

AI's ethics problem: Abstractions everywhere but where are the rules?

18 November 2019, by Nancy Cohen



Credit: CC0 Public Domain

Machines that make decisions about us: what could possibly go wrong? Essays, speeches, seminars pose that question year after year as artificial intelligence research makes stunning advances. Baked-in biases in algorithms are only one of many issues as a result.

Jonathan Shaw, managing editor, *Harvard Magazine*, [wrote](#) earlier this year: "Artificial intelligence can aggregate and assess vast quantities of data that are sometimes beyond human capacity to analyze unaided, thereby enabling AI to make hiring recommendations, determine in seconds the creditworthiness of loan applicants, and predict the chances that criminals will re-offend."

Again, what could possibly go wrong?

"AI development needs strong moral guidance if we are to avoid some of the more catastrophic scenarios envisaged by AI critics," said Stephen Fleischresser, lecturer, University of Melbourne's Trinity College, in [Cosmos](#).

Edd Gent in [SingularityHub](#): "Concerns around

privacy, transparency and the ability of algorithms to warp social and political discourse in unexpected ways have resulted in a flurry of pronouncements from companies, governments, and even supranational organizations on how to conduct ethical AI development."

So, the question becomes, who is minding the AI store? If we already imagine what could go wrong with the AI picture, who, which individual, which groups, if any, are trying to channel all the talk and take leadership for a working set of rules?

A good-and-timely [read](#) about this has emerged in *Nature Machine Intelligence*. The paper does not give you the single solution but it does shake up the conversation. As its title reads, "Principles alone cannot guarantee ethical AI."

The author, Brent Mittelstadt, is of the Oxford Internet Institute, University of Oxford, Oxford, UK and The Alan Turing Institute, British Library, London, UK.

Fleischresser described Mittelstadt as "an ethicist whose research concerns primarily digital ethics in relation to algorithms, machine learning, artificial intelligence, predictive analytics, Big Data and medical expert systems."

The problem has not been the lack of talk; "Artificial intelligence (AI) ethics is now a global topic of discussion in academic and policy circles," Mittelstadt stated. "At least 84 public-private initiatives have produced statements describing high-level principles, values and other tenets to guide the ethical development, deployment and governance of AI."

The problem is that with all the talk, well-intentioned stakeholders have not gone far beyond the starting gate. The supportive layer of lofty statements and boldfaced principles may be in place, but where is the top layer? "The truly difficult part of

ethics—actually translating normative theories, concepts and values into good practices AI practitioners can adopt—is kicked down the road like the proverbial can."

While high principles are laudable, "AI ethics initiatives have thus far largely produced vague, high-level principles and value statements that promise to be action-guiding, but in practice provide few specific recommendations...Declarations by AI companies and developers committing themselves to high-level ethical principles and self-regulatory codes nonetheless provide policymakers with a reason not to pursue new regulations."

Mittelstadt said "a principled approach may have limited impact on governance. What is missing: A unified regulatory framework that sets up "clear fiduciary duties towards data subjects and users." The absence of a fiduciary relationship, he said, means users "cannot trust that developers will act in their best interests when implementing ethical principles in practice."

The paper spelled out what the author identified as areas of concern for the future of AI ethics.

Legal and professional accountability was one of the key areas under discussion—that is, "the relative lack of legal and professional accountability mechanisms." Why this is a problem: "Serious, long-term commitment to self-regulatory frameworks cannot be taken for granted."

The author posed the fundamental question: Given these weaknesses in existing legal and professional accountability mechanisms for AI, one must ask: "is it enough to define good intentions and hope for the best? Without complementary punitive mechanisms and governance bodies to step in when self-governance fails, a principled approach runs the risk of merely providing false assurances of ethical or trustworthy AI."

Where do we go from here in AI ethics?

The author said this "principled" approach needs cooperative oversight. Think "Binding and highly visible accountability structures" along with "clear implementation and review processes" at a sectoral

and organizational level. "Professional and institutional norms can be established by defining clear requirements for inclusive design, transparent ethical review, documentation of models and datasets, and independent ethical auditing."

More information: Brent Mittelstadt. Principles alone cannot guarantee ethical AI, *Nature Machine Intelligence* (2019). DOI: 10.1038/s42256-019-0114-4

© 2019 Science X Network

APA citation: AI's ethics problem: Abstractions everywhere but where are the rules? (2019, November 18) retrieved 26 November 2020 from <https://techxplore.com/news/2019-11-ai-ethics-problem-abstractions.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.