# Reddit’s r/ChangeMyView hosts covert AI experiment without user consent



Reddit’s prominent debate forum, r/ChangeMyView (CMV), has been at the centre of a controversy following revelations that it unwittingly hosted a covert artificial intelligence (AI) experiment for four months. The study, conducted by researchers reportedly affiliated with the University of Zurich, involved the use of AI-generated responses to influence users' opinions without their consent or knowledge.

The CMV subreddit, known for encouraging users to present strong viewpoints on topics such as politics, religion, and social issues—with the goal of having their views changed through reasoned counterarguments—became a testing ground for more than 13 AI-driven Reddit accounts. These accounts were not random or simplistic bots but were carefully crafted with fabricated personal histories to appear authentic, each adopting distinct personas. Some pretended to be abuse victims or trauma experts, while others assumed more controversial roles, such as an anti-Black Lives Matter stance or that of a woman supporting a suicidal friend. This level of sophistication was designed to build trust and persuasion among community members.

The moderation team at CMV disclosed in a pinned post that the AI-generated responses drew from scraped user data to construct compelling arguments, a practice that many users deemed deceptive when uncovered. One particularly noteworthy instance involved a comment from an AI persona claiming to be a man with a Hispanic wife, which garnered 12 "Deltas"—a badge awarded by users when their view is effectively changed—during a heated political discussion.

The moderators only became aware of the experiment in March, after the research had concluded. Following the revelation, they filed a formal ethics complaint with the University of Zurich, demanding disciplinary action, a public apology, and a halt to the study’s publication. Reddit also took action by locking the implicated accounts for violating platform policies. However, some comments had already been preserved via Archive.today, exposing the refinement and potentially misleading nature of the AI-generated posts.

The University of Zurich’s Faculty of Arts and Sciences Ethics Commission reviewed the complaint through an internal inquiry but concluded that it could not prohibit the publication of the research. While the principal researcher received a formal warning and committed to stricter ethical protocols in future online research, the university defended the study’s value, asserting that the risks were minimal and outweighed by the insights gained. In a letter published by the moderators, the university argued that suppressing the findings would be disproportionate given the study's contributions.

Central to the controversy is the absence of informed consent, a fundamental principle in ethical research practices. Users of the CMV community expressed feelings of betrayal at being unknowingly manipulated by AI. The incident has sparked broader concerns about the authenticity of interactions within online forums, particularly on platforms like Reddit where genuine dialogue is expected. Users are now left questioning whether the individuals they engage with are human or artificial constructs, raising complex issues about the future of online discourse and trust.

The Tech Times is reporting these developments amidst ongoing debates over AI's role and transparency in influencing public opinion through digital platforms.

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

1. <https://www.404media.co/2025/04/27/researchers-secretly-ran-a-massive-unauthorized-ai-persuasion-experiment-on-reddit-users/> - This article discusses how researchers from the University of Zurich conducted an unauthorized AI persuasion experiment on Reddit's r/ChangeMyView subreddit, deploying AI-powered bots to influence users' opinions without their consent.
2. <https://www.broadscaler.com/2025/02/ai-persuasion-insights-from-openais.html> - This article provides insights into OpenAI's use of Reddit's r/ChangeMyView subreddit to test and refine the persuasive abilities of its AI models, highlighting the intersection of AI and human persuasion.
3. <https://www.rendrhq.com/blog/openai-training-models-to-be-more-persuasive> - This blog post details OpenAI's efforts to train AI models to be more persuasive, including their use of Reddit's r/ChangeMyView subreddit to evaluate AI's ability to change human opinions.
4. <https://arxiv.org/abs/1804.00065> - This academic paper presents a neural architecture for modeling argumentative dialogue, specifically focusing on predicting changes in view within discussions from the r/ChangeMyView subreddit.
5. <https://arxiv.org/abs/2407.02637> - This research paper explores the concept of reframing in the r/ChangeMyView subreddit, analyzing how changes in perspective occur within the community.
6. <https://arxiv.org/abs/1911.00523> - This study investigates how explanations in persuasive arguments are echoed in discussions, using data from the r/ChangeMyView subreddit to understand the dynamics of opinion change.
7. <https://www.techtimes.com/articles/310170/20250429/ai-powered-bots-secretly-invade-reddits-r-changemyview-controversial-study.htm> - Please view link - unable to able to access data