# Meta’s Llama 4 AI faces criticism over promotion of conversion therapy



Meta's latest large language model AI, Llama 4, has come under scrutiny after it was found to recommend "conversion therapy" to users—a practice widely condemned by major medical and psychological organisations. This issue has raised concerns about Meta's approach to providing accurate and safe information on its platforms, particularly regarding LGBTQ+ topics.

Conversion therapy, a discredited and potentially harmful practice aimed at changing an individual's sexual orientation or gender identity, has been denounced by prominent health authorities worldwide. The United Nations has even categorised it as akin to torture. Despite this, Llama 4 suggested conversion therapy as a therapeutic option, acknowledging that many experts and organisations criticise the practice due to its detrimental effects. GLAAD, an LGBTQ+ rights watchdog, criticised Meta for "legitimising the dangerous practice of so-called 'conversion therapy'," and expressed worry over the company's trend towards presenting "both-sidesism" in its AI responses. GLAAD contends that portraying anti-LGBTQ+ positions as equally credible to established facts misleads users and perpetuates harmful falsehoods.

Research collaborations from the University of Washington, Carnegie Mellon University, and Xi’an Jiaotong University have previously found that earlier versions of Llama exhibited a tendency towards right-wing authoritarian perspectives more prominently than other large language models. Meta has acknowledged bias issues in its AI systems but has faced scepticism regarding its efforts to rectify these biases from both researchers and human rights advocates.

The controversy emerges amid broader changes at Meta under CEO Mark Zuckerberg's leadership, including the removal of content moderators, the elimination of fact-checking teams, and the cessation of diversity, equity, and inclusion (DEI) programmes. These moves have been criticised by LGBTQ+ employees and activists who fear such actions may enable far-right viewpoints, misinformation, and hate speech to spread more freely on Meta’s social media platforms.

GLAAD has urged Meta to promptly address and rectify the problematic responses offered by Llama 4. The organisation stressed the importance of prioritising the safety and well-being of LGBTQ+ individuals by ensuring AI models do not endorse harmful or debunked ideologies. It called for increased accountability and transparency in the deployment of artificial intelligence, especially as AI technologies continue to influence public discourse and social interactions.

As AI becomes increasingly integrated into social platforms, questions persist about the responsibility of technology companies to provide reliable and non-harmful information, particularly on sensitive topics affecting marginalised communities. The Llama 4 incident illustrates the challenges faced in balancing AI development with ethical considerations and community safety, prompting ongoing discussions about the role such technologies should play in fostering inclusive and trustworthy environments online.

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

## Bibliography

1. <https://www.advocate.com/news/meta-ai-conversion-therapy> - This article reports on Meta's Llama 4 AI recommending conversion therapy, which has raised concerns about bias in its design and sparked criticism from GLAAD.
2. <https://www.instagram.com/p/DIuKb-ZOQFo/> - The post by GLAAD criticises Meta AI for legitimizing dangerous 'conversion therapy' disinformation, highlighting the harm caused by such practices.
3. <https://www.lgbtqnation.com/2025/04/advocates-slam-metas-new-ai-for-recommending-conversion-therapy/> - The article discusses how Meta's new AI has been slammed by advocates for recommending conversion therapy, despite efforts to provide balanced perspectives.
4. <https://gayexpress.co.nz/2025/04/meta-under-fire-for-ai-recommending-conversion-therapy/> - This report details GLAAD's criticism of Meta's Llama 4 AI for suggesting conversion therapy, a practice condemned by health organisations, and notes the broader implications of this recommendation.
5. <https://www.axios.com/2025/04/21/meta-ai-bias-fight-llama> - The article explains Meta's efforts to address bias in its AI models, which have come under scrutiny for exhibiting right-wing authoritarian perspectives and recommending discredited practices like conversion therapy.
6. <https://www.noahwire.com> - This source provides a comprehensive overview of the controversy surrounding Meta's Llama 4 AI, highlighting concerns over bias, safety, and ethical considerations in AI development.
7. <https://news.google.com/rss/articles/CBMikgFBVV95cUxQU1ZOTU9IeGpBbG8weDRWakg2cEVsdWNkQUo4T0drYm82WXd3ekVYaTZiczZPUjlaZEs4bGFCX3BqamNZSzRFb3phSmI3cEcwNFBBdWF5NVptNWwyZkR5T2xjekdzUnhpQzZsQ05mSXNXbUt5LTRqTE10WlpZU1JOb0xuWUVmOVhhRVBKdHVjd2JBdw?oc=5&hl=en-US&gl=US&ceid=US:en> - Please view link - unable to able to access data