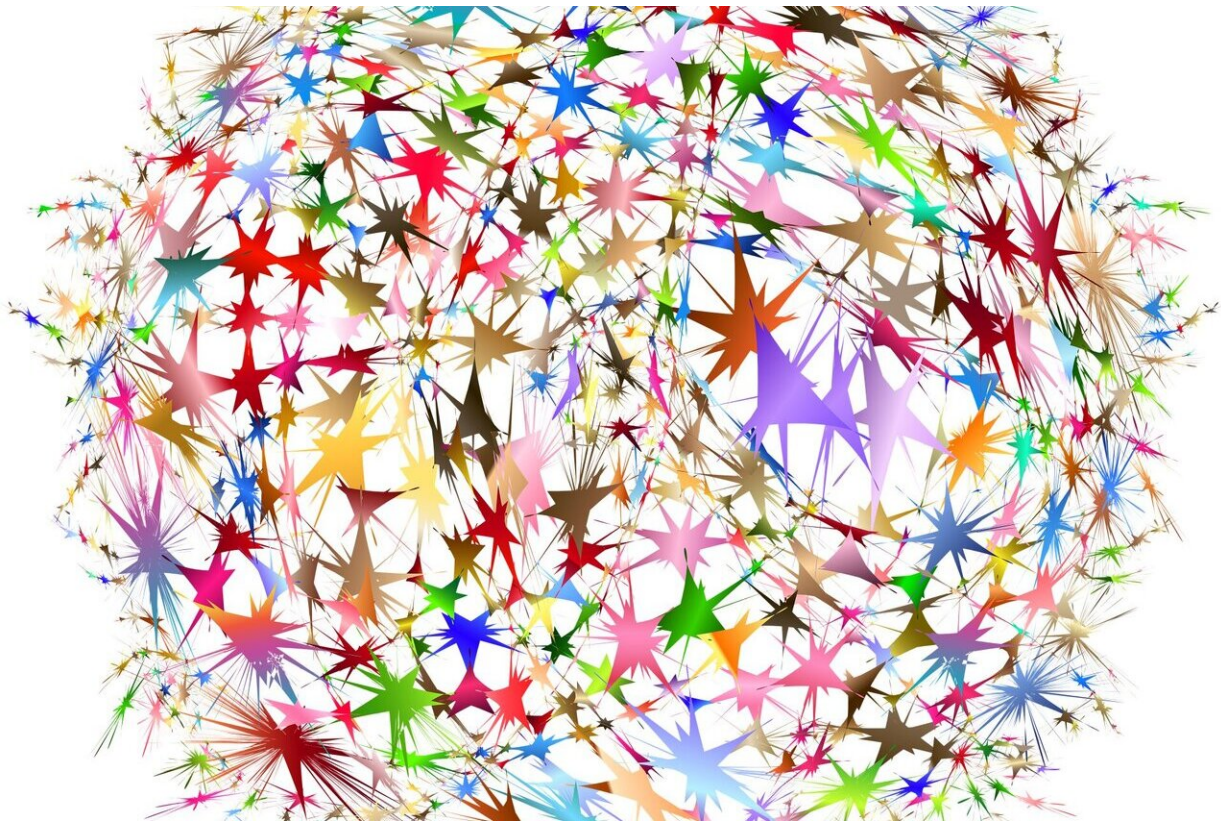


AI researchers trust international, scientific organizations most, study finds

August 9 2021



Credit: CC0 Public Domain

Researchers working in the areas of machine learning and artificial intelligence trust international and scientific organizations the most to shape the development and use of AI in the public interest.

But who do they trust the least? National militaries, Chinese tech companies and Facebook.

Those are some of the results of a new study led by Baobao Zhang, a Klarman postdoctoral fellow in the College of Arts and Sciences. The paper, "Ethics and Governance of Artificial Intelligence: Evidence from a Survey of Machine Learning Researchers," published Aug. 2 in the *Journal of Artificial Intelligence Research*.

"Both tech companies and governments emphasize that they want to build 'trustworthy AI,'" Zhang said. "But the challenge of building AI that can be trusted is directly linked to the trust that people place in the institutions that develop and manage AI systems."

AI is nearly ubiquitous in everything from recommending social media content to informing hiring decisions and diagnosing diseases. Although AI and machine learning (ML) researchers are well-placed to highlight new risks and develop technical solutions, Zhang said, not much is known about this influential group's attitudes about governance and ethics issues.

To find out more, the team conducted a survey of 524 researchers who published research at two top AI/ML conferences. The team then compared the results with those from a 2016 survey of AI/ML researchers and a 2018 survey of the U.S. public.

Zhang's group found AI and ML researchers place the most trust in nongovernmental scientific organizations and intergovernmental research organizations to develop and use advanced AI in the best interests of the public. And they place higher levels of trust in international organizations, such as the United Nations and European Union, than the U.S. public does.

AI and ML researchers generally place low to middling levels of trust in most Western technology companies and governments to develop and use advanced AI in the best interests of the public.

The survey respondents generally view Western tech companies as relatively more trustworthy than Chinese tech companies, with the exception of Facebook. This same pattern is also seen in their attitudes toward the U.S. and Chinese governments and militaries.

The findings also shed light on how AI and ML researchers think about [military](#) applications of AI. For example, the American public rated the U.S. military as one of the most trustworthy, while researchers, including those working in the U.S., place relatively low levels of trust in the militaries of countries where they do research. Though the survey respondents were overwhelmingly opposed to AI and ML researchers working on lethal autonomous weapons (74% somewhat or strongly opposed), they were less opposed to researchers working on other military applications of AI, particularly logistics algorithms (only 14% opposed).

AI and ML applications have increasingly come under scrutiny for causing harm such as discriminating against women job applicants, causing traffic or workplace accidents, and misidentifying Black people in [facial recognition software](#). Civil society groups, journalists and governments have called for greater scrutiny of AI research and deployment. The majority of researchers in the survey seem to agree that more should be done to minimize harm from their research.

More than two-thirds of respondents said research that focuses on making AI systems "more robust, more trustworthy and better at behaving in accordance with the operator's intentions" should be prioritized more highly than it is currently. And 59% think that ML institutions should conduct prepublication reviews to assess potential

harms from the public release of their research.

Zhang said she's happy to see the AI research community become more reflective of the societal and ethical impact of their work. Since she and her team conducted the survey, one of the leading ML conferences—the Conference and Workshop on Neural Information Processing Systems—began requiring a form of prepublication review for submissions.

"I think this is a move in the right direction," Zhang said, "and I hope prepublication review becomes a norm within both academia and industry."

As the authors note, "the findings should help to improve how researchers, private sector executives and policymakers think about regulations, governance frameworks, guiding principles, and national and international governance strategies for AI."

The paper's co-authors are Markus Anderljung, Noemi Dreksler and Allan Dafoe from the Centre for the Governance of AI; and Lauren Kahn and Michael C. Horowitz, from the University of Pennsylvania.

More information: Baobao Zhang et al, Ethics and Governance of Artificial Intelligence: Evidence from a Survey of Machine Learning Researchers, *Journal of Artificial Intelligence Research* (2021). [DOI: 10.1613/jair.1.12895](https://doi.org/10.1613/jair.1.12895)

Provided by Cornell University

Citation: AI researchers trust international, scientific organizations most, study finds (2021, August 9) retrieved 3 May 2024 from <https://techxplore.com/news/2021-08-ai-international->

[scientific.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.