Check for updates

ESSAY

REVISED Ethics as attention to context: recommendations for the ethics of artificial intelligence

[version 2; peer review: 1 approved, 2 approved with reservations]

Anais Resseguier[1], Rowena Rodrigues[2]

[1]Trilateral Research, Waterford, X91 W0XW, Ireland
[2]Trilateral Research, London, SW1X 7QA, UK

v2 **First published:** 26 Mar 2021, **1**:27
https://doi.org/10.12688/openreseurope.13260.1
Latest published: 13 Dec 2021, **1**:27
https://doi.org/10.12688/openreseurope.13260.2

Abstract

This article shows that current ethics guidance documents and initiatives for artificial intelligence (AI) tend to be dominated by a principled approach to ethics. Although this brings value to the field, it also entails some risks, especially in relation to the abstraction of this form of ethics that makes it poorly equipped to engage with and address deep socio-political issues and the material impacts of AI. This is particularly problematic considering the risk for AI to further entrench already existing social inequalities and injustices and contribute to environmental damage. To respond to this challenge posed by AI ethics today, this article proposes to complement the existing principled approach with an approach to ethics as attention to context and relations. It does so by drawing from alternative ethical theories to the dominant principled one, especially the ethics of care or other feminist approaches to ethics. Related to this, it encourages the inclusion of social sciences and humanities in the development, deployment and use of AI, as well as in AI ethics discussions and initiatives. This article presents this proposal for an ethics as attention to context and formulates a series of practical recommendations to implement this proposal concretely.

Keywords

AI ethics, ethics as attention, practical ethics, social impacts of AI, technology ethics, social sciences and humanities, technology and society

H2020 This article is included in the Horizon 2020 gateway.

Open Peer Review

Approval Status ? ✓ ?

	1	2	3
version 2 (revision) 13 Dec 2021			
version 1 26 Mar 2021	? view	✓ view	? view

1. **Nathalie Grandjean**, University of Namur, Namur, Belgium
University of Namur, Namur, Belgium

2. **Marcello Ienca**, ETH Zurich, Zürich, Switzerland

3. **Mathieu Corteel**, Sciences Po, Paris, France

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the Artificial Intelligence and the Social Sciences and Humanities collection.

Corresponding author: Anais Resseguier (anais.resseguier@trilateralresearch.com)

Author roles: Resseguier A: Conceptualization, Formal Analysis, Investigation, Methodology, Project Administration, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Rodrigues R**: Conceptualization, Funding Acquisition, Investigation, Project Administration, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This research was financially supported by the European Union's Horizon 2020 research and innovation programme under the grant agreement No 741716 (project SIENNA).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Resseguier A and Rodrigues R. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Resseguier A and Rodrigues R. **Ethics as attention to context: recommendations for the ethics of artificial intelligence [version 2; peer review: 1 approved, 2 approved with reservations]** Open Research Europe 2021, **1**:27 https://doi.org/10.12688/openreseurope.13260.2

First published: 26 Mar 2021, **1**:27 https://doi.org/10.12688/openreseurope.13260.1

REVISED **Amendments from Version 1**

A new version of this article was created in November 2021 taking into account the feedback of three reviewers. We are grateful to the reviewers for their thoughtful comments and suggestions that have helped improve the text. In particular, they helped clarify several key concepts of this paper ("materiality", "context" and Canguilhem's "normative capacity") and specify AI ethics processes currently happening, if at all, in ethical review, and the challenges of a one-off review process. The reviewers' comments also encouraged us to highlight more clearly the overall scope and ambition of this paper, i.e., its theoretical ambitions and the practical recommendations that draw from them (see endnote 50).

Any further responses from the reviewers can be found at the end of the article

Introduction

The ethics of artificial intelligence (AI) have generated a large amount of interest over the last few years[1]. The numerous ethics guidelines and other forms of ethics initiatives produced provide ample evidence of this[2]. Although the field has seen major developments in various arenas (including in academia, industry, and policy), it has also come under intense criticism for being too abstract and high-level, and therefore, unable to properly guide technological development, deployment and use[3]. Critics have highlighted that some of these initiatives lead to ethics washing[4] and/or contribute to reproducing structural inequalities in society[5].

Although this essay recognises the value of the ethics of AI and what they have produced over the last few years, it also acknowledges that efforts in this area are still needed, especially *to move beyond the abstraction of current AI ethics initiatives*. More recently, there have been numerous efforts at making ethics guidelines more fit for purpose by operationalising them to specific sectors of application or development contexts[6]. These are assuredly necessary and praiseworthy efforts.

Several experts on AI have critiqued AI ethics for failing to consider contextual elements, especially the socio-political context, which, as they note, is essential to address ethical challenges posed by this technology[7]. This essay shows that 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*. It points to the need to complement this approach and makes a series of practical recommendations. As such, it calls for *a shift of attention in AI ethics: away from high-level abstract principles to concrete practice, context and social, political, environmental materialities.* It draws extensively from critical approaches to AI and ethics, especially those emerging from feminist perspectives (including the ethics of care). It also derives from consultation with a number of stakeholders as part of SIENNA engagement[8].

This essay first undertakes a theoretical detour to better understand the nature of the ethics approach at stake in current AI ethics (section 2) and how to best complement it (section 3). This detour is necessary to build the conceptional groundwork for the practical recommendations developed in section 4, which contains the actual proposals to promote ethical AI. This essay, and the guidance it offers, are addressed to policymakers in government, AI ethicists, engineers (software engineers, data scientists, etc.), organisations developing AI, and, more generally, anyone concerned by the development, deployment and use of AI and the potential social and ethical impacts of this technology[9].

[1] This article includes in the concept of Artificial Intelligence systems that have a physical component, i.e., AI-powered robotics. This article was developed as part of the SIENNA project - *Stakeholder-informed ethics for new technologies with high socio-economic and human rights impact* (2017–2021). SIENNA has received funding from the European Union's H2020 research and innovation programme under grant agreement No 741716. This article and its content reflect only the view of the authors.

