Commentary

BIG DATA & SOCIETY

Big Data & Society
July-December 2020: 1–5

DOI: 10.1177/2053951720942541
journals.sagepub.com/home/bds
SAGE

AI ethics should not remain toothless! A call to bring back the teeth of ethics

Anaïs Rességuier[1] and Rowena Rodrigues[2]

Abstract

Ethics has powerful teeth, but these are barely being used in the ethics of AI today – it is no wonder the ethics of AI is then blamed for having no teeth. This article argues that 'ethics' in the current AI ethics field is largely ineffective, trapped in an 'ethical principles' approach and as such particularly prone to manipulation, especially by industry actors. Using ethics as a substitute for law risks its abuse and misuse. This significantly limits what ethics can achieve and is a great loss to the AI field and its impacts on individuals and society. This article discusses these risks and then highlights the teeth of ethics and the essential value they can – and should – bring to AI ethics now.

Keywords

AI ethics, law of AI, regulation of AI, ethics washing, EU HLEG on AI, ethical principles

Ethics has great powerful teeth. Unfortunately, we are barely using them in AI ethics – it is no wonder then that the ethics of AI is called toothless. This article reflects on the various ethics mechanisms that have emerged over the past couple of years to respond to the massive deployment and use of AI in society and its associated risks. These mechanisms include lists of principles, ethics codes, recommendations and guidelines. Yet, as many have shown, these ethics developments, while promising, are also problematic: their effectiveness is still to be demonstrated and they are particularly prone to manipulation, especially by industry. This is a loss to the AI field and significantly limits what ethics can achieve for society and individuals. However, as this article shows, the issue is not that ethics is worthless (or toothless) in the face of current AI deployment; it is rather that ethics is used (or manipulated) in such a way that it is rendered ineffective for AI ethics.

The article begins by highlighting the nature of ethics in the AI ethics field today – AI ethics is primarily principled, i.e. it follows a 'law' conception of ethics. It then shows how this approach to ethics fails to achieve what it pretends to achieve. This article next builds on the real value of ethics – its 'teeth' – which we define as a constantly renewed ability to see the new as it emerges. We show how this capacity to avoid cognitive and perceptive inertia that make us passive in the face of new developments is highly critical for AI ethics today. Finally, while we recognise that the legalistic approach to ethics is not completely off the mark, we argue that it is the *end* of ethics, not its teeth, not the most precious and critical aspects that ethics has to offer.

AI ethics today

There are many ongoing discussions and initiatives on the ethics of AI in various stakeholder quarters (policy, academia, industry and even the media). We can certainly rejoice about this. In particular, policymakers (e.g. European Commission and the European Parliament) and industry are showing much concern about getting things right to ensure the ethical and responsible development and deployment of AI in society.[1] It is now well recognised that things could go really wrong if AI is implemented without due regard and consideration for its potentially harmful impacts on individuals, on specific communities and on society as a whole (including, for example, bias and discrimination, injustice, privacy infringements, increase in surveillance, loss of autonomy, overdependency on

[1]Trilateral Research Ltd, Waterford, Ireland
[2]Trilateral Research Ltd, London, UK

Corresponding author:
Anaïs Rességuier, Trilateral Research Ltd, Marine Point, 2nd Floor, Belview Port, Waterford X91 W0XW, Ireland.
Email: anais.resseguier@trilateralresearch.com

technology, etc.).[2] We can then observe a turn to ethics to ensure that AI is deployed in a manner that respects dearly held societal values and norms, and puts them at the heart of responsible technology development and deployment (Hagendorff, 2020; Jobin et al., 2019). The 'Ethics guidelines for trustworthy AI' drafted by the High-Level Expert Group on AI – a group set up by the European Commission in 2018 – is one example of such recent ethics initiatives (High-Level Expert Group on Artificial Intelligence, 2019).

However, the manner in which 'ethics' is currently being used in the field of AI ethics is problematic. Today's AI ethics is dominated by what the British philosopher G.E.M. Anscombe calls a '*law conception of ethics*', i.e., a view on the ethics endeavour that makes it a sort of replica of law (Anscombe, 1958).[3] Viewing ethics as a 'softer version of the law' is common (Jobin et al., 2019: 389). However, this is only one approach to ethics, and one that, as Anscombe has shown, is problematic. For AI ethics, it is problematic in at least two ways.

