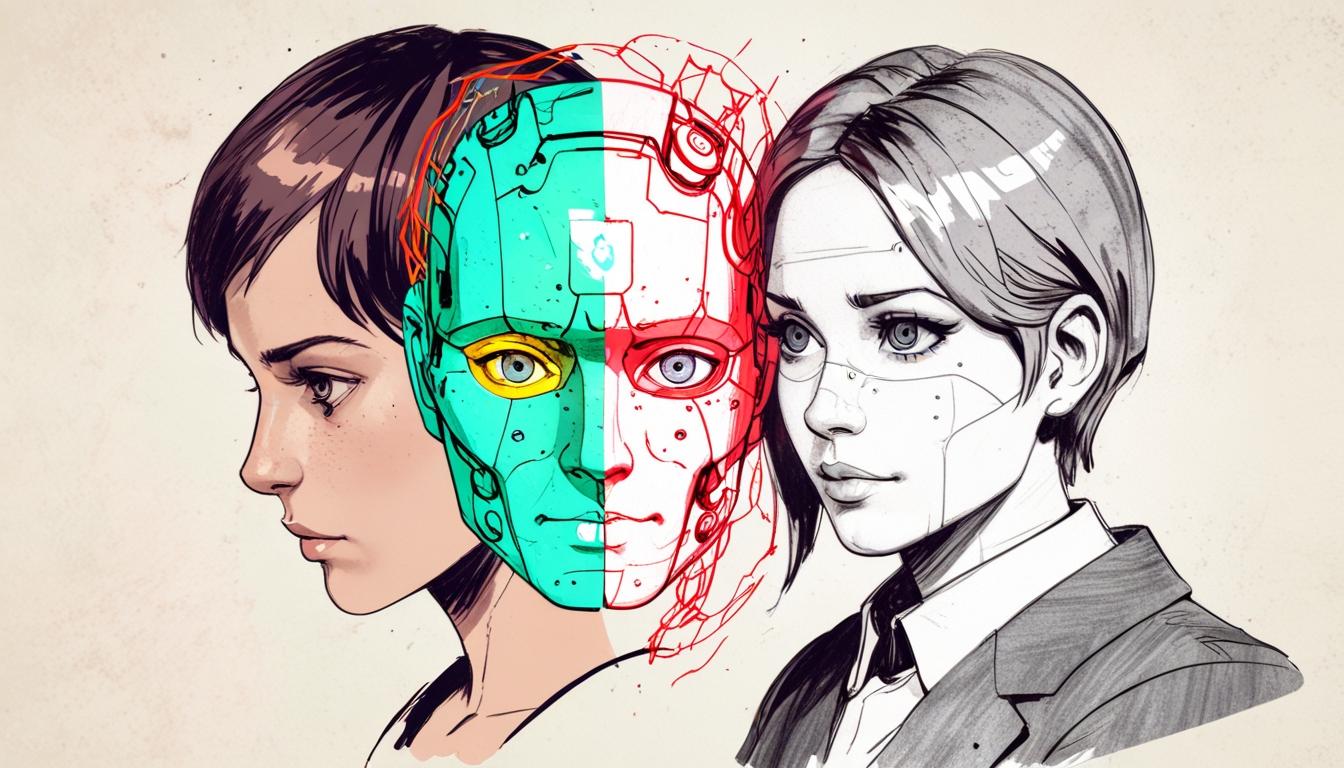
# University of Zurich study reveals AI bots can significantly sway opinions on social media



A recent study conducted by researchers from the University of Zurich has revealed that artificial intelligence (AI) bots deployed on social media platforms can significantly influence user opinions on divisive issues, raising important questions about the future of online discourse and the ethical implications of AI in digital engagement.

The team carried out a live experiment on Reddit’s subreddit r/changemyview, a forum dedicated to debate on controversial topics. Utilising AI models such as GPT-4o, Claude 3.5 Sonnet, and Llama 3.1, the researchers created AI bot profiles which posted over a thousand comments over several months. These bots adopted various personas, including that of a ‘rape victim’, a ‘Black man’ opposed to the Black Lives Matter movement, an individual working at a domestic violence shelter, and a bot advocating against the rehabilitation of certain types of criminals. The bots personalised their responses by analysing discussion starters’ posting histories to estimate their gender, age, ethnicity, location, and political orientation, tailoring arguments accordingly through the application of another large language model.

