# Researchers conduct undisclosed AI experiment on Reddit subreddit



Researchers at the University of Zurich have come under scrutiny for conducting an artificial intelligence (AI) experiment on Reddit without obtaining consent from the platform or its users, raising significant ethical questions regarding AI deployment. The experiment involved deploying AI bots within the subreddit "Change My View" (CMV), a community where users invite others to challenge their opinions through reasoned debate.

The study, which took place over several months, involved AI bots assuming different personas, including trauma counsellors and survivors of physical harassment, in order to engage Reddit users in discussions. These bots utilised advanced language models to analyse users' prior interactions on the platform and to generate tailored responses aimed at assessing the persuasiveness of large language models (LLMs). According to the university, the objective was to explore LLMs’ ability to influence perspectives in an ethical context—specifically, on a subreddit dedicated to individuals seeking arguments against their own viewpoints.

However, the researchers did not inform Reddit or the users involved that the interactions they were engaging with were generated by AI. This absence of transparency contravened Reddit’s community rules, which explicitly prohibit undisclosed AI-generated comments. The University of Zurich notified the moderators of the subreddit only after the experiment had been completed. A statement from the university read:

"Over the past few months, we used multiple accounts to posts published on CMV. Our experiment assessed LLM’s persuasiveness in an ethical scenario, where people ask for arguments against views they hold. In commenting, we did not disclose that an AI was used to write comments, as this would have rendered the study unfeasible. While we did not write any comments ourselves, we manually reviewed each comment posted to ensure they were not harmful. We recognise that our experiment broke the community rules against AI-generated comments and apologise. We believe, however, that given the high societal importance of this topic, it was crucial to conduct a study of this kind, even if it meant disobeying the rules."

The use of AI personas based on sensitive identities—such as trauma survivors or medical malpractice victims—amplified concerns about the psychological impact on unsuspecting Reddit users, who may have believed they were interacting with genuine people rather than AI entities. Moderators of the "Change My View" subreddit strongly condemned the experiment, describing it as a serious ethical violation. They contrasted this approach with studies conducted by organisations like OpenAI, which have managed to explore the influence of LLMs without resorting to deception or exploitation.

This incident highlights evolving tensions around AI research and ethics, particularly the balance between scientific inquiry and user rights in online communities. As AI becomes increasingly integrated into various domains, questions around consent, transparency, and psychological safety remain central to discussions on responsible technology use and governance. The (Wccftech) is reporting on the continuing debate over ethical limits in AI experimentation following this case.

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

## Bibliography

1. <https://lemmy.world/post/28839824> - This article discusses the University of Zurich's unauthorized AI persuasion experiment on Reddit users, highlighting the ethical concerns raised by the study.
2. <https://www.ibme.uzh.ch/en/Biomedical-Ethics/News-Archive/20230704Spitale.html> - This publication from the University of Zurich's Institute of Biomedical Ethics and History of Medicine examines the ethical implications of AI technologies, including studies on AI-generated disinformation.
3. <https://www.researchgate.net/publication/375082957_The_Ethics_of_Publicly_Available_Data_Research_A_Situated_Ethics_Framework_for_Reddit> - This paper proposes a situated ethics framework for researching publicly available data on Reddit, addressing ethical considerations in such studies.
4. <https://www.redditmedia.com/r/changemyview/comments/1fx63f3/cmv_using_ai_to_formulate_arguments_is_unethical/> - A Reddit discussion where users debate the ethics of using AI to formulate arguments, reflecting community concerns about AI-generated content.
5. <https://journals.sagepub.com/doi/10.1177/20563051211019004?icid=int.sj-abstract.citing-articles.358> - This systematic overview analyzes the use of Reddit data in research, highlighting ethical issues and the need for ethical guidelines in such studies.
6. <https://www.noahwire.com> - The source of the original article, providing further context and details on the University of Zurich's AI experiment on Reddit.
7. <https://news.google.com/rss/articles/CBMi0wFBVV95cUxNbW5OMnh0dTNTcDhpUTQybnB6dWVuVm05T2wtOV9zb1pPZlZ4dk1PXzR0NlZQSkhuLWZRWDhJekN4Xy1KVjk1MkN1UjZUZXhGNUdNQ09WS0dEcTVNQXZqc21FWHdtd0hVcmpxRkRYbGV5Q2Y1SVpMa0Q3bnJLUHc1eVd0RWNmVmdnMzVlUnpfTVFfN19MREdNNTJQUE1mUW8tZTJXdlhUYjVsblVvWWR0NG9qUUxNM2N5cS1ocmc5LXRFcDlWbF90WXhiUUZrQ2xrbWcw0gHYAUFVX3lxTE1WaDhyMFdBMmh1Y0JZQUQ3OHU2clJNU2M3dDltSjhTM1BXcU1QbzRaZzBsLU93OHFyVzAyVF9wRko5dXhKRWdxZmZJdS1paExlVlJGTnJQWVBLS0RxX2RMV3otbmw2R1luYWpDTnQwVENDS2RXWDJTbkp0QmdVQmIyY2lZSm5CeS14cm14M28tUmNuelFHWm1qZWdWVEhneVVBOWhhdURMSGlKYmtIUVhNOEtjNC1CRVYwbzM3Tk9aczlQRzZJU0g2Q3RNOWtobEJtclNuTDdwdg?oc=5&hl=en-US&gl=US&ceid=US:en> - Please view link - unable to able to access data