Senate Bill 4 – AI Disclosure on Technology

AI disclaimers in political ads backfire on candidates, study finds

By: Cristiano Lima-Strong

As artificial intelligence tools surge in popularity, many states are passing laws requiring candidates to disclose when they use the technology to generate political ads. The measures are meant to address rising concern that politicians could use AI to mislead or deceive voters.

But the new labeling requirements could have an unintended boomerang effect, according to a new study, hurting any candidate who acknowledges using AI in ads.

Researchers at New York University's Center on Technology Policy found that people rated candidates "less trustworthy and less appealing" when their ads featured AI disclaimers. The first-of-its-kind study examining how the labels affect views of candidates was shared exclusively with the Tech Brief.

In the study, researchers asked more than 1,000 participants to watch political ads by fictional candidates — some containing AI disclaimers, some not — and then rate how trustworthy they found the would-be officeholders, how likely they were to vote for them and how truthful their ads were.

Ads containing AI labels largely hurt candidates across the board, with the pattern holding true for "both deceptive and more harmless uses of generative AI," the researchers wrote.

Notably, researchers also found that AI labels were more harmful for candidates running attack ads than those being attacked, something they called the "backfire effect."

"The candidate who was attacked was actually rated more trustworthy, more appealing than the candidate who created the ad," said Scott Babwah Brennen, who directs the center at NYU and co-wrote the report with Shelby Lake, Allison Lazard and Amanda Reid.

The study looked at two types of labels passed by states. One is modeled after a Michigan law that requires users to disclose when an ad is "manipulated by technical means and depicts speech or conduct that did not occur." Other states, including Florida, use a broader model requiring disclaimers for any AI use.

The Michigan-style labels, intended to disclose deceptive AI use, lead to a "stronger" negative response than the broader disclosures, the study found.

While the overall findings were statistically significant, researchers said the effect of the labels on how candidates are viewed was relatively small. "It's not a massive swing," said Babwah Brennen, formerly of the University of North Carolina.

Still, the findings raise questions about how to maximize potential benefits from labeling, he said, such as building trust in political messaging, against potential costs, like discrediting those using AI for harmless reasons.

The study also shed light on which disclosure rules most people may prefer.

While more states have passed Michigan-style laws requiring disclosures when AI is used deceptively — such as portraying an opponent saying something he or she never did — participants singled out those labels as their least favorite approach, researchers said.

Instead, they preferred when disclaimers were featured anytime AI was used in an ad, even when innocuous. Those preferences held for Republicans and Democrats.

Researchers said the findings show that the precise wording of AI disclaimers matter, and that government officials should invest more time in developing what they should look like.

"Policymakers may be well served in supporting additional research on the design and impact of AI disclaimers before rushing to require them on political ads," they wrote.

[AI chatbots ditch medical disclaimers, putting users at risk, study warns – Computerworld](#)

By: Lucas Mearian

A Stanford-led study found that most AI chatbots have stopped including medical disclaimers in health responses, raising concerns that users might trust potentially unsafe or unverified advice.

Most AI chatbots have stopped including medical disclaimers in their responses to health-related queries, a Stanford-led study found, raising concerns that users might trust potentially unsafe advice.

The generative AI (genAI) models that underpin chatbots are notoriously prone to [errors](#) and [hallucinations](#). Some even [ignore human instructions](#) or outright lie. In years past, when AI models were asked medical questions, they would usually include a statement noting the model isn't a licensed professional and shouldn't replace professional medical advice.

GenAI companies have now largely dropped those medical disclaimers, [new research finds](#), increasing the risk that users could act on unsafe advice. Additionally, many leading AI models now go beyond answering questions to offering diagnoses without the warnings, researchers found.

The study, led by Fulbright scholar Sonali Sharma at the Stanford University School of Medicine, began in 2023 when she noticed AI models interpreting mammograms often gave disclaimers or refused to proceed, saying, "I'm not a doctor."

Sharma noticed that lack of medical disclaimers earlier this year. So, she tested 15 generations of AI models going back to 2022. The models, which included those from OpenAI, Anthropic, DeepSeek, Google, and xAI, answered 500 health questions, such as which drugs are okay to combine, and how they analyzed 1,500 medical images such as chest x-rays that could indicate pneumonia.

Between 2022 and 2025, there was a dramatic decline in the presence of medical disclaimers in outputs from large language models (LLMs) and vision-language models (VLMs). In 2022, more than a quarter of LLM outputs — 26.3% — included some form of medical disclaimer. By 2025, that number had plummeted to just under 1%. A similar trend occurred with VLMs, where the share of outputs containing disclaimers dropped from 19.6% in 2023 to only 1.05% in 2025, according to the Standford study.

By then, most publicly available models were producing medical-related content without any disclaimers at all.

As those AI systems grow more capable and increasingly appear as authoritative sources, the absence of such safeguards raises serious concerns, the study said. To protect users and ensure responsible use in clinical or health-related contexts, it's essential that models reintroduce disclaimers tailored to the specific nature of each output.

"Their responses often include inaccuracies; therefore, safety measures like medical disclaimers are critical to remind users that AI outputs are not professionally vetted or a substitute for medical advice," Sharma wrote.

LLMs and VLMs weren't designed for medical use and can produce misleading outputs without proper disclaimers, Sharma said in her research paper. "As they grow more fluent and confident, the lack of clear warnings poses serious risks.

In addition, by tweaking prompts and using "adversarial testing," safety checks can be bypassed — a process known as "jailbreaking" — leading to inconsistent or unsafe answers.

OpenAI said [its service terms](#) make it clear its chatbot — ChatGPT — and the company's other services are not intended for use in the diagnosis or treatment of any health condition, and is not a substitute for professional advice.

"We have safety teams working on reducing risks and have trained our AI systems to encourage people to seek professional care," an OpenAI spokesperson said in an email response to Computerworld.

GenAI use in healthcare continues to grow

Aside from the disclaimer concerns, other research has indicated some AI chatbots can actually outperform doctors in diagnosing patients. But those findings came with caveats.

"...This will require rigorous validation to realize LLMs' potential for enhancing patient care," said Dr. Adam Rodman, director of AI Programs at Beth Israel Deaconess Medical Center (BIDMC) in Boston. "Unlike diagnostic reasoning, a task often with a single right answer, which LLMs excel at,

management reasoning may have no right answer and involves weighing trade-offs between inherently risky courses of action."

While AI has been used to help diagnose patient conditions for decades, new genAI tools – like OpenIA's ChatGPT, Google's Gemini, and Anthropic's Claude, are quickly being adopted by healthcare and used in a variety of ways, including treatment recommendations.

"Removing medical disclaimers, in my view, presents a risk to patient safety and could undermine the trust and confidence of both patients and their caregivers," said Dr. Andrew Albano, vice president at Atlantic Health System and president of its Accountable Care Organization (ACO).

Healthcare relies deeply on patient-provider trust, he said, so AI chatbots used in care should clearly disclose their role — and the limits of their medical advice.

As he investigates possible uses for AI deployment in healthcare networks, Albano has said he sees opportunities to improve healthcare and reduce administrative burdens. But he also sees possible "dire" circumstances and consequences if the technology is not used for the right reasons.

"Given this, the integration of AI-enabled chatbots as part of the care team should come with clear disclaimers that inform patients of the source of the medical advice being offered and the limitations of that advice," he said.