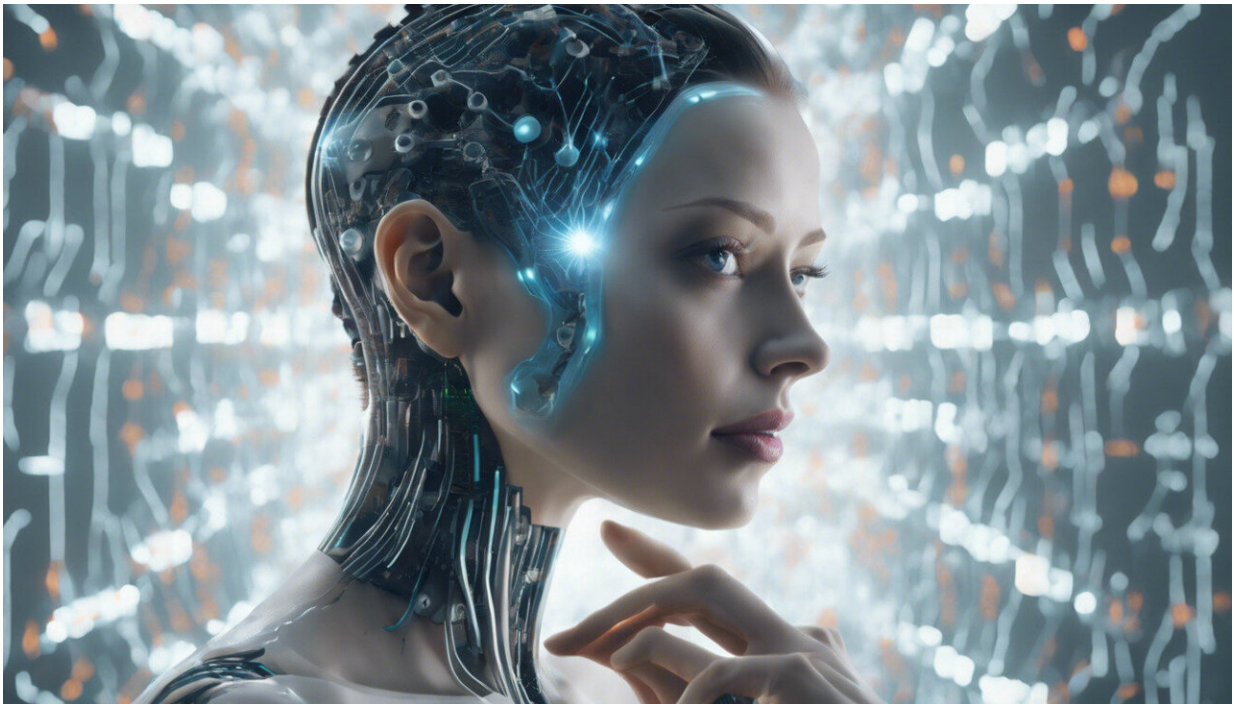


Researchers suggest historical precedent for ethical AI research

February 15 2024, by Chad Boutin



Credit: AI-generated image

If we train artificial intelligence (AI) systems on biased data, they can, in turn, make biased judgments that affect hiring decisions, loan applications, and welfare benefits—to name just a few real-world implications. With this fast-developing technology potentially causing life-changing consequences, how can we make sure that humans train AI

systems on data that reflects sound ethical principles?

A multidisciplinary team of researchers at the National Institute of Standards and Technology (NIST) is suggesting that we already have a workable answer to this question: We should apply the same basic principles that scientists have used for decades to safeguard [human subjects](#) research.

These three principles—summarized as "respect for persons, beneficence and justice"—are the core ideas of 1979's watershed [Belmont Report](#), a document that has influenced U.S. government policy on conducting research on human subjects.

The team has [published](#) its work in the February issue of the journal *Computer*. While the paper is the authors' own work and is not official NIST guidance, it dovetails with NIST's larger effort to support the development of trustworthy and responsible AI.

"We looked at existing principles of human subjects research and explored how they could apply to AI," said Kristen Greene, a NIST social scientist and one of the paper's authors. "There's no need to reinvent the wheel. We can apply an established paradigm to make sure we are being transparent with research participants, as their data may be used to train AI."

The Belmont Report arose from an effort to respond to unethical research studies, such as the [Tuskegee syphilis study](#), involving human subjects. In 1974, the U.S. created the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, and it identified the basic ethical principles for protecting people in research studies.

A [U.S. federal regulation](#) later codified these principles in 1991's

[Common Rule](#), which requires that researchers get informed consent from research participants. Adopted by many federal departments and agencies, the Common Rule was revised in 2017 to take into account changes and developments in research.

There is a limitation to the Belmont Report and Common Rule, however: The regulations that require application of the Belmont Report's principles apply only to government research. Industry, however, is not bound by them.

The NIST authors are suggesting that the concepts be applied more broadly to all research that includes human subjects. Databases used to train AI can hold information scraped from the web, but the people who are the source of this data may not have consented to its use—a violation of the "respect for persons" principle.

"For the [private sector](#), it is a choice whether or not to adopt ethical review principles," Greene said.

While the Belmont Report was largely concerned with the inappropriate inclusion of certain individuals, the NIST authors mention that a major concern with AI research is inappropriate exclusion, which can create bias in a dataset against certain demographics. Past research has shown that [face recognition algorithms](#) trained primarily on one demographic will be less capable of distinguishing individuals in other demographics.

Applying the report's three principles to AI research could be fairly straightforward, the authors suggest. Respect for persons would require subjects to provide informed consent for what happens to them and their data, while beneficence would imply that studies be designed to minimize risk to participants. Justice would require that subjects be selected fairly, with a mind to avoiding inappropriate exclusion.

Greene said the paper is best seen as a starting point for a discussion about AI and our data, one that will help companies and the people who use their products alike.

"We're not advocating more government regulation. We're advocating thoughtfulness," she said. "We should do this because it's the right thing to do."

More information: Kristen K. Greene et al, Avoiding Past Mistakes in Unethical Human Subjects Research: Moving From Artificial Intelligence Principles to Practice, *Computer* (2024). [DOI: 10.1109/MC.2023.3327653](https://doi.org/10.1109/MC.2023.3327653)

This story is republished courtesy of NIST. Read the original story [here](#).

Provided by National Institute of Standards and Technology

Citation: Researchers suggest historical precedent for ethical AI research (2024, February 15) retrieved 14 May 2024 from <https://techxplore.com/news/2024-02-historical-ethical-ai.html>

<p>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.</p>
--