[2] The following references give a listing of these: Jobin, Anna, Marcello Ienca, and Effy Vayena, "The Global Landscape of AI Ethics Guidelines", *Nature Machine Intelligence* 1, no. 9, 2019, pp. 389–99; Hagendorff, Thilo. "The Ethics of AI Ethics. An Evaluation of Guidelines", *Minds and Machines*, no. 30, 2019, pp. 99–120; Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Christopher Nagy, and Madhulika Srikumar, "Principled Artificial Intellience: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI", Cambridge, MA: Berkman Klein Center for Internet & Society at Harvard University, January 2020, https://cyber.harvard.edu/publication/2020/principled-ai. The organisation Algorithm Watch maintains a Global Inventory of AI Ethics Guidelines and had 160 items as of April 2020: https://inventory.algorithmwatch.org.

[3] Mittelstadt, B., "Principles Alone Cannot Guarantee Ethical AI", *Nature Machine Intelligence 1*, Nov 2019, pp. 501–507.

[4] Ethics washing corresponds to the use of ethics to avoid strict legal regulation. See in particular Wagner, B., "Ethics as an escape from regulation: From ethics-washing to ethics-shopping" in Emre Bayamlioglu, Irina Baraliuc, Liisa Janssens, and Mireille Hildebrandt, *Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen*, Amsterdam University Press, Amsterdam, 2019, pp. 84–89.

[5] D'Ignazio, Catherine, and Klein, Lauren F., *Data Feminism*, MIT Press, Cambridge, MA; London, England, 2020.

[6] See for instance the Ethics by design framework developed in SIENNA D5.4 "Multi-stakeholder Strategy and Practical Tools for Ethical AI and Robotics" (work in progress); Brey, Philip, Björn Lundgren, Kevin Macnish, and Mark Ryan, "Guidelines for the Ethical Use of AI and Big Data Systems", SHERPA project, July 2019. See also the High-Level Expert Group on AI, "Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment", European Commission, Brussels, July 2020, https://altai.insight-centre.org.

[7] Gebru, Timnit, "Race and Gender", in Markus D. Dubber, Frank Pasquale, and Sunit Das, *The Oxford Handbook of Ethics of AI*, Oxford, Oxford University Press, 2020.

[8] In particular, an earlier version of this piece was presented at the SIENNA online Workshop on Multi-Stakeholder Strategies for Ethical AI on 9 September 2020. It was entitled "Critical Insights from the Social Sciences and Humanities for the Ethics of AI". This piece includes feedback received by stakeholders on this occasion as well as during the EUREC/SIENNA online workshop on guidance documents for Research Ethics Committees on 26 and 27 October 2020.

[9] The authors of this article wish to express great gratitude to a number of people who have kindly reviewed and commented on various draft versions of this article: Nicole Santiago and Zachary Goldberg from Trilateral Research, Ana Valdivia (King's College London), Charalampia (Xaroula) Kerasidou (Lancaster University), and other SIENNA stakeholders (see Acknowledgment section in SIENNA Deliverable 5.4: "Multi-stakeholder Strategy and Practical Tools for Ethical AI and Robotics", February 2021).

Current AI ethics and the risk of abstraction

The principled approach to ethics in AI ethics

The numerous AI ethics initiatives, documents and guidelines produced over the past few years have primarily taken the shape of high-level abstract and prescriptive principles, such as "Ethics guidelines for trustworthy AI" of the High-Level Expert Group on AI (AI HLEG) set up by the European Commission, the "Recommendation of the Council on Artificial Intelligence" by the Organisation for Economic Co-operation and Development (OECD), or on the industry side, the Google AI principles[10]. However, as Resseguier and Rodrigues have argued, these AI ethics initiatives are dominated by a principled approach, i.e., a "law-conception of ethics", which is a view on ethics that considers it as a *replica* of the law, i.e. a softer version of the law[11]. Resseguier and Rodrigues point to the risk of misusing ethics as a replacement for legal regulation, and in particular the risk of ethics washing. While legal regulation might come with hard lines that could restrict innovation, regulation through ethical principles and guidelines offer more flexibility and leeway; hence, they are favoured by industry actors[12]. An issue with this principled approach to ethics is that it leads to ethics being used as a weaker form of regulation by actors with interest in avoiding hard lines[13]. Additionally, this approach to ethics comes with another pitfall that needs to be addressed: *its abstraction and the risk of disconnection from social, political and environmental materialities*.

The ethical theory of the "ethics of care" or "care ethics" that emerged in the 1980s with the work of Carol Gilligan has identified a fundamental gap in the principled approach to ethics, a conception that the ethics of care has called "ethics of justice" or "principlism": its neglect of context and actual practices, i.e. its abstraction[14]. This neglect is not simply a side-effect of this approach; *it is one of its constitutive features*. Indeed, as the ethics of care shows, the principled view on ethics is primarily and fundamentally characterised by its gesture of abstraction from concrete situations. It is through this abstraction that it can develop the general and impartial principles that it relies on and seeks to promote. This is particularly clear in the "veil of ignorance" promoted in the ethics of John Rawls. This veil aims at hiding any elements that constitute a person's specific socio-political situation in the world (including race, gender, socio-economic status, nationality, etc.) in order to formulate judgements from an unbiased and impartial standpoint, what Thomas Nagel has called the "view from nowhere"[15]. Although this approach to ethics has its value, its abstraction also brings with it several blind spots that are particularly problematical in the context of AI, even more so when it is applied to a field such as AI which presents itself as supposedly immaterial.

AI immaterial narrative and AI ethics

There is a dominant narrative surrounding digital technologies that presents these technologies as intangible. The supposed dematerialisation of technology is a narrative that has existed before digitalisation; however, it has become particularly pervasive with digitalisation, and especially AI[16]. A telling example of this is the concept of the "cloud" that makes data storage appear as deprived of a physical existence and, therefore, of practical impacts on the society and the environment. However, as we know, this is far from being true. The "cloud" relies on giant data centres that consume massive amount of energy. Similarly, studies have shown that the training of AI systems requires a high level of energy consumption as well as resource extraction, both of which have significant environmental costs[17].