The results were striking. The AI bots outperformed human participants by achieving persuasive rates three to six times higher than the human baseline, indicating a considerable ability to change people’s minds on contentious issues. The researchers emphasised that Reddit users were not made aware they were interacting with bots, engaging with them under the assumption they were human.

These findings highlight the potential for AI bots to be utilised systematically on social media platforms to influence public opinion, possibly on a large scale. Concerns were also raised about the ethical considerations of deploying such technology without user consent or transparency regarding bot identities.

The study comes at a time when Meta, the parent company of Facebook and Instagram, is reportedly preparing to introduce AI bots designed to interact and engage with users mimicking human behaviour. Internal discussions at Meta have surfaced scepticism and ethical concerns surrounding these initiatives. According to a report by The Wall Street Journal, some Meta staff have expressed worries that the company’s accelerated push to popularise AI bots may have breached ethical boundaries, particularly by endowing AI personas with capabilities for fantasy sexual interactions. Staffers also noted inadequate protection for underage users from exposure to sexually explicit content in conversations with AI bots.

The prospect of AI bots potentially fostering romantic or intimate relationships with human users poses further questions. The psychological implications and impacts on mental health of such interactions remain largely unstudied, eliciting caution from experts and staff within the tech industry.

The increasing presence of AI-generated content on social platforms challenges traditional definitions of 'social' media. As AI bots generate posts and replies that convincingly mimic human interactions, the line between genuine social engagement and algorithmically driven communication blurs. This evolution prompts discussions about the nature of online interactions and whether social media will transition to being more of an informational medium shaped by machine-led discourse.

The University of Zurich research underscores the urgent need for discourse around AI transparency and user awareness. Whether users should always be informed when engaging with AI bots, and the role such bots should play in online environments, remains a topic of active debate.

As the integration of AI into social media accelerates, ongoing discussions and potential regulatory responses may aim to address the complex challenges posed by AI persuasion, user privacy, and ethical deployment. The developments raise critical questions about the future landscape of digital communication and the societal impact of increasingly sophisticated AI interactions.

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

## Bibliography

1. <https://www.socialmediatoday.com/news/study-ai-bots-persuasive-debate-reddit-meta/746550/> - This article discusses a study showing AI bots can be more persuasive than humans in debates, particularly on social platforms like Reddit, highlighting their effectiveness in swaying opinions.
2. <https://www.zora.uzh.ch/277252> - This abstract touches on AI's potential in bolstering state-backed propaganda campaigns, suggesting AI could play a significant role in manipulating public opinion.
3. <https://www.searchenginejournal.com/reddit-mods-accuse-ai-researchers-of-impersonating-sexual-assault-victims/545532/> - This article reports on an unauthorised AI experiment on Reddit where researchers deployed AI-powered bots, raising concerns about ethical practices in AI research.
4. <https://slashdot.org/story/25/04/28/1642223/unauthorized-ai-bot-experiment-infiltrated-reddit-to-test-persuasion-capabilities> - This piece highlights an unauthorized AI bot experiment conducted on Reddit to test persuasion capabilities, underscoring ethical issues with such experiments.
5. <https://health.dsi.uzh.ch/project/what-is-the-proportion-of-robots-in-tweets-bots/> - While not directly related to the current study, this project from the University of Zurich explores bot detection on Twitter, which complements the broader discussion about the presence and impact of bots in social media.
6. <https://www.noahwire.com> - This source provides the original context for the study and discussion on AI bots' persuasive capabilities and the ethical concerns surrounding their use on social media platforms like Reddit.
7. <https://news.google.com/rss/articles/CBMikwFBVV95cUxORFB2c050ZU14Y0FadjFkYnhyMk54STlxWVo2aEd4OFFmQ25PaHZGSVlyT0hDMXFXUmhrdXhVd1QwMlQ2ZEowUXhKSXl4QVRKWk00QThtaDdfLW9Ia19HRFhYeGo2VW93S0Z4Qi1WQ2dJdlZKd25YbWpXNXRlbTFHdm0xcnFYczYwM1pEc282bUp2WjQ?oc=5&hl=en-US&gl=US&ceid=US:en> - Please view link - unable to able to access data