



In defense of ethical guidelines

Björn Lundgren^{1,2,3}

Received: 16 September 2022 / Accepted: 22 November 2022 / Published online: 17 January 2023
© The Author(s) 2023

Abstract

Recently, Luke Munn attacked “AI ethics” generally, or guidelines, principles, codes of ethics, ethical frameworks. In particular, he argued that ethical guidelines are useless. Here I respond to this critique, arguing that Munn’s criticism is mostly unfair and misguided, and that his own proposal is already implemented in various guidelines.

Keywords AI ethics · Ethical guidelines · Ethical principles · Operationalization of ethical principles

1 Introduction

Recently, Luke Munn argued that AI guidelines are useless. Although the title of his paper—“The uselessness of AI ethics”—seems to indicate that it is about AI ethics, his discussion and examples concern what we broadly call ethical guidelines (i.e., including ethical principles, codes of ethics, guidelines, frameworks, and so forth) [21].¹ In this article, I aim to defend ethical guidelines. Although I grant that some of Munn’s criticism is warranted for some guidelines, I believe that he draws the wrong conclusion. Ethical guidelines for AI development and use are in their methodological adolescence and, as I will argue, there are possible fixes to the problems that Munn identifies. Indeed, there are guidelines that are already responding to many of these worries, implying that Munn’s argument is partly based on cherry-picking. This is particularly problematic when it comes to Munn’s own *alternatives* to “AI ethics”, which simply knocks on an open door. Moreover, some of Munn’s critique is simply misguided.

The remainder of this article will be structured as follows. In the next section, I give a brief contextualization of AI ethics and ethical guidelines, aiming to explain how the former

can inform the latter, and how both can inform standards and regulations. In Sect. 3, I respond to Munn’s critique and his own proposal, aiming to show that ethical guidelines are already responding to the challenge and that the proposal is knocking on open doors. Lastly, I end the article with a summation and some brief comments on some of the main challenges for ethical guidelines, such as contextual variations, operationalizations, and enforcement.

In this commentary, I will rely heavily on a set of guidelines that in turn build on a large survey of guidelines [3], but see also [13] for a recent overview of ethical guidelines for AI. For many examples, I will cite from [3] because it builds on a large set of other guidelines (but I will also list other references when it is important).

2 AI ethics and ethical guidelines

In this section, I will briefly explain what ethical guidelines are and their relation to AI ethics, as well as how ethics and ethical guidelines can further inform ethical standards and regulations. Broadly speaking, “AI ethics” is a term used in various disciplines to denote the study of ethical aspects of AI. In ethics (a subdiscipline of philosophy), one normally distinguishes between normative ethics and applied ethics. The former is about theory building (what is right, wrong, permissible, impermissible, and so forth), while the latter is

✉ Björn Lundgren
b.a.lundgren@uu.nl; bjorn.lundgren@iffs.se;
bjorn.lundgren@philosophy.su.se

¹ Department of Philosophy and Religious Studies, Utrecht University, Utrecht, The Netherlands

² Institute for Futures Studies, Stockholm, Sweden

³ Department of Philosophy, Stockholm University, Stockholm, Sweden

¹ Arguably, despite his provocative title, this is the only way to take his arguments seriously. If we read Munn’s arguments as being about AI ethics, then his arguments just fall flat. For example, concerns about enforcement are not a matter for the study of ethics, since the aim is to answer questions such as what is right and wrong (see Sect. 2).

about ethical inquiries addressing a specific question (e.g., is Facebook’s algorithm controlling the newsfeed privacy invasive?) or questions about a specific domain (e.g., the study of ethical issues raised by AI and AI technologies). Simply put, AI ethics can be viewed as a field of inquiry that studies the ethical issues that are raised by AI and AI-technology. As I noted earlier (see footnote. 1), the theoretical nature of these endeavors makes it clear that if AI ethics were Munn’s target, then the critique is misguided because most of the critique is about practical enactment of policy, rather than a problem for theoretical studies.

