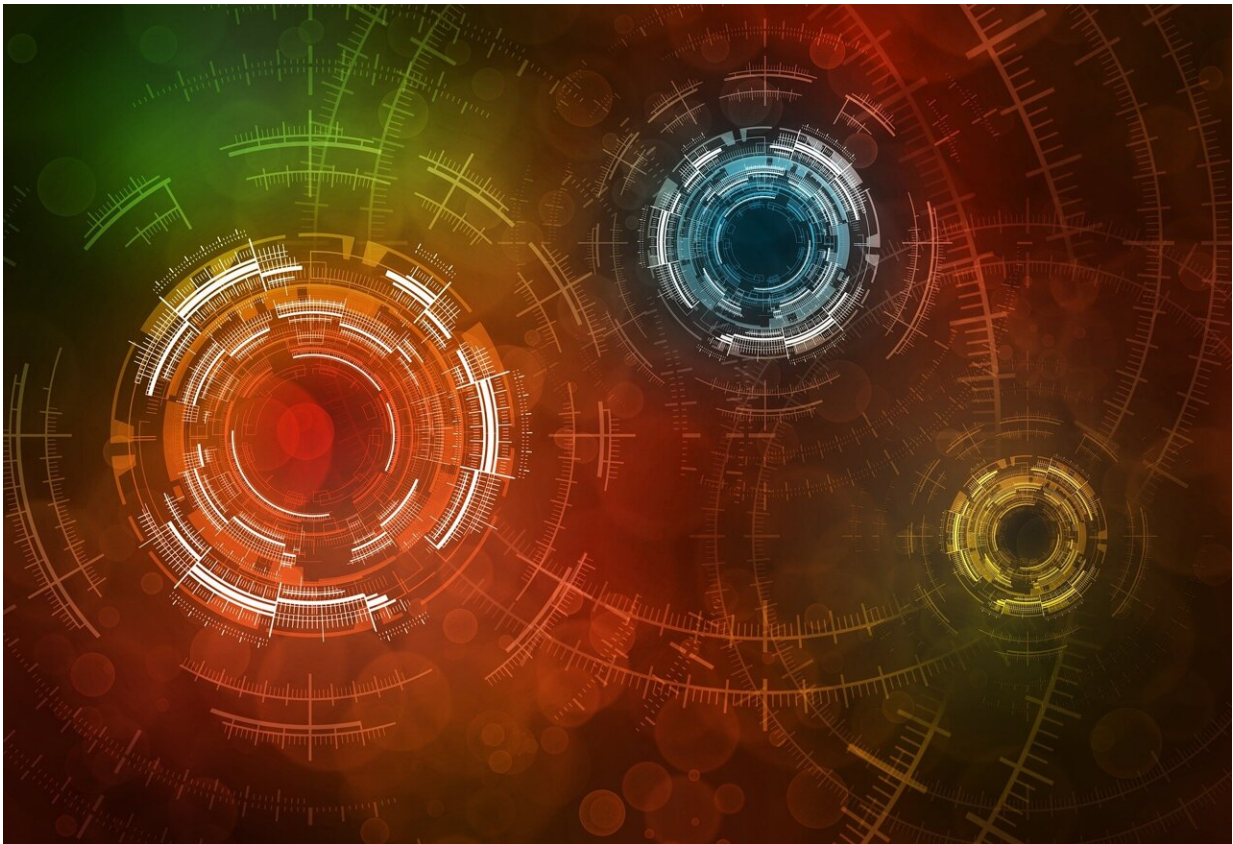


Researchers measure global consensus over the ethical use of AI

October 13 2023



Credit: Pixabay/CC0 Public Domain

To examine the global state of AI ethics, a team of researchers from Brazil performed a systematic review and meta-analysis of global guidelines for AI use. Publishing October 13 in the journal *Patterns*,

the researchers found that, while most of the guidelines valued privacy, transparency, and accountability, very few valued truthfulness, intellectual property, or children's rights. Additionally, most of the guidelines described ethical principles and values without proposing practical methods for implementing them and without pushing for legally binding regulation.