The situation of micro-workers in the AI ecosystem today provides another clear case that undermines the AI "immaterial narrative". Micro-workers are people who work for platforms, such as Amazon's Mechanical Turk, to prepare, label and verify the data that go into AI systems. The AI industry tends to hide the situation of these workers, pretending that all the hard work is fully automatised. However, this is fallacious: AI systems rely on the work of people around the world that are extremely poorly paid and are working with no social security protection[18]. Here as well, AI "immaterial narrative"

[10] Respectively: High-Level Expert Group on Artificial Intelligence, "Ethics guidelines for trustworthy AI", European Commission, Brussels, 2019, https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1; OECD, "Recommendation of the Council on Artificial Intelligence", adopted on 22 May 2019, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449; Google, Artificial Intelligence at Google: Our Principles, https://ai.google/principles/. For listings of AI ethics guidelines and other guidance documents, see references in footnote 1.

[11] Anscombe, Gertrude, E., M., "Modern moral philosophy", *Philosophy*, Vol. 33, Issue 124, 1958, pp. 1–19; Resseguier, Anais, and Rodrigues, Rowena, "*AI ethics should not remain toothless?* A call to bring back the teeth of ethics", *Big Data & Society*, July-Dec 2020, pp. 1–5.

[12] Benkler, Yochai, "Don't Let Industry Write the Rules for AI", *Nature* 569, 2019, p. 161.

[13] In addition, Mittelstadt points to issues of ineffectiveness the approach of ethics through codes and lists of principles in Mittelstadt, Brent, "Principles Alone Cannot Guarantee Ethical AI", *Nature Machine Intelligence* 1, November 2019, p. 504

[14] Gilligan, Carol, *In a Different Voice: Psychological Theory and Women's Development*, Harvard University Press, Cambridge, MA, 1982.

[15] Nagel, Thomas, *The View from Nowhere*, Oxford, Oxford University Process, 1989.

[16] Izoard, Celia, "Les Réalités Occultées Du 'progrès' Technique : Inégalités et Désastres Socio-Écologiques", *Ritimo*, no. 21, May 2020, pp. 27–33. See also Campolo, Alexender, and Kate Crawford, "Enchanted Determinism: Power without Responsibility in Artificial Intelligence", *Engaging Science, Technology, and Society*, Vol. 6, 2020, pp. 1–19. In this article, Campolo and Crawford point to a discourse surrounding deep learning systems as one of "exceptional, enchanted, otherworldly and superhuman intelligence" (p. 9).

[17] On the environmental impacts of AI: Strubell, Emma, Ananya Ganesh, and Andrew McCallum, "Energy and Policy Considerations for Deep Learning in NLP", *arxiv.org*, 2019. https://arxiv.org/abs/1906.02243; Crawford, Kate, and Joler, Vladan, "Anatomy of an AI System: The Amazon Echo as An Anatomical Map of Human Labor, Data and Planetary Resources", AI Now Institute, September 2018.

[18] Tubaro, Paola, Antonio Casilli, and Marion Coville, "The Trainer, the Verifier, the Imitator: Three Ways in Which Human Platform Workers Support Artificial Intelligence", *Big Data & Society*, Jan-June 2020, pp. 1–12.

hides highly problematical impacts on the society, in this case those related to working conditions and employment rights.

Hence, this "immaterial narrative" that serves to pretend that AI, because intangible, is deprived of questionable social or environmental impacts, is highly problematic. However, because of its neglect of situatedness and materialities to reach high-level abstract principles, the dominant approach in AI ethics today renders it ineffective at properly addressing these negative material implications of AI. In other words, the disconnected nature of AI ethics today does not allow an appropriate response to some of the key critical issues that AI raises for society and individuals[19]. It is therefore necessary to complement it with methods to ensure AI ethics turns to concrete practices, considers the socio-political context and materialities, and addresses impacts.

AI ethics and the neglect of structural inequalities

This need for AI ethics to engage with the socio-political reality is even more necessary to address a major issue of AI: the risk of this technology further entrenching already existing structural inequalities. As Klein and D'Ignazio have shown in *Data Feminism,* AI ethics at present appear inadequate to properly respond to this challenge. According to them, data ethics is a field that relies on "concepts that secure power", and that, as such, "maintain the current structure of power"[20]. This is highly problematic as it renders AI ethics unable to address the root causes of one of the most significant ethical issues of AI: biases and social inequalities, especially those pertaining to race and gender. Studies have shown that the most vulnerable populations are those who face the most problematic impacts of AI (and often who are subject to the deployment of AI technology without choice or ability to influence its design and development), while, on the contrary, those who benefit the most from AI are those who hold positions of power in relation to this technology[21]. For instance, the study "Gender Shades" shows this clearly, demonstrating the problematic consequences of producing facial recognition systems trained with white males as the norm[22]. The consequence is a technology that works best for white males, but poorly for black females, white females and black males lying in between. Here as well, the "point of view of nowhere" that characterises AI and AI ethics fails to offer a proper response to key issues in the field of AI, particularly those related to structural inequalities and injustice[23]. Hence, AI ethics needs to be complemented in a way that enables it to address and respond to these issues.

There is an interesting parallel to be drawn between (1) the criticism toward AI ethics as being elaborated from the perspective of the privileged members of society and (2) the criticism formulated by the ethics of care toward what it identified as the dominant form of ethics (the ethics of justice). On the side of the critique toward AI ethics, Klein and D'Ignazio point to the "privilege hazard" which they define as "the phenomenon that makes those who occupy the most privileged positions among us—those with good educations, respected credentials, and professional accolades—so poorly equipped to recognize instances of oppression in the world", i.e. an "ignorance of being on top"[24]. The ethics of care highlights a similar situation characterised by the "indifference" of those who are privileged, i.e. those who occupy positions of power[25].

Thus, the dominant form of ethics in current AI ethics is *primarily a principled one characterised by abstraction*. This approach to ethics makes AI ethics *poorly equipped to respond to some of the key issues of AI related to social, political and environmental materialities*. Therefore, the ethics of AI as it has been developed over the past few years *needs to be complemented by other means to ensure proper consideration of actual practices, situations and socio-political context*.

For an ethics of attention to context, situatedness and materialities

The second section pointed to one of the key risks of the dominant understanding of ethics in current AI ethics discourses and initiatives: its abstraction leading to a potential neglect of problematical social, political and environmental impacts of AI. The third section presents how this approach can be best complemented to respond to the identified limitation. It does so by promoting a different approach to ethics, one that *invites to a sharp attention to context, situatedness and materialities*. This section also responds to feedback from SIENNA's stakeholders and other AI experts who have repeatedly called for making AI ethics more practical and usable for organisations and developers[26].