However, AI ethics may—and arguably should—inform more practical endeavors, by supplying basic insight and constraints for such work. This is an example of when ethical guidelines are particularly relevant. Ethical guidelines are documents that supply action-guiding principles, which aim to ensure an ethical process or outcome in the appropriate area of application. For example, for AI one may focus on ethical guidelines for using and/or designing AI technology. Ethical guidelines for AI should take insight from AI ethics as a starting point, but more work must then be done to make the theoretical insights useful in practice.

While ethical guidelines have existed for a long time, ethical guidelines for AI are a relatively new endeavor (most of the AI guidelines are from the last five years or so—see, e.g., [3, 13]), so it is not surprising that one can—as Munn has—identify problems in some, or sometimes even most, of these guidelines. However, as I will argue, there are several guidelines that illustrate solutions to the problems Munn identifies and in many others we can see seeds of possible solutions, which will hopefully come to fruition as the methodology for developing ethical guidelines improves. As I will explain in the paper, one of the main challenges has to do with the operationalization of ethical principles, which is a fundamental methodological challenge. That is, one of the most fundamental challenges for ethical guidelines for AI is not first-order ethical inquiries (i.e., questions of what is right, wrong, permissible, impermissible, and so forth), but rather second-order problems regarding how we can translate known ethical principles into understandable goals. This is why I believe that Munn draws the wrong conclusion, even when his critique is on target. That is to say, given the very limited time-frame within which AI ethical guidelines have been produced in response to fairly complex problems, we cannot expect all solutions to be in place already. While ethical guidelines are doing a lot better than Munn would claim, they are still a work in progress, which is why I think the conclusion to draw is that we are in the methodological adolescence of ethical guidelines for AI. In the final section, I briefly discuss some of the main challenges.

Traditional ethical guidelines often made use of simple check-lists: do x, do y. However, when creating guidelines

for purposes that are more ethically challenging it may be difficult—or even impossible—to achieve a checklist that is jointly (1) understandable by non-experts and (2) applies in all relevant contexts. This is especially important for designing AI technology, which can be used in extremely different contexts, in which what is appropriate—for example, *viz-à-viz* the information handling of a system—may differ, for example, between a democracy and an oppressive regime [18, 19]). This is why some modern ethical guidelines have goals that require designers and users to make substantial choices within a limited set of ethical boundaries (I will return to this when addressing Munn’s critique in the next section).

Ethical guidelines are often based on a set of ethical principles, many of which largely overlap between different competing guidelines. For example, in a survey of 84 ethical guidelines for AI the following principles were identified in the majority of the guidelines (number of documents in parenthesis): transparency (73), justice and fairness (68), non-maleficence (60), responsibility (60), and privacy (47) [13], which illustrates that there is at least some international consensus emerging on basic principles.

An illustrative example of basic ethical principles can be found in [3], which also builds on a large set of ethical guidelines:

1. Human agency, liberty and dignity: Positive liberty, negative liberty and human dignity
2. Technical robustness and safety: Including resilience to attack and security, fall back plan and general safety, accuracy, reliability and reproducibility
3. Privacy and data governance: Including respect for privacy, quality and integrity of data, access to data, data rights and ownership
4. Transparency: Including traceability, explainability and communication
5. Diversity, non-discrimination and fairness: Avoidance and reduction of bias, ensuring fairness and avoidance of discrimination, and inclusive stakeholder engagement
6. Individual, societal and environmental wellbeing: Sustainable and environmentally friendly AI and big data systems, individual wellbeing, social relationships and social cohesion, and democracy and strong institutions
7. Accountability: Auditability, minimisation and reporting of negative impact, internal and external governance frameworks, redress, and human oversight

In [3], these high-level principles are the foundation for a set of over 70 requirements (some with underlying conditions) that specify in more detail what must be done to satisfy the higher level principles. For example:

Requirement 19: Clarify roles and responsibilities toward information use, security and privacy. In all phases (but especially in business understanding, data understanding, and data preparation), assess and ensure that:

- There are clear and precise descriptions of the roles and responsibilities of users toward information, media and network usage, security, and privacy;
- A common culture is established and encouraged that strongly promotes ethical behavior for all individuals in the enterprise, and establishes a low tolerance threshold for unethical behaviors.