"Establishing clear ethical guidelines and governance structures for the deployment of AI around the world is the first step to promoting trust and confidence, mitigating its risks, and ensuring that its benefits are fairly distributed," says social scientist and co-author James William Santos of the Pontifical Catholic University of Rio Grande do Sul.

"Previous work predominantly centered around North American and European documents, which prompted us to actively seek and include perspectives from regions such as Asia, Latin America, Africa, and beyond," says lead author Nicholas Kluge Corrêa of the Pontifical Catholic University of Rio Grande do Sul and the University of Bonn.

To determine whether a [global consensus](#) exists regarding the ethical development and use of AI, and to help guide such a consensus, the researchers conducted a [systematic review](#) of policy and ethical guidelines published between 2014 and 2022.

From this, they identified 200 documents related to AI ethics and governance from 37 countries and six continents and written or translated into five [different languages](#) (English, Portuguese, French, German, and Spanish). These documents included recommendations, practical guides, policy frameworks, legal landmarks, and codes of conduct.

Then, the team conducted a meta-analysis of these documents to identify the most common [ethical principles](#), examine their [global distribution](#),

and assess biases in terms of the type of organizations or people producing these documents.

The researchers found that the most common principles were transparency, security, justice, privacy, and accountability, which appeared in 82.5%, 78%, 75.5%, 68.5%, and 67% of the documents, respectively.

The least common principles were labor rights, truthfulness, intellectual property, and children/adolescent rights, which appeared in 19.5%, 8.5%, 7%, and 6% of the documents, and the authors emphasize that these principles deserve more attention. For example, truthfulness—the idea that AI should provide truthful information—is becoming increasingly relevant with the release of generative AI technologies like ChatGPT. And since AI has the potential to displace workers and change the way we work, practical measures are to avoid mass unemployment or monopolies.

Most (96%) of the guidelines were "normative"—describing ethical values that should be considered during AI development and use—while only 2% recommended practical methods of implementing AI ethics, and only 4.5% proposed legally binding forms of AI regulation.

"It's mostly voluntary commitments that say, 'these are some principles that we hold important,' but they lack practical implementations and legal requirements," says Santos. "If you're trying to build AI systems or if you're using AI systems in your enterprise, you have to respect things like privacy and user rights, but how you do that is the gray area that does not appear in these guidelines."

The researchers also identified several biases in terms of where these guidelines were produced and who produced them. The researchers noted a gender disparity in terms of authorship. Though 66% of samples

had no authorship information, the authors of the remaining documents more often had male names (549 = 66% male, 281 = 34% female).

Geographically, most of the guidelines came from countries in Western Europe (31.5%), North America (34.5%), and Asia (11.5%), while less than 4.5% of the documents originated in South America, Africa, and Oceania combined. Some of these imbalances in distribution may be due to language and public access limitations, but the team says that these results suggest that many parts of the Global South are underrepresented in the global discourse on AI ethics.

In some cases, this includes countries that are heavily involved in AI research and development, such as China, whose output of AI-related research increased by over 120% between 2016 and 2019.

"Our research demonstrates and reinforces our call for the Global South to wake up and a plea for the Global North to be ready to listen and welcome us," says co-author Camila Galvão of the Pontifical Catholic University of Rio Grande do Sul. "We must not forget that we live in a plural, unequal, and diverse world. We must remember the voices that, until now, haven't had the opportunity to claim their preferences, explain their contexts, and perhaps tell us something that we still don't know."

As well as incorporating more voices, the researchers say that future efforts should focus on how to practically implement principles of AI ethics. "The next step is to build a bridge between abstract principles of ethics and the practical development of AI systems and applications," says Santos.

More information: Nicholas Kluge Corrêa et al, Worldwide AI ethics: a review of 200 guidelines and recommendations for ai governance, *Patterns* (2023). [DOI: 10.1016/j.patter.2023.100857](https://doi.org/10.1016/j.patter.2023.100857).
[www.cell.com/patterns/fulltext ... 2666-3899\(23\)00241-6](https://www.cell.com/patterns/fulltext...2666-3899(23)00241-6)

Provided by Cell Press

Citation: Researchers measure global consensus over the ethical use of AI (2023, October 13) retrieved 13 May 2024 from <https://techxplore.com/news/2023-10-global-consensus-ethical-ai.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.