Ethics as attention in situation

This section argues for the *need for AI ethics to shift attention away from high level abstract and prescriptive principles*

[19] Another challenge that contributes to this issue of AI ethics failing to engage with the materiality of AI is that ethical evaluations of AI projects, when they happen, often take place at only one point in time with the ethics review at the start of a project. It is essential to do a proper ethical evaluation of an AI system that it is framed as an ongoing reflexive process, not simply a one off. Thank you to reviewer Nathalie Grandjean for highlighting this point.

[20] D'Ignazio, C. and Klein, L., op. cit., 2019, p. 60 and p. 61.

[21] See for instance Jansen, Philip, Philip Brey, Alice Fox, Jonne Maas, Bradley Hillas, Nils Wagner, Patrick Smith, Isaac Oluoch, Laura Lamers, Hero van Gein, Anais Resseguier, Rowena Rodrigues, David Wright, David Douglas, Ethical Analysis of AI and Robotics Technologies, SIENNA D4.4, Aug 2019.

[22] Buolamwini, Joy, and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification", in *Conference on Fairness, Accountability, and Transparency*, 81, New York: PLMR, 2018, pp. 1–15. See also: Birhane, Abeba, and Cummins, Fred, "Algorithmic Injustices: Towards a Relational Ethics", arXiv: 1912.07376, 2019.

[23] In "Race and Gender", Gebru condemns this "point of view of nowhere" in AI and its negative impacts on the most vulnerable and marginalised populations. Gebru, T., op. cit., 2020.

[24] D'Ignazio, C. and Klein, L., op. cit., 2019, p. 29 and p. 28 (respectively).

[25] Molinier, Pascale, "De la civilisation du travail à la société du Care", *Vie Sociale* 14, no. 2, 2016, p. 138.

[26] Workshop on the analysis of present and future ethical issues in AI and robotics, Uppsala (Sweden), 13–14 June 2019 and Online workshop on Multi-stakeholder Strategies for Ethical AI, 8–9 September 2020.

to practical contexts and relations. This shift is one of the lessons learned in bioethics, a field that was at first heavily dominated by high level principles, such as in the famous *Principles of Bioethics* by Beauchamp and Childress that derive answers to ethics challenges from four basic principles (autonomy, beneficence, non-maleficence, and justice). This approach has been challenged for being too abstract, top-down and insufficiently attentive to particulars[27]. Aren't we reproducing the same issue in AI ethics today as bioethics in the 1980s with the dominating principled approach (or "principlism")? As indicated by a participant to the SIENNA workshop on multi-stakeholder strategies for ethical AI (Sept 2020), AI ethics should draw from lessons learned in bioethics[28].

The ethics of care has developed resources to move beyond the principled approach to ethics. It has called for ethics to "modify its field: from an enquiry into general concepts to the study of particular situations, individual's moral configurations"[29]. It moves away from the "view from nowhere" that characterises the principled approach and invites to a sharp attention to concrete practical reality, situations and relations.

This attention to specific situations is more generally a key contribution from feminist theory. As Klein and D'Ignazio put it: "one of the central tenets of feminist thinking is that all knowledge is situated"[30]. The recognition of the situatedness of knowledge is a particularly essential recognition for the fields of AI and data science, fields that are often "framed as an abstract and technical pursuit" and that, as such, leave aside the "social context, ethics, values, or politics of data"[31]. Data science has the tendency to pretend it is always the same regardless of its context of application, whether it is astrophysics, criminal justice or carbon emission[32]. This is deceptive and the ethics of AI needs to be able to challenge and address this fallacious claim. To do so, it needs to advocate for attention to social, political, environmental context and materialities[33].

For instance, when dealing with criminal data, it is essential for AI ethics to recall the historical structures of injustices and inequalities through which these data have been produced[34]. Similarly, when handling health data, one should have in mind that white males are significantly over-represented in such sets of data, leading to the risk of poor quality healthcare for women and the non-white population[35]. The case of the machine learning technique of word embedding is another telling example of the need to consider the social background, as Bolukbasi *et al*. show clearly in their article "Man is to Computer Programmer as Woman is to Homemaker?"[36]

Ethics defined as attention means attention to socio-political realities and materialities at, at least, two different levels: (1) that of the development of an AI system (both from the production of the algorithm and the datasets that were used for the training of the algorithm) and (2) that of the context of application and intended use. The proposal formulated by Gebru *et al.* documenting the "motivation, composition, collection process, recommended uses, and so on" of a database used for machine learning would be particularly useful in that regard[37]. Another expression of ethics as attention to context is developed by Asaro in an article in which he calls for an ethics of care approach to AI ethics in predicting policing. He highlights the need for AI ethics "to seek likely ways in which political, economic, or social pressures may have influenced historical datasets, to consider how it [sic] may be shaping current data collection practices, and to be sensitive to the ways in which new data practices may transform social practices and how that relates to the communities and individuals a system aims to care for"[38]. To do so, he recommends the inclusion of "domain experts as well as critical social scientists as members of design teams" and the recognition of the "necessity of their expertise in shaping ultimate system design"[39].

Ethics as attention to relations of power

Another crucial insight from the ethics of care that AI ethics can draw on is the need to pay closer attention to *relations*

[27] Arras, John, D., "The Way we Reason Now: Reflective Equilibrium in Bioethics", in Bonnie Steinbock, *The Oxford Handbook of Bioethics*, Oxford University Press, Oxford, 2009.

[28] For an excellent comparative analysis of AI ethics versus medical ethics, see Mittelstadt, Brent, op. cit., 2019.

[29] Molinier, Pascale, Sandra Laugier, and Patricia Paperman, *Qu'est-ce que le care? Souci des autres, sensibilité, responsabilité*, Payot & Rivages, Paris, 2009, p. 23. Authors' translation.

[30] D'Ignazio, C. and Klein, L., op. cit., 2020, p. 152.

[31] *Ibid.*, p. 66.

[32] *Ibid.* Similarly, Campolo and Crawford have pointed to the "epistemological 'flattening' of complex social contexts into clean 'signal' for the purposes of prediction" that machine learning brings about. Campolo and Crawford, op. cit., 2020, p. 10.