Misusing ethics as a replacement for regulation

First, it is problematic because of its potential for misuse as a replacement for regulation (whether through law, policies or standards). Many articles have argued the following point over the last couple of years: AI ethics is, by itself, deficient in regulating behaviours and practices for proper development and deployment of AI (Article 19, 2019; Greene et al., 2019; Hagendorff, 2020; Jobin et al., 2019; Klöver and Fanta, 2019; Mittelstadt, 2019; Wagner, 2018). Wagner points, for instance, to the case of a member of the Google DeepMind ethics team at the *Conference on World Affairs 2018* repeatedly claiming 'how ethically Google DeepMind was acting, while simultaneous avoiding any responsibility for the data protection scandal at Google DeepMind' (Wagner, 2018). Ochigame (2019) states very critically that the discourse of 'ethical AI', 'was aligned strategically with a Silicon Valley effort seeking to avoid legally enforceable restrictions of controversial technologies'.

Considering it has no means to ensure compliance, ethics does indeed fall short in this respect. As Hagendorff puts it, ethics 'lacks mechanisms to reinforce its own normative claims' (2020: 99). If ethics is about regulation, then indeed ethics has no teeth. This is the view of the human rights organisation Article 19 for which, although ethics initiatives 'put forth admirable goals', 'their common lack of accountability and enforcement mechanisms' makes these initiatives ineffective (Article 19, 2019: 18). Ultimately and unsurprisingly, ethics is then blamed for being toothless.

However, it is essential to be clear here: the issue is not that ethics is asked to do something for which it is too weak, or too soft. It is rather that it is asked to do something that it is *not* designed to do. Blaming ethics for having no teeth to ensure compliance with whatever it calls for is like blaming the fork for not cutting meat properly: this is not what it is designed to do. The objective of ethics itself is not to impose particular behaviours and to ensure these are complied with. The problem arises when it is used to do so. This is particularly evident in AI ethics, where ethical principles, norms or requirements are called for to regulate AI and ensure that it does not harm individuals and the society at large (e.g. AI HLEG).

Some argue that this misuse of ethics is an intended one, driven by the desire to ensure that AI will not be regulated by law, i.e. more flexibility is possible and no hard lines will be put in place restricting industrial and commercial interests related to this technology (Klöver and Fanta, 2019). This critique has been addressed to the AI HLEG guidelines, for instance. Article 19 points to the fact that:

> during deliberations at the European High-Level Expert Group on Artificial Intelligence (EU-HLEG), industry was heavily represented, but academics and civil society did not enjoy the same luxury. And while some non-negotiable ethical principles were originally articulated in the document, these were omitted from the final document due to industry pressure. (Article 19, 2019: 18)

Using ethics to prevent the implementation of legal regulation that is actually necessary is a serious and worrying abuse and misuse of ethics. This then leads to ethics washing and its cousins: ethics shopping, ethics shirking, etc. (Floridi, 2019; Greene et al., 2019; Wagner, 2018).

We are not using the teeth of ethics in AI ethics today

Second, because the AI ethics field tends to be dominated by this 'law conception of ethics', it does not actually truly use what ethics has to offer, its actual proper teeth, despite the great need for them. What are these ethics teeth and what can they bring to the field?

The real teeth of ethics consist of a *constantly renewed ability to see the new* (Laugier, 2013). Ethics is primarily a form of attention, a continuously refreshed and agile attention to reality as it evolves.

The ethics of care has particularly highlighted attentiveness as a fundamental aspect of ethics (Tronto, 1993: 127).[4] Ethics in that sense is a powerful tool against cognitive and perceptive inertia that hinders our capacity to see what is different from before or in different contexts, cultures or situations and what, as a result, calls for a change in behaviour (regulation included). This is especially needed for AI, considering the profound changes and impacts it has, and is bringing to society, and to our very ways of being and acting.

This constantly refreshed capacity to perceive the world is the one that helps us not to be boiled alive like the frog: it helps us notice small changes as they unfold. In the context of AI, the increasingly hot water is a combination of evolutions that include an expansion and deepening of surveillance by governments and private companies, an increasing dependency on technology and the deployment of biased systems that lead to discrimination towards women and minorities. The progressive evolutions these bring to society need to be closely studied and resisted when their negative impacts outweigh their benefits.