This illustrates how ethical guidelines work on a much more practical level than AI ethics and how they must therefore be informed by, and adapt to, the empirical conditions (e.g., as can be seen in the requirement above, the requirement specifies in which development phases the conditions are especially important).

AI ethics can inform ethical guidelines for AI by supplying the basic fundamental principles, alternatives for analyzing, defining, and applying these principles, as well as a roadmap to the ethical challenges that must be addressed by an ethical guideline. Moreover, these processes can also inform ethical standards, which in turn can underpin regulations [26]. An illustrative example is the work done by the IEEE Standards Association, which created a discussion document [12] that served as a starting point for the creation of several standards (see [26] for a brief discussion of the process).² By creating standards, one enables a possible foundation for soft and hard law, or self-regulation, in case such standards are broadly accepted. Given that global standards have often been successful (e.g., consider the global smoking ban on airplanes) [15], this provides a strong argument for the practical value of ethical guidelines, irrespective of any criticism. Nevertheless, let us now turn to Munn’s specific criticism.

3 Responding to the critique and Munn’s proposal

In this section, I present and critically assess Munn’s critique and proposal in three subsections.

² See <https://ethicsinaction.ieee.org/p7000/> for the latest updates.

3.1 Meaningless principles and the principle/practice gap

The first form of criticism is that the “principles are highly abstract and ambiguous, becoming incoherent.” Munn diagnoses the problem as having to do with conceptual disagreement, which he thinks has the consequences that “Commendable values like ‘fairness’ and ‘privacy’ break down when subjected to scrutiny, leading to disparate visions and deeply incompatible goals” [21].

I will start by addressing the difficulty of creating principles based on concepts such as fairness and privacy. While it is true that there are well-known difficulties and conceptual disagreements about how to define, analyze, or conceptualize ethical concepts in general—and privacy, in particular—this is not necessarily a problem for creating a good ethical guidelines. Take the concept of privacy. Although there is substantial conceptual disagreement, there is arguably also a lot of agreement about what falls under the concept of privacy. There are a few potential ways to avoid problems caused by conceptual disagreement. First, one can build on existing frameworks (such as the GDPR). Second, one can focus on issues of agreement and apply requirements that fit with the most stringent conception in any given situation. Third, based on one of the two first choices one can require that a data protection policy is developed under certain specifications, that it is communicated internally, that there are safe-guards to ensure its enforcement, and that all of this is transparent to outsiders. Similarly, in the case of fairness, one can make similar demands on transparency on a choice of fairness measures and a method of testing and validation.

An illustrative example that follows part of my third suggestion is the following requirement from [3]:

Requirement 19: Clarify roles and responsibilities toward information use, security and privacy. In all phases (but especially in business understanding, data understanding, and data preparation), assess and ensure that:

- There are clear and precise descriptions of the roles and responsibilities of users toward information, media and network usage, security, and privacy;
- A common culture is established and encouraged that strongly promotes ethical behavior for all individuals in the enterprise, and establishes a low tolerance threshold for unethical behaviors.

Requiring precise descriptions of roles and responsibilities is a further step in setting a clear plan for how to enact privacy protections. Moreover, the second bullet point illustrates the possibility (and need) to deal with some ethical concerns by establishing a common culture.

Turning to the problem of incompatible goals, these are *prima facie* problematic given that they seem to give contradictory guidance and hence are not action-guiding. However, the *prima facie* presumption would be an erroneous conclusion. Guidelines with incompatible goals are not necessarily wrongheaded. Arguably AI ethics is an ethics of trade-offs and value-conflicts [17], which ethical guidelines should recognize. More importantly, ethical guidelines can include goals that are *prima facie* incompatible and still avoid incoherence and provide action-guidance. As previously mentioned, a sensible option is to require developers or users of AI systems to make choices, enact them, and communicate them. An illustrative example, which forces a choice between trade-offs, is the following from [3]:

Requirement 41: Trade-offs. In every phase, assess and ensure that:

- Trade-offs between explainability/transparency and best performance of the system are appropriately balanced based on the systems context of application (e.g., in healthcare the accuracy and performance of the system may be more important than its explainability; whereas, in policing, explainability is much more crucial to justify behaviors and outcomes of law enforcement; and in other areas, such as recruitment, both accuracy and explainability are similarly valued).