[33] By "context", we mean more generally the field of application of the AI, i.e., the context in terms of place, time and culture, while by "materiality", we mean more specifically the physical dimension of AI, e.g., devices, data centers, pipelines, transportation networks, and CO2 emissions that come into play in the development and use of AI.

[34] For a famous study of how AI systems used in the criminal systems have led to further entrenching already existing structures of inequalities in the US context, see: Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner, "Machine Bias", *ProPublica*, May 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[35] For a general study on the poor representation of women in data see: Criado Perez, Caroline, *Invisible Women. Exposing Data Bias in a World Designed for Men*, London, Chatto & Windus, 2019.

[36] Bolukbasi, Tolga, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai, "Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings", ArXiv.Org, 2016. Word embedding is a machine learning technique that represents text data as vectors.

[37] Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daume III, and Kate Crawford, "Datasheets for Datasets" *ArXiv.Org*, 19 March 2020. https://arxiv.org/abs/1803.09010.

[38] Asaro, Peter M, "AI Ethics in Predictive Policing. From Models of Threat to an Ethics of Care" *IEEE Technology and Society Magazine*, June 2019, p. 50.

[39] *Ibid*.

of power and inequalities. When the ethics of care argues that ethics is a matter of attention to situations, this includes relations. It is important to note that these relations do not only refer to interpersonal relationships, although these are at stake as well. It also refers to broader relations in the society, including relations of power, i.e. power asymmetries between different social groups. This is one of the key contributions of the ethics of care to ethical theory. It is also a call for realism. As Laugier puts it: "Care takes us back to this requirement of realism in the sense of the need to see what is in front of our eyes: the reality of inequality before the idealness of principles"[40]. This is particularly essential in the context of AI considering the high economic interests involved and how these come to shape policy agenda on the regulation of AI.

Several experts have shown that the dominant form of ethics has failed to pay sufficient attention to the power imbalance at stake in the discussions on the regulation of AI, especially with regards to the concentration of power in the hands of a few big tech companies. For instance, Wagner shows how ethics has been used "as an escape from regulation"[41]. Article 19, a non-governmental organisation, argues that ethics initiatives have often "proven to be a strategy of simply buying time to profit from and experiment on societies and people, in dangerous and irreversible ways."

In the face of this, it is essential for ethicists to (1) avoid developing tools and resources that may serve this form of misuse of ethics, and (2) combat these misuses. To do so, ethicists need to be clear about the power relations and economic interests at stake. For instance, with these interests at stake, they need to consider when self-regulation is a meaningful and potentially effective response and when it is not. For instance, Access Now, another non-governmental organisation, has demonstrated in a recent report, "taking an 'ethics'-based approach to facial recognition and other dangerous applications of AI would leave millions exposed to potential human rights violations, and with little to no recourse." As a SIENNA stakeholder put it: we need to be clear on what we can realistically expect from ethics. This is even more important bearing in mind the power of the big technology companies, especially GAFAM (Google, Apple, Facebook, Amazon, and Microsoft)[42].

Paying attention to relations of power in the AI field also entails considering the severe diversity issue in the AI community. The 2019 report by the AI Now Institute "Discriminating Systems. Gender, Race, and Power in AI" powerfully points to this issue and calls on the AI industry to "acknowledge the gravity of its diversity problem"[43]. This "diversity problem" is not only an employment issue; it has direct implications on the type of AI systems produced. As West *et al.* write, "these patterns of discrimination and exclusion reverberate well beyond the workplace into the wider world", i.e. they lead to the development of technological products that are more beneficial to males than females or non-binary, to white rather than non-white[44]. Once again, this is essential to take into account for an AI ethics that directly engages with its situatedness and develops adequate tools to ensure AI systems are assessed with their social, political, and environmental impacts in consideration, so that the negative ones may be properly addressed and mitigated.

Enabling ethical agency

Hence, this essay argues that 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. But a question then arises: how can there be any ethics guidance, if ethics is primarily a matter of attention to specific situations? In other words, isn't this approach to ethics as attention rendering any recommendations inadequate as they would take away from the specificities of the context? Although this essay argues that ethics is primarily a matter of attention to specific situations, guidance is still needed. However, the status of such ethics guidance needs to be specified (the guidance formulated in this piece is precisely of this nature).

The recommendations formulated here do not aim to say what one should do or should not do, but rather aim to *promote the conditions of possibility of doing ethics, i.e., of being able to identify the right course of action from within a particular situation*. In that sense, the form of ethics that is promoted here does not determine the 'right' or the 'good' from a distanced position; it is not prescriptive or imperative. Rather, it aims to provide tools and resources to ensure actors, in their situated position – such as a developer in an organisation, within her own role, position, organisation, socio-political-environmental context, and confronted with a particular engineering challenge – can make ethical choices. Rather than determining from a distanced position what the ethical thing to do is, it should be ensured *actors are in a position to determine what is the right thing to do*. To use the terminology proposed by Canguilhem, it is not a matter of determining norms, i.e.

[40] The author's translation. Original French language: "Le care nous ramène a cette exigence de réalisme, au sens de la nécessité de voir ce qui est sous nos yeux: la réalité de l'inégalité, avant l'idéalité des principes." Laugier, Sandra, "Le care comme critique et comme féminisme", *Travail, genre et sociétés*, vol. 26, no. 2, 2011, p. 185–186.

[41] Wagner, B., op. cit., 2019

[42] As we write this essay (Dec 2020), there are ongoing policy initiatives in the EU to regulate big technology companies. These companies are fighting hard to have a seat at the table of discussions in order to limit the extent of these regulations as much as possible. Satariano, Adam and Stevis-Gridness, Matina, "Big Tech Turns Its Lobbyists Loose on Europe, Alarming Regulators", *The New York Times*, 14 Dec 2020. https://www.nytimes.com/2020/12/14/technology/big-tech-lobbying-europe.html

[43] West, Sarah Myers, Whittaker, Meredith, and Crawford, Kate. 'Discriminating Systems. Gender, Race, and Power in AI'. AI Now Institute, 2019. https://ainowinstitute.org/discriminatingsystems.html. p. 3. Hagendorff has analysed AI ethics guidelines and notes that only 37,1% of these have female authors, this number comes down to 7.7% for the guidelines developed in the FAT ML community. Hagendorff, T., op. cit., 2019.