In that sense, ethics entertains a very close link with social sciences, as an effort to see what we do not otherwise see. Ethics helps us look concretely at how the world changes. It helps clean up the lens through which we see the world so that we can be more attentive to its transformations (and AI does bring many of these). It is essential for ethics to support us in this regard. It helps us to be less passive towards these changes and puts us in a better position to then steer them in ways that do not harm individuals and society, and that help us live better. In his piece on the 'Ethics of AI ethics', Hagendorff makes a similar proposition by questioning the dominant deontological approach to ethics in AI ethics (what we have called in this article a legalistic approach to ethics) whose purpose is primarily 'to limit, control, or steer' (2020: 112). He presses the need to turn to virtue ethics for the AI field, an ethics that aims at 'broadening the scope of action, uncovering blind spots, promoting autonomy and freedom, and fostering self-responsibility' (Hagendorff, 2020: 112). Other ethical theory frameworks that would bring significant value to the AI ethics debate today include in particular the Spinozist approach that focuses on increase or decrease of agency and capacity of action.

Ethical principles, norms and values are an 'end' of ethics, not ethics itself

So, are we simply getting it wrong in AI ethics, which today, as we have shown, is dominated by a 'law-conception of ethics'? Is the current legalistic approach to ethics completely off the mark? Not exactly. The issue is rather that principles, norms, values – this law conception of ethics that is dominant today in AI ethics – are rather an *end* of ethics, not ethics itself. The end here means two things at the same time.

First (1), it is an end of ethics in the sense of its *finality*, *where ethics ultimately leads*, i.e., shaping the laws, decisions, behaviours, and actions in ways that are close to what the society values. This is where we find ethics as the production of principles (such as in the AI HLEG requirements) or the operationalisation of ethical principles and values or norms to specific contexts. This process of operationalisation of ethical norms can be seen, for instance, in the Ethics appraisal procedure of the European Commission's research funding programme[5] or in ethics impact assessments that examine how a particular new process or technology might affect ethical norms and values.[6] These are undoubtedly useful undertakings that lead to positive impacts on society and individuals. Ethics as the production of principles is also helpful to shape policies and regulatory frameworks. Current policy and legislative developments at the EU level such as the European Commission 'White Paper on Artificial Intelligence' (February 2020) and the European Parliament proposed 'Framework of ethical aspects of artificial intelligence, robotics, and related technologies' (April 2020) are strongly informed by the AI HLEG guidelines. Here, ethics clearly provides guidance on rights and wrongs; what should be done and what should be avoided.

However, it is essential to remember that ethics as ethical principles is also an end of ethics in another sense (2): *where it stops*, where the reflection is interrupted, where this constantly renewed attention terminates. So, when ethics settles in different principles, norms or requirements, it has reached its end. If we have reached a sufficient level of certitude and confidence in what are the right decisions and actions, then there is no need for ethics.

Ethics is about navigating murky and risky waters; to do so, it needs to be watchful. For instance, in the field of AI, ethical principles do not, by themselves, help practically explore challenging issues such as fairness in highly complex socio-technical systems. These need to be investigated in-depth to ensure we are not implementing systems that go against dearly held norms and values. Without a continuous process of questioning what is or may be obvious, of digging behind what seems to be settled, of keeping alive this interrogation, ethics is rendered ineffective. And thus, the settling of ethics into established norms and principles comes down to its termination.

Considering the radical, massive and widespread impact of AI on the society, it is critical to keep ethics agile and alive. AI ethics is in a deep and vital need of the continuously renewed process of questioning the world and the lenses through which we see it – consciously, constantly and iteratively.

Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: This article reflects only the views of the authors and does not intend to reflect those of the European Commission. The European Commission is not responsible for any use that may be made of the information it contains.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This article was developed as part of the SIENNA project (Stakeholder-informed ethics for new technologies with high socio-economic and human rights impact) which has received funding under the European Union's H2020 research and innovation programme under grant agreement No 741716.

