Against Ethical AI

Guidelines and Self Interest

Donald McMillan
donald.mcmillan@dsv.su.se
Department of Computer and System Sciences
Stockholm University
Stockholm, Sweden

Barry Brown
barry@dsv.su.se
Department of Computer and System Sciences
Stockholm University
Stockholm, Sweden

ABSTRACT

In this paper we use the EU guidelines on ethical AI, and the responses to it, as a starting point to discuss the problems with our community's focus on such manifestos, principles, and sets of guidelines. We cover how industry and academia are at times complicit in 'Ethics Washing', how developing guidelines carries the risk of diluting our rights in practice, and downplaying the role of our own self interest. We conclude by discussing briefly the role of technical practice in ethics.

CCS CONCEPTS

• **Social and professional topics → Codes of ethics**; *Computing / technology policy*; *Governmental regulations*.

KEYWORDS

Ethics, Algorithms, Artificial Intelligence, Human Rights, Policy

ACM Reference Format:
Donald McMillan and Barry Brown. 2019. Against Ethical AI: Guidelines and Self Interest. In *Proceedings of the Halfway to the Future Symposium 2019 (HTTF 2019), November 19–20, 2019, Nottingham, United Kingdom.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3363384.3363393

1 INTRODUCTION

As interest in the applications of artificial intelligence has grown, so have ethical guidelines for AI, and research around the ethics of computing systems. The list of guidelines curated by AlgorithmWatch is currently at 83 and growing[2]. These efforts are addressing a real problem: clearly, machine learning systems are being deployed in a host of settings where there is concern for the potential harm that such systems can cause. Moreover, the potential of such systems for even greater harm is prescient. Basic technologies such as facial recognition have such a broad range of applications, many of which are close to being inherently harmful — such as differentiating membership of an ethnic group [12].

Yet, as we argue in this short paper, there are aspects of ethical guidelines that are not about benefiting any actual user (or subject) of AI systems. Ethical guidelines can work as a mechanism to minimise institutional blame. As with research ethics, it is not clear that increased guidelines result in more ethical researcher behaviour. Indeed, McNamara et al. [10] showed that the ACM's ethical guidelines had little to no effect on the choices made by developers. In nearly all cases, guidelines stay at the level of pronouncements, statements of value that would take large (and undefined) effort to interpret in any actual situation. Moreover, these ethical guidelines often have contradictions or could have damaging unintended consequences.

Part of the problem that we identify here is that there is little consideration of *self-interest* in the discussions and descriptions of ethics as proposed. Self-interest encompasses motivations, which cause direct benefits to oneself or one's immediate kin, family or companions. Organisations, companies and groups can act self-interested, and while we would not claim that this is the dominant motivation behind human action and activity, clearly it should not be ignored. We discuss here how self-interest manifests in the generation of 'ethics work' as a form of protectionism, how these resulting manifestos and guidelines by there unenforceable nature can foreshadow the dilution of human rights overall, and encourage the move beyond value statements to influencing and enforcing policy and law.

2 ETHICS WASHING

The proposed Ethics guidelines for trustworthy AI from the EU was drawn up by a panel of 52 experts drawn from industry and academia, with a public consultation period which garnered input from over 500 other interested parties from the EU and beyond. The makeup of this panel of experts has been one initial source of criticism. Thomas Metzinger [11], an academic member of the panel himself, questioned the overwhelming number of industry-based or industry-funded members included. Metzinger reports how the pressure from industrial members removed text he championed, which included 'red lines' and 'non-negotiable' limits to AI and its impacts on subjects. This was done supposedly in favour of a 'positive vision' yet Metzinger himself points to this as one reason he now sees the whole exercise as one of 'ethics washing'. As described by Wagner [16], ethics washing is the use of working groups, guidelines, and manifestos as a counterbalance to calls for legal and regulatory frameworks which would ensure the safety of the public.