[44] West, S. *et al.*, 2019,

the right thing to do, but of developing, actors' "*normative capacity*", which we can also define as ethical agency[45].

This approach avoids the risk of ethics guidance being patronising. This is one issue that was raised by a SIENNA stakeholder from a big technology company: the risk of the ethicist "coming on high horse" with predetermined ideas of what engineers should do. As Miller and Coldicutt show, technology workers are sensitive to the impacts of their products: "79% agree it's important to consider potential consequences for people and society when designing new technologies"[46]. They have called this capacity "personal moral compass"[47]. For AI ethics, it is essential to recognise this already existing sensitivity to issues and ability to respond to ethical challenges. However, as participants in the SIENNA workshop on multi-stakeholder strategies for ethical AI highlighted, this is not sufficient. This ethical sensitivity and ability need to be further enhanced, promoted and protected. This is precisely the objective of the practical recommendations below. Promotion and protection of ethical sensitivity and ability can take various shapes. One of these is by conducting *impact assessment* of AI technologies and products. Miller and Coldicutt point to the fact that 81% of people in AI "would like more opportunities to assess the potential impacts of their products"[48]. These can also take the shape of operationalised guidance documents, such as those developed in SIENNA with the ethics by design methodology, or research ethics guidance documents.

It is essential to clarify that these are only tools to promote ethical sensitivity and the ability to make the right decision. Human expertise and sense of the specificity of the situation at stake (e.g. *this* AI system developed in such and such context with such and such objective) remains essential. In other words, high-level principles, norms and guidance, although they can help provide the broad lines, cannot fully dictate what should be done and what should not be in a specific situation. There is always, necessarily, the need to pay attention to the specificities of a particular situation. Therefore, for instance, in addition to research ethics guidelines, research ethics also needs human expertise in order to ensure the guidelines (necessarily general) are properly applied and interpreted with a specific situation (necessarily particular). The EUREC[49]/SIENNA online workshop on 26–27 October 2020 that brought together members of European research ethics committees and SIENNA consortium partners on the topic of research ethics guidelines made this clear: expertise in interpreting and applying ethics guidelines to specific research cases is essential to research ethics[50].

Another way of ensuring AI products/technologies are developed with care for their socio-politico-environmental impacts, is through the *protection of whistle-blowers and unions*. The need to protect whistle-blowers was mentioned by a participant in the Multi-Stakeholder Strategies for Ethical AI workshop on 8–9 September 2020. The protests by Google employees against the Maven project, in which Google was developing AI technology for the US military drone programme, is a strong example of this. Google employees saw issues with a powerful technology company such as Google getting into "the business of war"[51]. Considering ethics implies asking hard questions, and those who have the courage to do so should be protected. In other words, if we want workers "to be vectors of change", they need to be able to raise concerns when they see something that is problematic in their organisation and be protected appropriately from the ensuring fall out.

Conclusion

To conclude, this essay has shown that current AI ethics guidance and initiatives tend to be dominated by a principled approach to ethics. Although this brings value to the field, it also entails some risks, especially in relation to the abstraction of this form of ethics that makes it poorly equipped to engage with and address deep socio-political issues and practical impacts. As such this essay has sought to complement the existing principled approach to ethics with an approach to ethics as attention to context and relations. Below are some more practical recommendations to promote ethical AI by drawing from an approach to ethics as attention[52].

Practical recommendations for AI ethics

AI ethicists should:

→ **Engage with social scientists and their research on social impacts of AI** in the short, medium and long term

[45] Canguilhem, Georges, *The Normal and the Pathological*, Zone Books, 1991. This notion of "normative capacity" shares commonalities with that of "capability" commonly discussed by Amartya Sen and Martha Nussbaum. However, the value of Canguilhem's notion for this article is the direct link it makes with norms and the production of norms. This is why it is used here instead of the notion of "capability".

[46] Miller, Catherine, and Rachel Coldicutt, "People, Power and Technology: The Tech Workers' View," London, Doteveryone, 2019, p. 16. https://doteveryone.org.uk/report/workersview.

[47] *Ibid.*

[48] *Ibid*., p. 19

[49] EUREC is the European Network of Research Ethics Committees: http://www.eurecnet.org. EUREC Office is a partner organisation in the SIENNA project.

[50] The ethics guidelines the workshop focused on included guidelines identified as the most important ones used by research ethics committees in Europe.

[51] Wakabayashi, Daisuke, and Shane, Scott, "Google Will Not Renew Pentagon Contract That Upset Employees", The New York Times, 1 June 2018. https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html. More recently, the AI ethics researcher Timnit Gebru has been forced out by her employer at Google for raising a number of ethical issues related to AI technology used by the company. Hao, Karen, "We read the paper that forced Timnit Gebru out of Google. Here's what it says", *MIT Technology Review*, 4 Dec 2020. https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/

[52] Please note that several of these recommendations are relevant to diverse groups and are therefore repeated. Additionally, further work would be needed to operationalise them more specifically and concretely in practical fields of applications and by different actors. This list seeks to give a flavor of the many different ways the approach of 'ethics as attention to context' proposed in the article can help make the ethics of AI more effective than the principled approach currently dominating and how it can be applied by different actors of the field. Developing these recommendations further would take us beyond the objective and scope of this article.

→ **Engage with data scientists and software engineers** to better understand the way AI systems are developed and the data collected, cleaned, processed and interpreted

→ **Draw from ethics theories beyond principlism**, especially the ethics of care, virtue ethics, or Spinozist ethics.

→ **Ensure diversity** in the composition of AI ethics teams (especially encouraging inclusion of non-white females and non-binary individuals).