ORCID iD

Anaïs Rességuier https://orcid.org/0000-0002-0461-0506

Notes

1. The organisation Algorithm Watch has compiled an inventory of all ethics guidelines that have been developed globally. As of April 2020, the list comprised 160 documents. See Algorithm Watch, 'AI ethics Guidelines Global Inventory': https://inventory.algorithmwatch.org
2. The EU-funded research project SIENNA (Stakeholder-Informed Ethics for New Technologies with high Socio-Economic and Human Rights Impact) provided an overview of the ethical issues raised by AI and robotics (Jansen and Brey, 2019). See also Boddington, 2017. On the risk related to loss of autonomy, see Rodrigues and Rességuier (2019).
3. The following article makes a similar argument in relation to the ELSI (Ethical, Legal, Social Implications) field in Canada and what the authors call its 'juridification': López and Lunau (2012).
4. As the reviewer of this piece rightly noted, this view on ethics is strongly influenced by the conception of ethics in the ethics of care, an approach to ethics informed by the experience of women, as initially developed by the ethicist and psychologist Carol Gilligan and further expanded by the political scientist Joan Tronto (Gilligan, 1982; Tronto, 1993). A future piece will detail the great contribution the ethics of care can bring to AI ethics.
5. European Commission, H2020 Online Manual: Ethics: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/ethics_en.htm
6. See for instance, European Committee for Standardization (2017): https://satoriproject.eu/media/CWA17145-23d2017.pdf

References

Anscombe GEM (1958) Modern moral philosophy. *Philosophy* 33(124): 1–19.

Article 19 (2019) Governance with teeth: How human rights can strengthen FAT and ethics initiatives on artificial intelligence. Available at: https://www.article19.org/wp-content/uploads/2019/04/Governance-with-teeth_A19_April_2019.pdf (accessed 18 February 2020).

Boddington P (2017) *Towards a Code of Ethics for Artificial Intelligence*. Cham: Springer.

European Committee for Standardization (2017) CEN Workshop Agreement: Ethics assessment for research and innovation – Part 2: Ethical impact assessment framework (by the SATORI project). Available at: https://satoriproject.eu/media/CWA17145-23d2017 (accessed 13 July 2020) .

European Parliament JURI (April 2020) Framework of ethical aspects of artificial intelligence, robotics and related technologies, draft report (2020/2012(INL)). Available at: https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?lang=&reference=2020/2012 (accessed 13 July 2020).

Floridi L (2019) Translating principles into practices of digital ethics: Five risks of being unethical. *Philosophy & Technology* 32: 185–193.

Gilligan C (1982) *In a Different Voice: Psychological Theory and Women's Development*. Cambridge: Harvard University Press.

Greene D, Hoffmann A and Stark L (2019) Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. In: *Proceedings of the 52nd Hawaii international conference on system sciences*, Maui, Hawaii, 2019, pp.2122–2131.

Hagendorff T (2020) The ethics of AI ethics. An evaluation of guidelines. *Minds and Machines* 30: 99–120.

High-Level Expert Group on Artificial Intelligence (2019) Ethics guidelines for trustworthy AI. Available at: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Jansen P, Brey P, Fox A, Maas J, Hillas B, Wagner N, Smith P, Oluoch I, Lamers L, van Gein H, Resseguier A, Rodrigues R, Wright D, Douglas D (2019) Ethical analysis of AI and robotics technologies. August, SIENNA D4.4, https://www.sienna-project.eu/digitalAssets/801/c_801912-l_1-k_d4.4_ethical-analysis–ai-and-r–with-acknowledgements.pdf (accessed 13 July 2020).

Jobin A, Ienca M and Vayena E (2019) The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1(9): 389–399.

Klöver C and Fanta A (2019) No red lines: Industry defuses ethics guidelines for artificial intelligence. Available at:

https://algorithmwatch.org/en/industry-defuses-ethics-guidelines-for-artificial-intelligence/

Laugier S (2013) The will to see: Ethics and moral perception of sense. *Graduate Faculty Philosophy Journal* 34(2): 263–281.

López JJ and Lunau J (2012) ELSIfication in Canada: Legal modes of reasoning. *Science as Culture* 21(1): 77–99.

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

Ochigame R (2019) The invention of "Ethical AI" how big tech manipulates academia to avoid regulation. *The Intercept*. Available at: https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/?comments = 1

Rodrigues R and Rességuier A (2019) The underdog in the AI ethical and legal debate: Human autonomy. In: *Ethics Dialogues*. Available at: https://www.ethicsdialogues.eu/2019/06/12/the-underdog-in-the-ai-ethical-and-legal-debate-human-autonomy/

Tronto J (1993) *Moral Boundaries: A Political Argument for an Ethic of Care*. New York: Routledge.

Wagner B (2018) Ethics as an escape from regulation: From ethics-washing to ethics-shopping. In: Bayamlioglu E, Baraliuc I, Janssens L, et al. (eds) *Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen*. Amsterdam: Amsterdam University Press, pp. 84–89.