This illustrates how the guideline requires a choice in cases of trade-offs or value-conflicts, and it also provides some guidance by explaining what *prima facie* matters more in which situations.

The same set of guidelines also includes general communication requirements. Consider the following requirement for AI developers to ensure that “information to stakeholders, (end-)users, and other affected persons, about the system’s capabilities and limitations, is communicated in a clear, understandable, and proactive manner, that enables realistic expectation setting” [3].

Moreover, this is in line with a challenge that AI ethics, on the one hand, must be contextually sensitive, but, on the other hand, this creates difficulties for ethical guidelines [18]. For example, what context should we adapt our guidelines to? What if the context changes after the technology is adapted (e.g., perhaps the technology itself leads to contextual changes through disruption; see [11] for a discussion on socially disruptive technologies)? Recently, I argued that these types of worries can be resolved—as least for some applications—by adapting the ethical requirements to what is needed in the most challenging context [19]. For example, privacy-protections could be enacted that are sufficiently strong for the context in which the challenges to

individuals’ privacy are the most pervasive. This may raise concerns about overly demanding guidelines, but as I also suggest, we must also temper our ethical restrictions to not limit the balance between the risks and the opportunity for good-making features of the given technology. Moreover, for many applications privacy preservation does not create a problem for the features of the technology, but for the market model of the company that promotes it. For example, for many products one need not surveil user behavior to supply a good service. However, if the market model is built on personalized ads, certain forms of surveillance may be necessary to satisfy the market model.

When it comes to abstractness, ambiguity, or vagueness, Munn actually discusses the solution later in his article, when he turns to the gap between principles and practice. As he correctly notes, a problem with guidelines is that they often include fuzzy abstract higher-order principles, which is part of the reason why ethical guidelines should be operationalized (i.e., simply put, turning requirements into conditions that can be properly evaluated).

However, before turning to that issue, it is worthwhile to note a misunderstanding on Munn’s part. The critique of fuzzy high-order principles seems to miss the fact that the indeterminacy of high-order principles sometimes has a function in capturing a value that is technically impossible to satisfy in every specified way. Consider the example of fairness. If an AI-system is designed to identify a property, *P*, that is unequally distributed among two groups, and if it is not possible to perfectly predict the prevalence of *P*, then there are competing fairness measures that cannot all be satisfied. For example, either the property is identified more often in one of the groups, or the probability assessments that a singular individual has the property will differ between the groups [10, 16]. That is to say, two very sensible measures of fairness cannot, under the given circumstances, be mutually satisfied. In such cases, a choice must be made and it is not always the case that the same choice is suitable in every given context, which is why higher-order principles can sometimes be used to capture the overarching value, while the guidelines should enforce a choice as exemplified by the previously cited Requirement 41 from [3].

Nevertheless, Munn is correct in identifying that higher-order principles alone cannot give sufficient guidance. As argued above, in some cases, that is because we have competing values and trade-offs, which is an inherent part of AI ethics [17], but in other cases this is resolved by guidelines that also provide more detailed requirements (as exemplified in Sect. 2, some guidelines include both higher-order principles and more detailed specified requirements).