→ **Pay attention to the "privilege hazard"**, i.e. the risk of people in position of privilege failing to notice instances of oppression and injustices perpetuated by AI technologies

→ **Recognise, build upon and further develop AI developers' ethical sensitivity** (rather than impose guidance that may be top-down and disconnected from practice)

→ **Develop/share use cases on ethical and social impacts of AI** (especially negative ones) to make the impacts of AI more concrete and understandable

→ **Engage more with the impacted communities, especially the most vulnerable among them,** and consider social and ethical impacts from their perspective

→ Take into consideration the **environmental impacts of AI** (throughout the whole life cycle, from resource extraction to end-of-life disposal)

→ When carrying out impact assessments, consider **impacts on labour**, especially the conditions of "micro-workers" and other precarious workers in the AI industry

The AI industry should:

→ Engage with **social science studies** on the social and material impacts of AI

→ Promote **inclusion of social scientists** in AI research and development projects

→ Recognise that a technological product is **never "neutral"** and always reflects specific worldviews and intentions

→ Conduct/require **assessments of the ethical, social, environmental, and human rights impacts of AI** before, during and after deployment of AI systems

→ **Encourage prudence and honesty about the capabilities of the AI system being sold and do not gloss over the limitations**

→ **Ensure diversity** in the composition of the team researching, developing, using, or assessing AI (especially encouraging inclusion of non-white females and non-binary).

→ **Pay attention to the "privilege hazard"**, i.e. the risk of people in position of privilege failing to notice instances of oppression and injustices perpetuated by AI technologies

→ **Provide a safe environment for whistle-blowers and union members** in the AI industry

→ Develop and ensure respect for **research ethics in the private research and development (R&D) AI sector.**

→ Promote and create r**esearch ethics committees/AI ethics officers** for AI R&D in the private sector

→ **Engage more with impacted communities, especially the most vulnerable among them,** and consider social and ethical impacts from their perspective

→ **Recognise that not all social problems can be solved with technology**, i.e. be wary of technosolutionism

AI researchers and developers should:

→ Engage with **social science studies** on the social and material impacts of AI

→ Promote the **inclusion of social scientists** in AI projects

→ Recognise that a technological product is **never "neutral"** and always reflects specific worldviews and intentions

→ Conduct/require **assessments of the ethical, social, environmental, and human rights impacts of AI** before, during and after deployment

→ **Consider the possibility of *not* developing an AI system** when the ethical and social issues identified are too severe and difficult to mitigate. Also account for the remaining uncertainties surrounding the ethical and social impacts of AI.

→ **Ensure diversity** in the composition of the team developing, using or assessing AI (especially encouraging inclusion of non-white females and non-binary individuals).

→ **Pay attention to the "privilege hazard"**, i.e. the risk for people in position of privilege to fail to notice instances of oppression and injustices perpetuated by AI technologies

→ **Report on the development of AI technologies which are harmful to society and violate human rights,** and report incidents to AI incident databases and regulators or civil society organisations, particularly where no impact assessment of such technologies has taken place.

→ **Engage more with the impacted communities, especially the most vulnerable among them,** and consider social and ethical impacts from their perspective

→ **Recognise that not all social problems can be solved with technology**, i.e. be wary of technosolutionism

Policy-makers engaged in AI regulation and governance should:

→ **Not let the AI hype hide the ethical and social issues this technology brings about and the difficulty to solve them** (e.g. exacerbation of structural inequalities, privacy risks and surveillance) as well as the remaining uncertainties on the ethical and social impacts of AI

→ **Protect whistle-blowers and unions** in the AI industry and develop adequate protection mechanisms and safeguards where this is missing.

→ **Encourage research ethics in private-sector AI R&D**

→ **Encourage the creation/use of research ethics committees** for AI R&D in the private sector

→ Require/encourage/mandate the conduct of **assessments of the ethical, social, environmental, and human rights impacts of AI** before, during and after deployment.

→ **Engage more with the impacted communities, especially the most vulnerable among them,**and consider social and ethical impacts from their perspective

→ **Recognise and promote awareness of the fact that not all social problems can be solved with technology**, i.e. be wary of technosolutionism

→ **Be wary of the AI hype**

AI public research funding organisations should:

→ **Fund cutting-edge social science studies** on AI and its impacts in the short, medium, and long term.

→ Require/encourage the conduct of **assessments of the ethical, social, environmental, and human rights impacts of AI** research and development

→ **Engage more with impacted communities, especially the most vulnerable among them,** and consider social and ethical impacts from their perspective

→ **Recognise and promote awareness of the fact that not all social problems can be solved with technology**, i.e. be wary of technosolutionism

→ **Be wary of the AI hype**

AI STEM (science, technology, engineering, mathematics) educators should:

→ **Train AI researchers and developers** on the non-neutrality of AI and familiarise them with AI ethical and social impacts and impacts on human rights

→ **Develop/share use cases on the ethical and social impacts of AI** (especially negative ones) to make the impacts of AI more concrete and understandable

→ **Recognise and promote awareness of the fact that not all social problems can be solved with technology**, i.e. be wary of technosolutionism

Data availability

No data are associated with this article.

Open Peer Review

Current Peer Review Status: ? ✓ ?

Version 1

Reviewer Report 06 October 2021

https://doi.org/10.21956/openreseurope.14330.r27581

© **2021 Corteel M.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Mathieu Corteel
Laboratoire interdisciplinaire d'évaluation des politiques publiques (LIEPP), Sciences Po, Paris, France

The article develops a materialist perspective to consider the socio-political relations through the ethics of Artificial Intelligence (AI). Against the ideological discourses of AI promoters (immaterial narrative), the authors adopt a realist position based on attention to context.

They emphasize the fact that the ethics of AI is always to be remade according to social, political and environmental aspects. In order to improve the ethics of AI, they suggest shifting from a top-down ideological approach to a bottom-up materialist approach.

The main merit of the article is therefore to highlight the ineffectiveness of abstract "ethical" principles promoted by AI industry. However, the article takes the form of a statement of intent rather than a useful tool for implementing a new AI ethics.

1. The exhaustive list of “practical recommendations”, at the end of the article, suffers from the same defect as abstract and formal theories of ethics: it proposes a list of moral precepts and recommendations (ensure diversity, inclusion of social scientist, take into consideration the environmental impact...) without methods of implementation. The authors are, in a way, victims of what they denounce.

2. The ethics of care is invoked as a guarantor of effectiveness against discrimination and concentration of power, but it is not explicitly detailed as a pragmatic way to adapt AI ethics to concrete contexts.

3. There is no specific evidence in the article demonstrating the relevance of materialist ethics or ethics of care to improving the practices of data science, data-driven decision making, AI design or social applications of AI. How do we make ethics an implemented capability in the practice of citizens, engineers and organizations developing and using AI?

