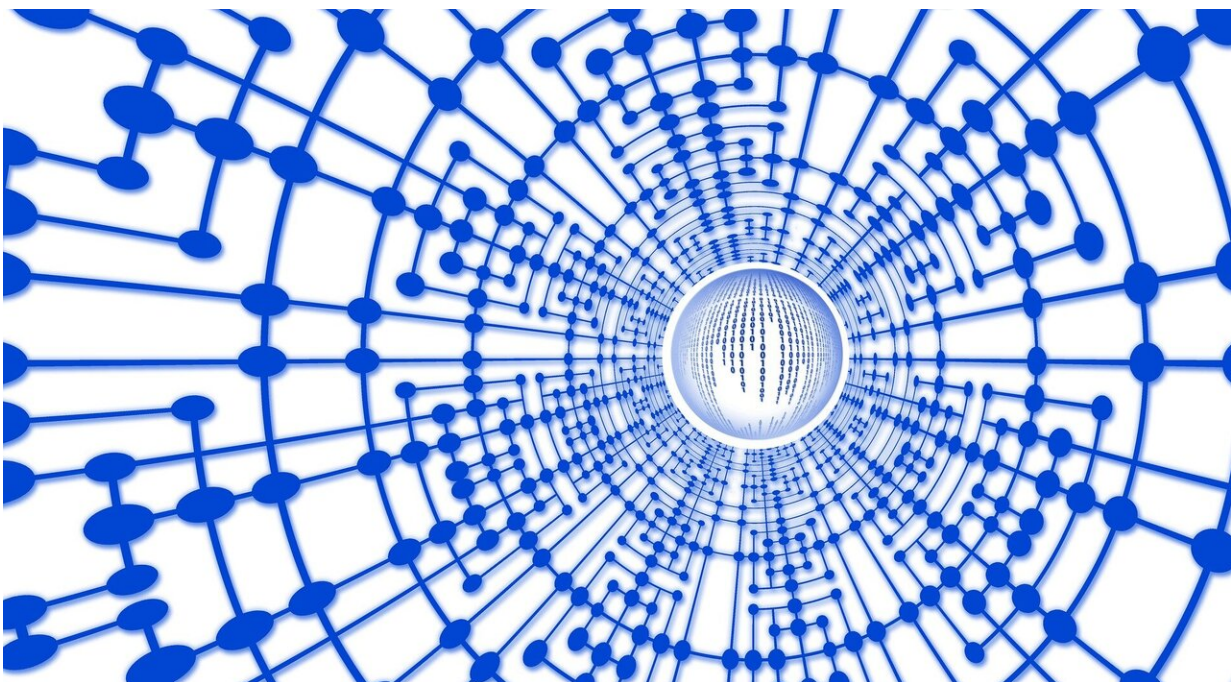


# Researchers develop new training technique that aims to make AI systems less socially biased

June 25 2024

---



Credit: CC0 Public Domain

An Oregon State University doctoral student and researchers at Adobe have created a new, cost-effective training technique for artificial intelligence systems that aims to make them less socially biased.

Eric Slyman of the OSU College of Engineering and the Adobe researchers call the novel method [FairDeDup](#), an abbreviation for fair deduplication. Deduplication means removing redundant information from the data used to train AI systems, which lowers the high computing costs of the training.

Datasets gleaned from the internet often contain biases present in society, the researchers said. When those biases are codified in trained AI models, they can serve to perpetuate unfair ideas and behavior.

By understanding how deduplication affects bias prevalence, it's possible to mitigate negative effects—such as an AI system automatically serving up only photos of white men if asked to show a picture of a CEO, doctor, etc. when the intended use case is to show diverse representations of people.

"We named it FairDeDup as a play on words for an earlier cost-effective method, SemDeDup, which we improved upon by incorporating fairness considerations," Slyman said. "While prior work has shown that removing this redundant data can enable accurate AI training with fewer resources, we find that this process can also exacerbate the harmful social biases AI often learns."

Slyman presented the FairDeDup algorithm last week in Seattle at the IEEE/CVF [Conference on Computer Vision and Pattern Recognition](#).

FairDeDup works by thinning the datasets of image captions collected from the web through a process known as pruning. Pruning refers to choosing a subset of the data that's representative of the whole [dataset](#), and if done in a content-aware manner, pruning allows for informed decisions about which parts of the data stay and which go.

"FairDeDup removes redundant data while incorporating controllable,

human-defined dimensions of diversity to mitigate biases," Slyman said. "Our approach enables AI training that is not only cost-effective and accurate but also more fair."

In addition to occupation, race and gender, other biases perpetuated during [training](#) can include those related to age, geography and culture.

"By addressing [biases](#) during dataset pruning, we can create AI systems that are more socially just," Slyman said. "Our work doesn't force AI into following our own prescribed notion of fairness but rather creates a pathway to nudge AI to act fairly when contextualized within some settings and user bases in which it's deployed. We let people define what is fair in their setting instead of the internet or other large-scale datasets deciding that."

Collaborating with Slyman were Stefan Lee, an assistant professor in the OSU College of Engineering, and Scott Cohen and Kushal Kafle of Adobe.

**More information:** Eric Slyman et al, [FairDeDup: Detecting and Mitigating Vision-Language Fairness Disparities in Semantic Dataset Deduplication](#) (2024)

Provided by Oregon State University

Citation: Researchers develop new training technique that aims to make AI systems less socially biased (2024, June 25) retrieved 29 June 2024 from <https://techxplore.com/news/2024-06-technique-aims-ai-socially-biased.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.