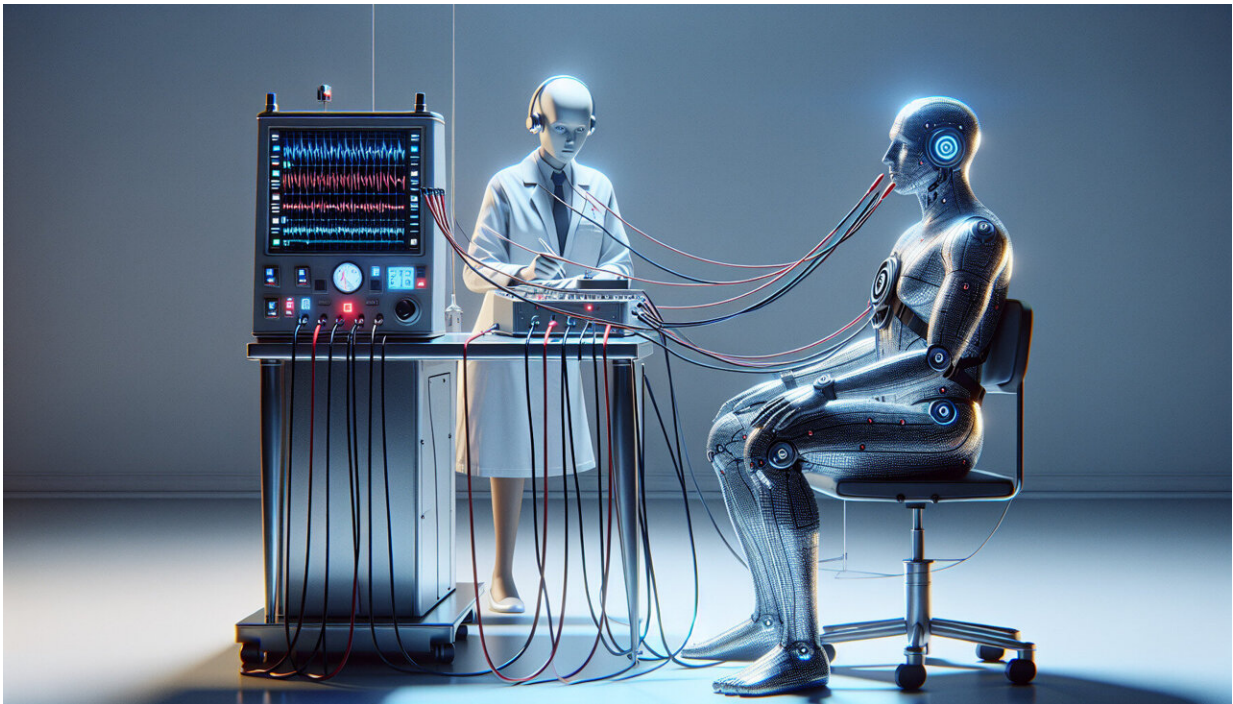


New technique helps AI tell when humans are lying

March 18 2024, by Matt Shipman



Credit: AI-generated image.

Researchers have developed a new training tool to help artificial intelligence (AI) programs better account for the fact that humans don't always tell the truth when providing personal information. The new tool was developed for use in contexts when humans have an economic incentive to lie, such as applying for a mortgage or trying to lower their

insurance premiums.

"AI programs are used in a wide variety of business contexts, such as helping to determine how large of a mortgage an individual can afford, or what an individual's [insurance premiums](#) should be," says Mehmet Caner, co-author of a paper on the work. "These AI programs generally use mathematical algorithms driven solely by statistics to do their forecasting. But the problem is that this approach creates incentives for people to lie, so that they can get a mortgage, lower their insurance premiums, and so on.

"We wanted to see if there was some way to adjust AI algorithms in order to account for these economic incentives to lie," says Caner, who is the Thurman-Raytheon Distinguished Professor of Economics in North Carolina State University's Poole College of Management.

To address this challenge, the researchers developed a new set of [training](#) parameters that can be used to inform how the AI teaches itself to make predictions. Specifically, the new training parameters focus on recognizing and accounting for a human user's economic incentives. In other words, the AI trains itself to recognize circumstances in which a human user might lie to improve their outcomes.

In proof-of-concept simulations, the modified AI was better able to detect inaccurate information from users.

"This effectively reduces a user's incentive to lie when submitting information," Caner says. "However, small lies can still go undetected. We need to do some additional work to better understand where the threshold is between a 'small lie' and a 'big lie.'"

The researchers are making the new AI training parameters publicly available, so that AI developers can experiment with them.

"This work shows we can improve AI programs to reduce [economic incentives](#) for humans to lie," Caner says. "At some point, if we make the AI clever enough, we may be able to eliminate those incentives altogether."

The research is [published](#) in the *Journal of Business & Economic Statistics*.

More information: Mehmet Caner et al, Should Humans Lie to Machines? The Incentive Compatibility of Lasso and GLM Structured Sparsity Estimators, *Journal of Business & Economic Statistics* (2024). [DOI: 10.1080/07350015.2024.2316102](https://doi.org/10.1080/07350015.2024.2316102)

Provided by North Carolina State University

Citation: New technique helps AI tell when humans are lying (2024, March 18) retrieved 9 May 2024 from <https://techxplore.com/news/2024-03-technique-ai-humans.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>
--