In HCI some have suggested that there be some sort of removal of accreditation for those who break such guiding principles [3, 9], others have even called for AI practitioners taking something akin to the Hippocratic oath medical professionals are expected to swear

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
HTTF 2019, November 19–20, 2019, Nottingham, United Kingdom
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-7203-9/19/11…$15.00
https://doi.org/10.1145/3363384.3363393

[5, 15]. Yet swearing an oath as a programmer, data scientist, or AI operator does not bring the level of protection of those at the receiving end of the 'digital treatment'. Indeed, the oath itself has been shown to be subservient, while complimentary, to the laws of the state [8] in ensuring the protection of rights. As pointed out by Sloan [13], discussions of ethics and technology have a tendency to position the harms as *social* while the solutions are *technological*. "That the social problem is deeply entangled with the existing fault lines of social stratification falls somewhat outsize the ontology of 'ethical algorithms"'[13]. This narrowing of scope of ethical conversation can lead to something that has the potential to be a 'dilution of rights'. The danger that in overlaying such unenforceable 'digital' rights on basic 'human' rights, the corresponding human rights will also become devalued or unenforced. Returning to the EU guidelines, it is notable that while they state that they are founded upon the EU declaration of Fundamental Human Rights, they only address a subset of those rights in the context of algorithmic harm. Wagner [16] notes that these regulations cover the rights to human dignity while ignoring rights such as the freedom of assembly or cultural rights.

How to manifest ethical guidelines in the work of actual design is usually left unspecified. Design requires trade offs and, in practice, it can be difficult to see how to balance conflicting ethical principles, or even how to respect these principles without producing designs that are poorer in some critical aspects. It is not clear that mandating websites to ask before setting a tracking cookie on a users machine has resulted in better outcomes for users, even if it fits better with the ethical principle of consent.

3 SELF-INTEREST

This brings us to the concept of motivation and self-interest in the generation and application of ethical guidelines, manifestos, principles, and oaths. It is easy to be pessimistic and see all such action in the light of 'ethics washing', avoidance of regulation, and manipulation of public opinion. It can also be seen, on the other hand, as evidence that a large number of people in and around technology are interested, motivated, and hopeful that the future technology we build and deploy can be shaped to embody the best of our goals and principles.

Starting on the side of pessimism, it is important to acknowledge and work with the self-interest that drives so much of human action. The corporations bankrolling the countless hours spent on generating and discussing these ethical guidelines are themselves not driven only by concerns of what is the most ethical action to take. There are any number of theories and models that point to understandings of how such corporate actors, and the individuals of which they comprise, align social good and business success. Freeman et al. 's Stakeholder model [6] talks of decisions being made by weighing the good of those within and without the business, which is often criticised [4] for naive optimism on leaving the weighing and inclusion of such stakeholders to the individual. Friedman, on the other hand, puts decision making firmly for the benefit of shareholders [7] bound only by the necessity to stay within the law (although he later added that "basic rules of society...and ethical custom" should also be respected). For Friedman, a manager deciding to spend company resources on developing a set of ethical guidelines should see this as a direct benefit to the profits and, therefore, shareholders of the company. In the more optimistic Stakeholder model, the decision to spend employee time on creating ethical guidelines should balance the benefit of society with the benefit of the employees, shareholders, and customers. In either case, the self-interest of the companies involved is central. This does not have to manifest only in the avoidance of regulations [16], but could be seen as a form of lobbying to ensure even incrementally more favourable regulations towards their product, service, or sector at the expense of competitors [7]. Yet, as seen in the makeup of the EU's panel for ethical AI, even in a governmental setting, such confounding motivations are largely ignored.

On the positive side, this corporate self-interest provides a large number of highly skilled and experienced practitioners time and energy to devote to the discussion of these issues. Going further, it can be expected that for those who end up being involved at least one of the many motivations (from prestige, to corporate visibility, to ensuring the continued good fortunes of their employer) would be that they *want*, and in fact in many cases have pushed through internal structures and strictures, to be involved. Currently, the balance of these motivations has been left mostly to the moral and ethical fibre of the individual (harking back to the naivety of expectation in the Stakeholder model of governance), yet by through careful frameworks and regulation such expertise can be brought to bear. This is by no means an easy task, one dimension of political lobbying as the practice of providing 'friendly' experts to inform on policy is fraught with problems [14] — yet these are acknowledged and those accepting the opinions do so with some understanding of the competing motivations of those they are talking to. In future ethical consultations involving industry such actors should be seen in the same light, embracing that their self-interest may not align with the goals of society as a whole yet involving them (as members of that society) in the process.

