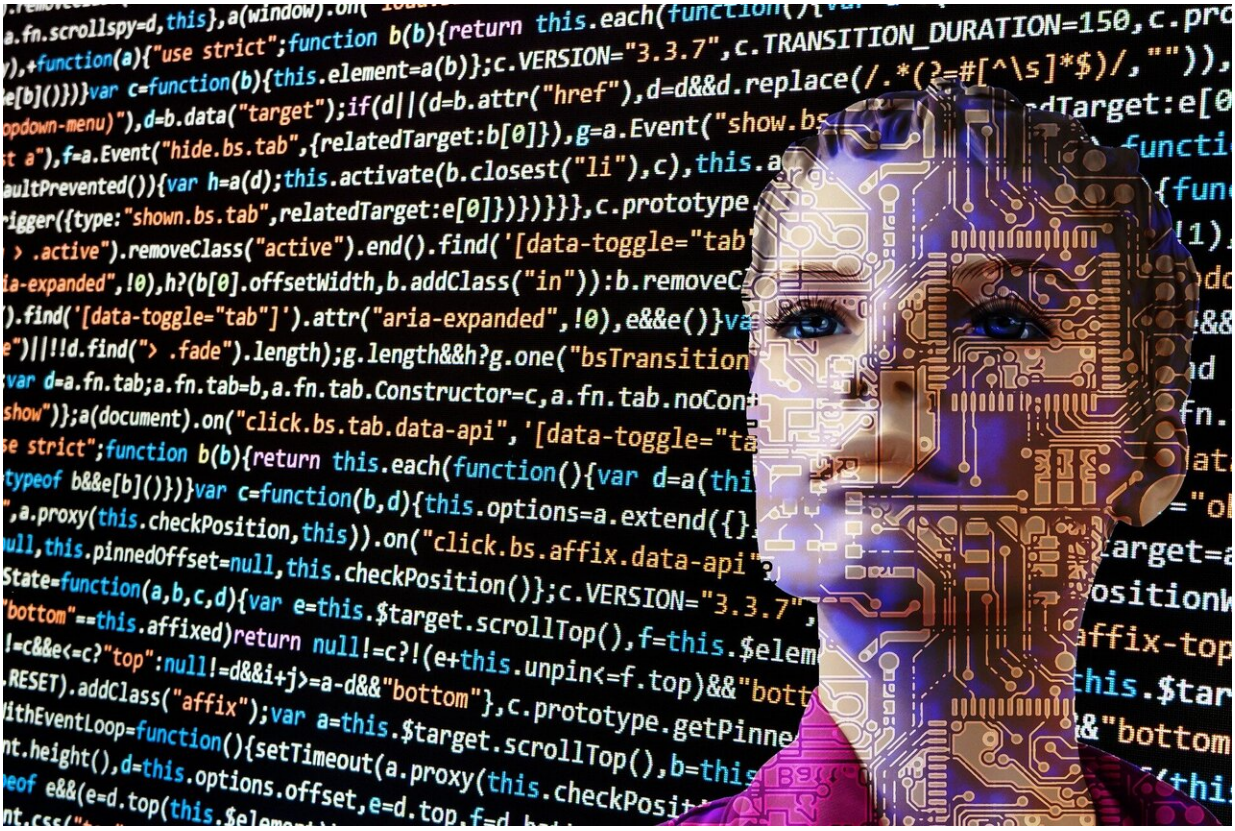
What is the difference between AI ethics, responsible AI, and trustworthy AI?

November 14 2023, by Tyler Wells Lynch



Credit: CC0 Public Domain

AI is everywhere—driving cars, diagnosing illnesses, making credit decisions, ranking job candidates, identifying faces, assessing parolees. These headlines alone should be enough to convince you that AI is far

from ethical. Nonetheless, terms like "ethical AI" prevail alongside equally problematic terms like "trustworthy AI."

Why are these phrases so thorny? After all, they're just words—how dangerous can they be? Well, words matter, and if we're ever to achieve a future where AI is worthy of our trust, then we at least need to agree on a common vocabulary.

Co-chairs of the AI Ethics Advisory Board at the Institute for Experiential AI (EAI): Cansu Canca and Ricardo Baeza-Yates explain the differences between these terms and why they matter.

The problem with 'trustworthy AI'

For Ricardo Baeza-Yates, who is also the director of research at EAI, it all comes down to a fundamental distinction between human and computational abilities. Artificial intelligence is not human, so we should avoid terms like "trustworthy AI" that not only humanize AI but also imply a level of dependability that simply does not exist.

"We know that AI does not work all the time, so asking users to trust it is misleading," Baeza-Yates explains. "If 100 years ago someone wanted to sell me an airplane ticket, calling it 'trustworthy aviation,' I would have been worried, because if something works, why do we need to add 'trustworthy' to it? That is the difference between engineering and alchemy."

Cansu Canca, ethics lead at EAI, adds that "trustworthy AI" seems to direct the attention to the end goal of creating trust in the user. By doing so it circumvents the hard work of integrating ethics into the development and deployment of AI systems, placing the burden on the user.

"Trust is really the outcome of what we want to do," she says. "Our focus should be on the system itself, and not on the feeling it eventually—hopefully—evokes."

The problem with 'ethical AI'

Ethical AI faces a similar problem in that it implies a degree of moral agency. Humans intend certain ethical outcomes. They can make value judgments and reorient their behavior to account for goals that do not translate to the world of algorithms.

"AI can have an ethical outcome or an unethical outcome," Cansu says. "It can incorporate value judgments, but it's not an ethical being with intent. It's not a moral agent."

Ethics, in that sense, is strictly the domain of human beings. Challenges emerge when people start to design systems with autonomous decision-making capabilities, because those systems are only as ethical as the intent of the people who create them.

'Responsible AI'

Ricardo and Cansu both prefer the term "responsible AI" while acknowledging that it, too, is imperfect. "Responsibility is also a human trait, but law has extended the concept of responsibility to institutions, so we use it in that sense," says Ricardo.

"In a way, 'responsible AI' is a shorthand for responsible development and use of AI, or responsible AI innovation," Cansu adds. "The phrase is still open to the interpretation that AI itself will have some responsibility, which is certainly not what we mean. We are trying to emphasize that responsible AI is about creating structures and roles for

developing AI responsibly, and that responsibility will always lie in these structures and the people who design the systems."

Cansu and Ricardo both see AI ethics as a component of responsible AI. Within that subdomain we find the perennial ethical question, "What is the right thing to do?" And in the larger domain around it we find room for innovation—an exploratory, interdisciplinary space for designers, developers, investors, and stakeholders that ultimately (hopefully) points towards an ethical core.

"We philosophers collaborate with developers and designers to find the ethical risks and mitigate them as they develop AI systems and design AI products," Canca says.

Provided by Northeastern University

Citation: What is the difference between AI ethics, responsible AI, and trustworthy AI? (2023, November 14) retrieved 3 May 2024 from https://techxplore.com/news/2023-11-difference-ai-ethics-responsible-trustworthy.html