The solution to the problem of supplying more detailed requirements often depends on the operationalization of specific requirements; that is, making it possible for non-experts

(i.e., non-ethicists) to determine whether a requirement is properly satisfied. While Munn discusses operationalizations, he thinks that “operationalizing AI ethics promises to be difficult or even impossible, a daunting challenge underestimated by a technically focused industry and even by ethicists” [21]. However, it is not clear what Munn takes “operationalization” to mean. What he says seems to imply that a principle is operationalized only when we have a very precise technical solution to how it can be implemented: “Operationalization is not simply a perfunctory matter of ‘translating’ an ethical value into a technological outcome. There are tensions and trade-offs that must be worked through and worked out into the material form of a data model or a digital product.” [21]

However, the purpose of an ethical guideline is not to provide a complete design for a product. That is, in order for an operationalization to be workable the ethical guidelines do not need to include specific technical solutions (even if they can). As I mentioned above, all that is needed is that the requirements satisfaction conditions can be determined by a human being, so that it is non-ambiguous whether a specific criterion has been satisfied or not. That is to say, the guidelines must make it clear to the designer (or user) when they have satisfied an ethical requirement, but the guidelines do not need to supply a pathway for how to achieve that goal. The examples that I have already given are non-ambiguous in this sense. Consider again, for example, the requirement that “information to stakeholders, (end-)users, and other affected persons, about the system’s capabilities and limitations, is communicated in a clear, understandable, and proactive manner, that enables realistic expectation setting” [3]. Although such a demand leaves some room for interpretation (as anything would), it is clearly determinable, judging by a reasonable standard. (For specific guidelines that are at least partly operationalized, see, e.g., [3, 19, 20].)

Relatedly, Munn worries that these “are highly contested issues, with high stakes. What is fair and who gets to decide it?” [21] This concern can be resolved by guidelines that require—as I stated before—the developer and user to make choices, enact them, and explicitly communicate them (cf., e.g., [3]). While the developer or user is making choices for the end-user in such cases, it is then clear to the end-user whether they want to use a product or service that is based on such choices. It is not a perfect fix, especially not in a fairly homogenous market place, but it is not as problematic as Munn seems to think. (I will return to this when I discuss Munn’s alternative proposal, since “his alternative”, which has already been used in various guidelines, relates to how guidelines can be improved.)

3.2 Toothless principle and the context of enactment

The next issue is about enforcement and the problem that guidelines are toothless. While I agree that ethical guidelines in isolation are toothless, that is not a charge against the guidelines as such. That just illustrates that enforcement may be needed. The implementation of systems for enforcement can be done in different ways. One example would be hard law that prescribes room for soft law in which ethical guidelines can play a role. It is also clear that the market can manage some forms of self-enforcement. An illustrative comparison is the CE marking. The CE marking is a certification on commercial products, which symbolizes that they abide by important EU standards (e.g., on health, safety and the environment). The certification is made by the company itself [5], which illustrates that an enforcement mechanism that partly depends on good actors in the market can be achieved. Indeed, the principle of self-certification also applies to many ISO-standards.

The idea here is not to enforce a specific set of ethical principles or guidelines, but that we could have a set of legal standards that manufacturers can choose to adhere to and certify that they do. Another possibility is that companies, through binding legal announcements, make clear that they have adhered to some specific ethical guidelines, thus creating a situation of legal liability in case they fail to upload the guidelines. Moreover, as previously mentioned in Sect. 2, guidelines can help with the creation of standards, which can then also underpin regulations, which provides yet another foundation for enforcement.

Of course, one of my proposed solutions depends on a desire to provide self-enforcement, which relates to a worry that Munn has about what he calls “isolated principles”. That is, the enactment of these principles is done in a context (i.e., the tech industry). The problem, Munn, thinks, is that the tech industry has a toxic culture and hence, “Unethical AI is the logical byproduct of an unethical industry.” [21]

This is not necessarily a problem for guidelines, since guidelines can and do address this by requiring ethical standards among developers. For example, stakeholder engagement, requirements on the cooperative culture, and so forth are available in many guidelines (see, e.g., [3, 4, 6, 9, 12, 22, 24, 25]—see also [3] for further examples). Consider, for example:

Requirement 55: Diversity.

In the acquisition and design, deployment and implementation, and monitoring phases, assess and ensure:

- A process to include the participation of different stakeholders in the use and review of the system;

- That efforts are made so that a wide diversity of the public, including different sexes, ages, and ethnicities, are represented;
- If this is applied within your organization, then inform and involve impacted workers and their representatives in advance [3].