4 CONCLUSION

Moving forward, we plan to explore in more depth the multiple relationships between guidelines and actual cases of technical practice. As Phil Agre wrote [1] – the relationship between technical practice, and attempts to guide that technical practice are complex. Agre wrote of the need for a "a split identity – one foot planted in the craftwork of design and the other foot planted in the reflexive work of critique". We take this to mean that assessing ethical guidelines will require both close attention to practice and critique, at the intersection of academia, industry, and governance. In this short paper we have started to critically discuss and assess the role of ethical guidelines in AI. We point to the role of self interest in generating these discussions as a way of questioning why particular entities might produce guidelines, and to what likely use they might or might not be put.

REFERENCES

[1] Philip E. Agre. 1997. Lessons Learned in Trying to Reform AI. In *Social Science, Technical Systems, and Cooperative Work*, Geoffrey Bowker, Susan Leigh Star, and Les Gasser (Eds.). Routledge, Mahwah, N.J.

[2] AlgorithmWatch. 2019. AI Ethics Guidelines Global Inventory.

[3] Barry Brown, Alexandra Weilenmann, Donald McMillan, and Airi Lampinen. 2016. Five Provocations for Ethical HCI Research. In *Proceedings of the 2016 CHI*

Conference on Human Factors in Computing Systems (CHI '16). ACM, New York, NY, USA, 852–863. https://doi.org/10.1145/2858036.2858313

[4] Thomas L. Carson. 2003. Self–Interest and Business Ethics: Some Lessons of the Recent Corporate Scandals. *Journal of Business Ethics* 43, 4 (April 2003), 389–394. https://doi.org/10.1023/A:1023013128621

[5] Oren Etzioni. 2018. A Hippocratic Oath for Artificial Intelligence Practitioners.

[6] William M. Evan and R. Edward Freeman. 1988. *A Stakeholder Theory of the Modern Corporation: Kantian Capitalism.*

[7] Milton Friedman. 2009. *Capitalism and Freedom.* University of Chicago press.

[8] Lisa Hasday. 2013. The Hippocratic Oath as Literary Text: A Dialogue Between Law and Medicine. *Yale Journal of Health Policy, Law, and Ethics* 2, 2 (Feb. 2013).

[9] Donald McMillan, Alistair Morrison, and Matthew Chalmers. 2013. Categorised Ethical Guidelines for Large Scale Mobile HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1853–1862. https://doi.org/10.1145/2470654.2466245

[10] Andrew McNamara, Justin Smith, and Emerson Murphy-Hill. 2018. Does ACM's Code of Ethics Change Ethical Decision Making in Software Development?. In *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2018)*. ACM, New York, NY, USA, 729–733. https://doi.org/10.1145/3236024.3264833

[11] Thomas Metzinger. 2019. Ethics washing made in Europe. https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html.

[12] S. Md. Mansoor Roomi, S. L. Virasundarii, S. Selvamegala, S. Jeevanandham, and D. Hariharasudhan. 2011. Race Classification Based on Facial Features. In *2011 Third National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics*. 54–57. https://doi.org/10.1109/NCVPRIPG.2011.19

[13] Mona Sloane. 2019. Inequality Is the Name of the Game: Thoughts on the Emerging Field of Technology, Ethics and Social Justice. In *Proceedings of the Weizenbaum Conference 2019 "Challenges of Digital Inequality - Digital Education, Digital Work, Digital Life"*. 9. https://doi.org/10.34669/wi.cp/2.9

[14] Bart Slob and Francis Weyzig. 2010. Corporate Lobbying and Corporate Social Responsibility: Aligning Contradictory Agendas. In *Business, Politics and Public Policy: Implications for Inclusive Development*, José Carlos Marques and Peter Utting (Eds.). Palgrave Macmillan UK, London, 160–183. https://doi.org/10.1057/9780230277243_7

[15] Brad Smith and Harry Shum. 2018. Artificial Intelligence and Its Role in Society. In *The Future Computed*. Microsoft Corporation.

[16] Ben Wagner. 2018. Ethics as an Escape from Regulation: From Ethics-Washing to Ethics-Shopping? In *Being Profiled*, Emre Bayamlioğlu, Irina Baraliuc, Liisa Janssens, and Mireille Hildebrandt (Eds.). Amsterdam University Press, 84–89. https://doi.org/10.2307/j.ctvhrd092.18