In order to improve the article, I suggest that the authors:

1. Explain more specifically how to implement the “practical recommendations” through effective ethical capabilities in the design and use of AI. This includes case studies that

demonstrate the relevance of these recommendations to ethically improving decision and action in context.

2. Detail the pragmatic basis and method of their materialist perspective by considering what the practice of ethics is in an AI application setting.

3. Clarify the need for interdisciplinarity to improve the design and adaptation of AI to diverse social and cultural contexts.

4. Indicate which disciplines and methods are most relevant to participate in this heuristic innovation.

Is the topic of the essay discussed accurately in the context of the current literature?
Yes

Is the work clearly and cogently presented?
Yes

Is the argument persuasive and supported by appropriate evidence?
No

Does the essay contribute to the cultural, historical, social understanding of the field?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Philosophy of Science, History of Science, Medical Ethics, Ethics of Technology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 26 Nov 2021

Anais Resseguier

Many thanks for the detailed review and comments. Please find below a response to each of them. On the first 3 comments:

1. There might be some confusion regarding the scope of this article. We added a note on Endnote 50 to clarify and hope that this helps sort this out.

2. We are not sure what might have given the impression that we 'invoke' the ethics of care as 'a guarantor of effectiveness against discrimination and concentration of power'. The contribution of the ethics of care here is that it makes it possible to approach the ethics of AI differently from the currently dominating approach which is characterised by "principlism". The ethics of care gives theoretical tools to challenge this dominant approach and to develop a new approach (new approach that is then

complemented by more practical recommendations). As indicated in the paper, we see the ethics of care as having 'developed resources to move beyond the principled approach to ethics', and the 'view from nowhere' that characterizes this approach. We see value in the ethics of care as it invites to 'pay closer attention to *relations of power and inequalities*.'

3. This comment might relate to the overall objective and scope of this article. The main objective of this article is twofold: 1. Critically present the current landscape of AI ethics and where/how we think it might be failing/ineffective (this is done in 1st section). 2. Provide a different approach to thinking about the ethics of AI (through ethics as attention to context) (this is done in the 2nd section). The Practical recommendations at the end highlight some practical steps to operationalise this different approach to ethics. Hence, we believe we started looking into the implementation of this approach in this last section; however, it is not the core objective of this article and is therefore not developed further. Developing this further would require another article, that could draw on the theoretical foundations laid down in the present article.

Regarding the suggestions to improve the article: Points 1 and 2: Here too, we see this as being outside the scope of this article. We trust the responses above will help clarify the scope. A case study would certainly be highly valuable but falls outside the scope of the present article. Points 3 and 4: We are slightly unsure what the suggestions are about and what they try to achieve. The article is interdisciplinary and draws from philosophy, ethics and social sciences (as made obvious in this paper). It also draws from discussions with experts coming different disciplines consulted as part of the SIENNA Horizon 2020 project, as explicitly stated in the paper. Hence, we do not see the need for further clarifications on this in the article.

Competing Interests: No competing interests were disclosed.

Reviewer Report 06 October 2021

https://doi.org/10.21956/openreseurope.14330.r27579

© 2021 Ienca M. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Marcello Ienca
ETH Zurich, Zürich, Switzerland

This article provides an overview of current ethics guidance documents and initiatives for artificial intelligence (AI). The authors argue that the principlist approach which currently dominates the AI ethics discourse is ill-equipped to address the deep socio-political issues and the material impacts of AI such as social inequalities, injustices and environmental damage. Drawing from alternative ethical theories such as the ethics of care and feminist approaches to ethics, the authors propose

to complement the existing principled approach with an approach to ethics as attention to context and relations. The article is well informed about the pertinent scientific literature, soundly argued and clearly written. I recommend indexing in its current form.

Is the topic of the essay discussed accurately in the context of the current literature?
Yes

Is the work clearly and cogently presented?
Yes

Is the argument persuasive and supported by appropriate evidence?
Yes

Does the essay contribute to the cultural, historical, social understanding of the field?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioethics, neuroethics, ethics of Artificial Intelligence, theoretical cognitive science and the philosophy of technology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 26 Nov 2021

Anais Resseguier

Many thanks for the review and the kind comments on the article.

Competing Interests: No competing interests were disclosed.

Reviewer Report 01 June 2021

https://doi.org/10.21956/openreseurope.14330.r26881

© 2021 Grandjean N. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Nathalie Grandjean
[1] CRIDS (Information Research Center, Law and Society), University of Namur, Namur, Belgium
[2] ESPHIN, University of Namur, Namur, Belgium

Generally I find the article excellent because it highlights an ethics of attention to context that is sorely lacking in AI ethics. However, I would like to address three remarks to the authors, in order

to clarify some points that seem obscure or insufficiently developed.

1. I fully agree with the argument of the narratives of digital technologies as immaterial; on the other hand, it seems to me that this immateriality is also at play in the fact that ethical evaluations of AI are formulated in project contexts, i.e. when AI technologies are in the process of being made... which also makes them evanescent, since they are not yet finished.

2. It seems to me that the paper would benefit from better clarification of the difference (both theoretical and practical) between context and materiality; as a reader, it is difficult to understand how the authors formally distinguish between them.

3. Finally, the authors mention the concept of "normative capacity" (Canguilhem) but do not mention the concept of "capability" (Sen, Nussbaum), commonly discussed in the literature. It seems to me that this choice should be clarified.

Is the topic of the essay discussed accurately in the context of the current literature?
Yes

Is the work clearly and cogently presented?
Yes

Is the argument persuasive and supported by appropriate evidence?
Yes

Does the essay contribute to the cultural, historical, social understanding of the field?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Ethics, philosophy of technique/technology, gender studies, ethics of care, ethics of attention.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 26 Nov 2021

Anais Resseguier

Thank you very much for the thoughtful review and feedback that helped clarify important aspects of the paper. Regarding the first point, see footnote added on p. 4 that acknowledges the particular context of application of AI ethics review, when there is any, and how this relates to the point raised about the immateriality of AI. Points 2 and 3 have now been clarified. See footnote on p. 6 and endnote 43.

Competing Interests: No competing interests were disclosed.