The same guideline also requires that “a culture is established and encouraged in which open and structured communication is provided to stakeholders, in line with their requirements”, “a culture of proactive problem management”, “a proactive risk- and self-aware culture”, “a compliance-aware culture”, “a culture that embraces internal audit”, and more [3].

Moreover, the drive for transparency requirements (as previously noted, it is available in 73/84 surveyed guidelines, see, e.g., [1, 3, 4, 12, 14, 22, 23] for specific example), for auditing requirements (see, e.g., [1–4, 9, 12, 14]—see also [7] for a discussion on AI governance and audits), and for accountability or responsibility requirements (see, e.g., [1–4, 9, 12, 14, 22, 23]) serves to respond to the concern of so-called “ethics-washing” or “bluwashing” [8], in which companies misuse ethical guidelines to make it seem as if they are ethical when they are not.

These responses relate directly to Munn’s own proposal, so instead of digging too deeply into the issue, I will turn to Munn’s proposal to argue that he is knocking on open doors.

3.3 Munn’s proposal

Munn suggests that AI ethics should focus more on AI justice. In practice, Munn thinks this means “engaging with groups that bear the brunt of AI impacts but are not typically consulted: children, people of color, LGBT-QIA + communities, migrants, and other groups” [21]. As Munn argues, this implies a need for stakeholder engagement, which, as previously mentioned, is already a central part of many ethical guidelines, some of which specifically mention minorities and underrepresented groups (e.g., [3]). Munn also returns to the issue of bias, which is a standard part of ethical guidelines.

Of course, the worry here is more substantial. That is, if we have an unethical culture, then guidelines will not solve that. But this is exactly why some guidelines go beyond the satisfaction of specific technical goals, and require, for example, that “a common culture is established and encouraged that strongly promotes ethical behavior for all individuals in the enterprise, and establishes a low tolerance threshold for unethical behaviours” [3]. This is not a silver bullet, but these type of suggestions do more to address the issue than Munn’s very limited call for more work on AI justice.

As Munn notes, these problems are arguably related to the issue of power relations. However, it is not clear how

Munn thinks that we should resolve abusive or otherwise problematic power relations. The simple call for avoiding domination is already present in some guidelines. For example, the Montréal Declaration for a Responsible Development of Artificial Intelligence states that “AIS development must help eliminate relationships of domination between groups and people based on differences of power, wealth, or knowledge.” [25]

Munn also mentions transparency, auditing, and accountability, but none of this is new. As previously mentioned, it is already part of many guidelines: for transparency requirements (see, e.g., [1, 3, 4, 12, 14, 22, 23]), auditing requirements (see, e.g., [1–4, 9, 12, 14]), and accountability or responsibility requirements (see, e.g., [1–4, 9, 12, 14, 22, 23]). Indeed, as mentioned in Sect. 2, it is illustrative to consider that transparency (and related concepts) is the most commonly shared principle in ethical guidelines, occurring in 73 out of 84 sampled guidelines [13]. As such, Munn’s suggestion is just knocking on open doors.

Munn also draws attention to unfair biases, which are also already addressed in some ethical guidelines. Consider, for example:

Requirement 45b: Use bias assessment.

In management and deployment and implementation phases, assess and ensure that:

- A strategy or a set of procedures is established to avoid creating or reinforcing unfair bias during the use of the system regarding the use of input data, and that the strategy is based on an assessment of the possible limitations stemming from the composition of the used data sets;
- Use of the system is guided by an awareness of cultural bias to prevent or exacerbate any potential harmful bias [3].

Thus, while I agree with Munn’s call for AI justice, his critique is misguided in the sense that his proposals are already enacted in many guidelines. Moreover, as exemplified, many guidelines contain solutions to these problems that are far more developed than the brief ideas that Munn mentions.³

³ It is perhaps ironic that Munn’s proposals suffer from his own critique, since he also talks of fuzzy high-level ethical values.

