (2018). Amazon’s face recognition falsely matched 28 members of Congress with mugshots. *American Civil Liberties Union*, 26 July. <https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28>. Accessed 16 September 2021.
- Swierstra, T. (2015). Identifying the normative challenges posed by technology’s ‘soft’ impacts. *Etikk i Praksis*, 9(1), 5–20.
- Swierstra, T., Stemerding, D., & Boenink, M. (2009). Exploring techno-moral change: The case of the obesity pill. In P. Sollie & M. Duwell (eds.), *Evaluating new technologies* (pp. 119–38). Dordrecht: Springer.
- Winner, L. (1980). Do artifacts have politics? *Daedalus*, 109(1), 121–36.

Author biography



Jana Mišić is a PhD candidate in ethics and technology at the University of Twente and the Rathenau Institute, The Hague. She is also a Marie Skłodowska-Curie fellow within the PROTECT project. Her research focuses on ethical risk frameworks for algorithmic decision-making in the public sector.