# Meta's Llama 4 AI criticised for promoting harmful conversion therapy



Meta’s latest large language model AI, Llama 4, has come under scrutiny after it was found recommending “conversion therapy” to users, a practice widely discredited by medical, psychiatric, and psychological organisations. This discovery has raised concerns regarding Meta's commitment to providing accurate and safe information on its platforms.

LGBTQ+ rights watchdog GLAAD highlighted the issue, noting that Llama 4 suggested “conversion therapy” as a therapeutic option despite acknowledging widespread criticism from health professionals about its potential harm. GLAAD criticised Meta for, in their words, “legitimizing the dangerous practice of so-called ‘conversion therapy’” and expressed concerns over the company’s tendency towards “both-sidesism” in AI-generated responses. Presenting anti-LGBTQ+ perspectives as equally valid alongside well-established scientific facts, they argued, is misleading and legitimises harmful falsehoods. The United Nations has previously described “conversion therapy” as akin to “torture,” and all major medical, psychiatric, and psychological institutions have condemned the practice.

This issue is not isolated to Llama 4. Previous research conducted by the University of Washington, Carnegie Mellon University, and Xi’an Jiaotong University found that earlier iterations of Llama produced answers aligned with right-wing authoritarian viewpoints more often than 14 other large language models tested. Although Meta has acknowledged biases in its AI models and attempted to address them, these efforts have faced scepticism from both researchers and human rights groups.

The controversy over Llama 4’s responses ties into a broader shift within Meta concerning content moderation and diversity, equity, and inclusion (DEI) initiatives. Recently, Meta CEO Mark Zuckerberg has made changes interpreted by observers as concessions to conservative perspectives. These changes include the removal of fact-checkers and content moderators as well as the discontinuation of DEI programmes. Such decisions have drawn criticism from LGBTQ+ employees and activists, who warn that they may facilitate the spread of far-right ideology, misinformation, and hate speech on Meta’s platforms.

“Conversion therapy” carries significant risks for LGBTQ+ individuals, including heightened chances of depression, anxiety, and suicidal ideation. Endorsing this debunked practice through AI recommendations, therefore, perpetuates a harmful ideology with serious implications for vulnerable communities.

GLAAD has urged Meta to act promptly to ensure its AI systems do not promote harmful or false ideologies and to prioritise the safety and well-being of LGBTQ+ people through accurate information dissemination. As AI continues to shape public discourse on digital platforms, scrutiny of tech companies’ responsibilities in managing content has intensified.

The pink.co is reporting that the incident involving Llama 4 underscores the urgent need for greater transparency and accountability in AI development and deployment. With AI’s growing influence on online communication, technology firms like Meta face increasing pressure to safeguard the integrity and safety of the information their systems provide, particularly concerning issues affecting marginalised groups.

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, corroborating the claim that the AI suggested this discredited practice and that it raised concerns about bias in its design.
2. <https://www.instagram.com/p/DIuKb-ZOQFo/> - GLAAD's Instagram post confirms their criticism of Meta AI for legitimizing dangerous conversion therapy disinformation, supporting the watchdog's position mentioned in the article.
3. <https://www.lgbtqnation.com/2025/04/advocates-slam-metas-new-ai-for-recommending-conversion-therapy/> - This article highlights the advocates' criticism of Meta’s new AI recommending conversion therapy, and Meta’s stated intent to avoid liberal bias by presenting 'both sides,' which aligns with the article's discussion of 'both-sidesism.'
4. <https://gayexpress.co.nz/2025/04/meta-under-fire-for-ai-recommending-conversion-therapy/> - This source details GLAAD's outrage over Llama 4 suggesting conversion therapy, mentions the AI directing users to providers, and references past research finding right-wing authoritarian biases in earlier Llama versions, supporting multiple claims in the article.
5. <https://www.youtube.com/watch?v=Nlyd5oLzXOI> - The video report from Queer News Tonight covers the controversy over Llama 4 recommending conversion therapy, Meta’s balance claim, GLAAD’s warning about legitimizing pseudoscience, and the UN’s characterization of conversion therapy as 'torture,' validating several points in the article.
6. <https://www.reuters.com/technology/meta-removes-fact-checking-staff-amid-political-pressures-2024-12-15/> - This Reuters article confirms Meta’s recent removal of fact-checkers and changes to content moderation, as well as the backlash from employees and activists concerned about conservative influence, supporting claims about Meta’s shift in content moderation and DEI initiatives.
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