Santa Clara High Technology Law Journal

Volume 36 | Issue 4 Article 1

5-18-2020

IS “ETHICAL AI” A RED HERRING?

Walker, Joshua

Recommended Citation
Walker, Joshua, *IS “ETHICAL AI” A RED HERRING?*, 36 SANTA CLARA HIGH TECH. L.J. 445 ().
Available at: https://digitalcommons.law.scu.edu/chtlj/vol36/iss4/1

IS “ETHICAL AI” A RED HERRING?

By Joshua Walker

INTRODUCTION

Proposed: *“Ethical AI” is a red herring. The term attenuates the very effect it is nominally designed to induce.*

There are very real dangers, and very real operational biases, in advanced software algorithms—indeed, in almost any software. This is true both because: (i) humans are inherently biased, and humans still generally write software, directly or indirectly; and (ii) data used to develop most classifiers and other algorithmic processes are also derived, directly or indirectly, from said biased humans. Both the creative matrix and the authors—the data and the humans—necessarily and naturally introduce bias into AI.

These concrete dangers require immediate operational controls. And many advocates of “ethical AI”, a powerful incipient movement, rightfully point this out with the best of intentions. The debate is needful.

But our problem is textual.

I. THE TEXTUAL PROBLEM

As any 1L has learned, the word “ethical,” like the word “moral,” is highly subjective. Thus, the ethical AI movement is saying:

> Let’s eliminate subjectivity in certain advanced software tools by imposing inherently subjective standards on such things.

The term defeats itself, and there is another lexical problem: “Ethical AI” is something of an oxymoron—or at least a misnomer. According to many leading technical votaries in machine learning, natural language processing, and neural network methodologies: There is a vast gulf between effective automation and self-awareness.

- Obviously, ethical (though better yet legal) standards can, in theory, be effectively *applied* to artificial intelligence modalities, thus making such AI, together with such exogenous controls, “ethical.”
- However, the term “ethical AI” is ambiguous (in how the adjective applies to the noun). It may imply, particularly to a layperson, that an

autonomous piece of software/robot is, itself, intrinsically ethical—consciously applying and developing a moral sense like a human (hopefully) does.

- Implying consciousness and self-awareness by statistical/logical/etc. based software tools is (i) bad marketing (over-stating and disingenuous), and (ii) likely to lead to even further misunderstandings about what AI currently is.
- Most definitions of AI centrally include the term algorithm—which itself is generally a [human defined] step-by-step, mechanistic process which does *not* require thought. In other words, "artificial intelligence" is neither. So neither intrinsically moral nor immoral either.
- No piece of software, given present technological means, is going to automatically define and refine its own moral "North Star" as we understand it.

II. THE RED HERRING PROBLEM

But the term itself is not its biggest problem; it is the operational consequences that flow therefrom.

What do we do to embody societal views of ethics in practice? We pass and promulgate laws. Unlike the vast majority of ethics codes and standards, laws have consequences. Laws are specific. Ethics are, generally, neither. One argues that while "ethical AI" may be a sop to popular fear and anger, it does not generally require, evoke, or encourage the specific types of action that society may direct.

A process—technological or legal—is designed to achieve a certain result. What is the objective of an AI ethics movement with no objective object? And ask an engineer constructing a process whether they think it is a good idea to measure the performance such process by subjective terms ("good", "bad", "green"; an "ethical airplane") and she will look at you funny. So why here?

Are some proponents of "ethical AI" trying to deflate concern about artificial intelligence by promulgating a nominal remedy that may have few concrete consequences for violation or aberration?

Is this movement designed to deflate popular fear and anger in a way that does not interfere with advanced software experimentation, or rapid AI commercialization efforts? Is "ethical AI" a red herring?

Technophiles and researchers naturally fear undue meddling in a highly competitive, dynamic area. As a prominent jurist and academic once said: "To a tech company, lawyers are like a bunch of zombies, walking around, killing everything they touch." This is a real fear. Lawyers are not always viewed as agents for advancing or catalyzing innovation—for advancing speed or enhancing the agility required for commercial and scientific success in this space. Technologists, and their commercializing cousins, naturally want to avoid fines if they cross a trip wire, or cause a harm of one kind or another. This is particularly true for certain AI methods, like neural network approaches, where the engineers and data scientists may themselves have limited understanding of *why* the AI produced a given result or classification.

"Ethical AI" gives every company a way to address public and governmental concern—while dislocating attorneys from the center of the conversation. *Zombie be gone!* More importantly, the relatively consequence-free autonomy that independent, subjectively evaluated "ethics bodies" feature is likely seen as a boon to pure technological innovation and commercial exploitation.

The problem is that this (perhaps speculative) fear of lawyers must be counterbalanced against the very real public fears regarding artificial intelligence. Far more importantly, the interests of victims of discrimination must be front and center—whether the modality of discrimination is a punch in the face or a slap on the internet. The slap is real, so should be the consequence for the slapper. We already have laws about bias—constitutional and statutory. The way they should be *applied* to new technologies always has to be addressed, in concrete cases, but that doesn't mean AI should be in a "law free" zone any more than the internet should be. (For more on that, see the last decade or two.)

But aren't there problems with scale and speed of traditional legal code and concomitant analysis as applied to rapidly evolving, sometimes opaque, and even partially autonomous systems?

In a word, "yes". The zombie meme is popular with engineers for a reason. But the answer isn't for law to be sublimated into gas, or gassy, altruistic sounding phraseology. It is the reverse. **It is legal engineering.** One proposes the creation of legal AI and legal engineering subsystems for commercial AI—legal-centric software that scales with the operations it constrains. But this, alas, requires more paper and/or time.