

Olivia P. Walker

Revised 6 November 2017

Examining the Risks of Artificial Intelligence

Technological advances have changed the way people work, consume information, and live. Innovations in technology gave impetus to the development of Artificial Intelligence (AI), which can be defined as computational systems and programming models that enable machines to acquire and apply knowledge in a manner previously exclusive to humans. In recent years, AI has generated controversy. Proponents cite AI's track record of improving operational efficiency and productivity through automation and big data analytics. Conversely, AI opponents blame automation for unemployment and rising inequality; some have even suggested AI is a threat to humanity. AI is not a threat to humanity; the algorithms and programmers' methods for developing AI technology are the real threat to humanity. This is the point I argue in this paper.

Employers, courts, banks, insurance companies, and educational institutions use AI technology to determine parole eligibility, credit worthiness, insurance premiums, teacher quality, and to make hiring decisions, among other things. The problem is:

AI researchers sometimes describe their approach as "heuristic programming." An early CMU Ph.D. thesis defined a heuristic program as one that "does not always get the right answer." Heuristic programs are based on "rules of thumb," that is, rules based on experience but not supported by theory. (Parnas 28)

This means algorithms that inform AI – and determine credit and employment worthiness – are based on correlative assumptions rather than evidence-based research. The threat that this poses to society is well documented.

Research detailing instances of algorithmic bias in AI is widely published. Researchers from Microsoft and Google have identified bias embedded in algorithms to include “flawed and misrepresentative systems used to rank teachers, and gender-biased models for natural language processing” (Knight). The potential detriment of such flaws to society is profound. Notably, according to Parnas, utilizing “rules of thumb” as a methodological approach to AI development exposes society to detrimental, yet avoidable outcomes (27).

It is widely understood that AI technology is as “intelligent” as the person(s) developing it. Meaning, the extent to which AI poses a threat to humanity is dependent upon human activity during AI development. While strategies to mitigate algorithmic bias have been identified, they are rarely utilized. Researchers have found “that crucial stakeholders, including the companies that develop and apply machine learning systems and government regulators, show little interest in monitoring and limiting algorithmic bias” (Knight).

Artificial intelligence is not a threat to humanity; the algorithms and programmers’ methods for developing AI technology are threats to humanity. Algorithms used in AI are a primary basis upon which institutions make crucial decisions about people, such as credit worthiness and fitness for employment. Even still, AI developers continue to employ heuristic programming, which is an approach based on faulty assumptions, and not grounded in theory. Recent studies have identified instances of algorithmic bias, disproving the reliability and validity of AI systems. Notwithstanding, government regulators and industry stakeholders have largely failed to prioritize the development of: AI bias mitigation strategies, mechanisms to assure integrity in AI development processes, and AI program monitoring and evaluation techniques.

Works Cited

- Knight, Will. "Biased Algorithms are Everywhere, and No One Seems to Care." *MIT Technology Review*, 12 July 2017, www.technologyreview.com/s/608248/biased-algorithms-are-everywhere-and-no-one-seems-to-care. Accessed 4 Nov 2017.
- Parnas, David L. "The Real Risks of Artificial Intelligence." *Communications of the ACM*, vol. 60, no. 10, Oct. 2017, pp. 27-31. *DeepDyve*, doi: 10.1145/3132724.