4 Final comments

While Munn is correct in saying that there are limitations to what ethical guidelines and principles can do, most of his critique is based on cherry-picking. As I have shown, there are already solutions to the concerns he raises. Moreover, his own proposals are already an important part of many recent guidelines (e.g., his call for transparency completely misses the target, since it is a concept that occurs in almost every guideline). As such his criticism is misguided and his solutions are knocking on open doors. Some might think: what about all of Munn's examples? The fact that some guidelines are bad does not prove that all of them are. More importantly, it does not prove that the challenges cannot be overcome. Indeed, as I have shown in the article, ethical guidelines are already far more progressed than what is reflected in Munn's critique and his own proposals. Such guidelines illustrate both how the critique misses the target and a path forward given the challenges we actually are facing.

Indeed, there are clear challenges for ethical guidelines. The contextual sensitivity of technology raises various challenges (e.g., what is appropriate in one context can be inappropriate in another). As previously mentioned, I have suggested (in [19] that for some applications this implies that we ought to adapt our requirements to the context that requires the highest standard of ethical protections (e.g., of privacy). Thus, although contextual differences can be challenging, there are plausible pathways for ethical guidelines to resolve these issues.

Perhaps two of the most challenging issues are enforcement and operationalization. However, as I have argued, the former is not a problem for the guidelines, but a matter that must be solved by other means (e.g., through regulations or self-enforcement through standards). Munn is correct in noting that operationalization is challenging, but he seems to think that it is practically impossible, which—as I have argued—is rested on a mistake. Operationalization need not be as technical as Munn suggests. Indeed, it arguably should not be, since tying guidelines to particular technical solutions would limit the potential for benefiting from technical progress. Nevertheless, the work on standards, regulations, audits, and so forth shows that we are progressing in the right direction.

Another concern is that we have too many guidelines. However, that just reflects the methodological adolescence of ethical guidelines. Arguably, since we already see progress toward some consensus building [3, 13], as time progresses we will hopefully have fewer guidelines. However, it is also important to mention that there is, and should be, a space for a variety of guidelines, since we can have guidelines for different purposes (design or use) and

for different parts of an organization (developing team, marketing team, executives, etc.).

Lastly, although Munn points to some salient problems, such as inherently unethical corporate cultures, his proposals for how to resolve these pressing matters are—as I exemplified in Sect. 3.3—less developed than what is already present in some ethical guidelines.

Acknowledgements The author wants to thank Birna van Riemsdijk for helpful comments prior to submission. The author also wants to thank two reviewers for AI & Ethics for helpful comments on my submitted version.

Funding This work is part of the research programme Ethics of Socially Disruptive Technologies, which is funded through the Gravitation programme of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.004.031).

Declarations

Conflict of interest The author has no competing interests to report.

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

References

1. Amnesty International and Access Now.: The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems. <https://www.torontodeclaration.org/declaration-text/english/>. Accessed 24 Nov 2022 (2018)
2. The Bertelsmann Stiftung and iRights.Lab.: Rules for the design of algorithmic systems. <https://algorules.org/en/home>. Accessed 24 Nov 2022 (2020)
3. Brey, P., Lundgren, B., Macnish, K., and Ryan, M.: Guidelines for the development and use of SIS. Deliverable D3.2 of the SHERPA project. (2019). <https://doi.org/10.21253/DMU.11316833.v3>.
4. Brey, P., Jansen, P., Maas, J., Lundgren, B., Resseguier, A.: D4.7: An Ethical framework for the development and use of AI and robotics technologies. https://sienna-project.eu/digitalAssets/801/c_801912-1_1-k_d4.7_ethical-framework-for-ai--robotics-1-.pdf. Accessed 24 Nov 2022 (2020)
5. Commission Notice—The 'Blue Guide' on the implementation of EU products rules 2016 (Text with EEA relevance). Official Journal, C 272, 1–149. CELEX: [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52016XC0726\(02\)](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52016XC0726(02)). Accessed 24 Nov 2022 (2016)
6. European Commission: High-level expert group on artificial intelligence (AI HLEG). Ethics guidelines for trustworthy AI. <https://>

- ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trust-worthy-ai. Accessed 24 Nov 2022 (2019)
7. Falco, G., Shneiderman, B., Badger, J., et al.: Governing AI safety through independent audits. *Nat. Mach. Intell.* **3**, 566–571 (2021). <https://doi.org/10.1038/s42256-021-00370-7>
 8. Floridi, L.: Translating principles into practices of digital ethics: five risks of being unethical. *Phil. Technol.* **32**, 185–193 (2019). <https://doi.org/10.1007/s13347-019-00354-x>
 9. Floridi, L., Cowls, J., Beltrametti, M., et al.: AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Mind. Mach.* **28**, 689–707 (2018). <https://doi.org/10.1007/s11023-018-9482-5>
 10. Herlitz, A.: Predictive Fairness. In: Lundgren, B., Nuñez Hernández, N.A. (eds.) *Philosophy of computing*. Philosophical studies series, vol. 143, pp. 141–161. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-75267-5_5
 11. Hopster, J.: What are socially disruptive technologies? *Technol. Soc.* **67**, 101750 (2021). <https://doi.org/10.1016/j.techsoc.2021.101750>
 12. The IEEE global initiative on ethics of autonomous and intelligent systems. Ethically aligned design: a vision for prioritizing human well-being with autonomous and intelligent systems, Version 2. IEEE. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf. Accessed 24 Nov 2022 (2018).
 13. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* **1**, 389–399 (2019). <https://doi.org/10.1038/s42256-019-0088-2>
 14. Information Commissioner's Office (ICO): Big data, artificial intelligence, machine learning and data protection. <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>. Accessed 24 Nov 2022 (2017).
 15. Kerwer, D.: Rules that many use: standards and global regulation. *Governance* **18**, 611–632 (2005). <https://doi.org/10.1111/j.1468-0491.2005.00294.x>
 16. Kleinberg, J., Mullainathan, S., Raghavan, M.: Inherent trade-offs in the fair determination of risk scores. *arXiv:1609.05807v2*. (2017).
 17. Lundgren, B.: Ethical machine decisions and the input-selection problem. *Synthese* **199**, 11423–11443 (2021). <https://doi.org/10.1007/s11229-021-03296-0>
 18. Lundgren, B.: Improving on and assessing ethical guidelines for digital tracking and tracing systems for pandemics. *Ethics Inf. Technol.* **23**(Suppl 1), 139–142 (2021). <https://doi.org/10.1007/s10676-020-09561-z>
 19. Lundgren, B.: Ethical requirements for digital systems for contact tracing in pandemics: a solution to the contextual limits of ethical guidelines. In: Macnish, K., Henschke, A. (eds.) *The Ethics of Surveillance in Times of Emergency*. Oxford University Press, Oxford (2023)
 20. Morley, J., Cowls, J., Taddeo, M., Floridi, L.: Ethical guidelines for COVID-19 tracing apps. *Nature* **582**, 29–31 (2020). <https://doi.org/10.1038/d41586-020-01578-0>
 21. Munn, L.: The uselessness of AI ethics. *AI Ethics* (2022). <https://doi.org/10.1007/s43681-022-00209-w>
 22. OECD: Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>. Accessed 24 Nov 2022
 23. The Public Voice.: Universal Guidelines for artificial intelligence. <https://thepublicvoice.org/ai-universal-guidelines/>. Accessed 24 Nov 2022 (2018).
 24. UNI Global Union. Top 10 principles for ethical artificial intelligence. http://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf. Accessed 24 Nov 2022
 25. University of Montreal. (2018). Montréal declaration for a responsible development of artificial intelligence. https://www.montr.ealdeclaration-responsibleai/_files/ugd/ebc3a3_c5c1c196fc164756afb92466c081d7ae.pdf. Accessed 24 Nov 2022
 26. Winfield, A.F.T., Jirotko, M.: Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Phil. Trans. Royal Soc. A* **376**, 20180085 (2018). <https://doi.org/10.1098/rsta.2018.0085>

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