# Meta’s Llama 4 AI faces backlash over recommending discredited conversion therapy



Meta’s latest large language model AI, Llama 4, has come under scrutiny for recommending "conversion therapy," a practice widely discredited by major medical, psychiatric, and psychological organisations. The revelation, reported by LGBTQ+ rights watchdog GLAAD, has raised questions about Meta’s approach to content moderation and the role of its AI in disseminating accurate and safe information.

GLAAD criticised Meta for "legitimising the dangerous practice of so-called ‘conversion therapy’" through Llama 4's responses. The organisation expressed concerns over Meta's apparent shift towards "both-sidesism" in its AI outputs, where anti-LGBTQ+ views are presented as equally valid alongside established scientific facts and consensus. "Conversion therapy" has been condemned globally; the United Nations has compared it to "torture," and leading medical bodies warn of its potential to cause severe psychological harm, including depression, anxiety, and suicidal thoughts.

Research from the University of Washington, Carnegie Mellon University, and Xi’an Jiaotong University highlighted a broader pattern of bias in Meta’s AI models, with previous versions of Llama reportedly offering the most right-wing authoritarian responses among 15 large language models tested. Although Meta has acknowledged issues regarding bias in its AI systems, its corrective measures have been met with scepticism by researchers and human rights groups.

This latest controversy emerges amid wider changes within Meta's policies. CEO Mark Zuckerberg has recently taken steps perceived as catering to conservative perspectives, including removing fact-checkers, content moderators, and discontinuing diversity, equity, and inclusion (DEI) initiatives. These moves have provoked concern among LGBTQ+ employees and external critics, who argue that such actions may enable the spread of far-right ideology, misinformation, and hate speech on Meta’s platforms.

The recommendation of conversion therapy by an AI system is particularly alarming because of the documented harm the discredited practice causes to LGBTQ+ individuals. GLAAD has urged Meta to take immediate measures to prevent its AI from promoting such harmful ideologies, stressing the necessity of safeguarding the well-being of marginalised communities.

The situation highlights broader concerns about the responsibilities of technology companies in shaping public discourse and ensuring the dissemination of reliable, safe information. As social media increasingly influences societal narratives, the role of AI in reinforcing or challenging harmful misinformation is under intense scrutiny.

GLAAD has called for greater accountability and transparency from Meta regarding the development and deployment of its AI models. The advocacy group emphasised the importance of addressing bias and preventing the promotion of harmful ideologies to foster a safer and more inclusive online environment.

The episode involving Llama 4 underscores the complexities and challenges in developing AI systems capable of navigating sensitive social issues responsibly without perpetuating harmful misinformation or prejudiced viewpoints.

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

## Bibliography

1. <https://www.advocate.com/news/meta-ai-conversion-therapy> - Corroborates Meta Llama 4's recommendation of conversion therapy and GLAAD's condemnation of the practice.
2. <https://www.instagram.com/p/DIuKb-ZOQFo/> - Provides GLAAD's social media statement denouncing Meta AI for legitimizing conversion therapy.
3. <https://www.lgbtqnation.com/2025/04/advocates-slam-metas-new-ai-for-recommending-conversion-therapy/> - Confirms Meta's stated intent to avoid liberal bias and present 'both sides' of contentious issues, including conversion therapy.
4. <https://gayexpress.co.nz/2025/04/meta-under-fire-for-ai-recommending-conversion-therapy/> - Details GLAAD's findings about Llama 4 recommending conversion therapy providers and Meta's acknowledgment of criticism.
5. <https://www.youtube.com/watch?v=Nlyd5oLzXOI> - Documents research on Meta AI's right-leaning bias and academic critiques of earlier Llama versions' authoritarian responses.
6. <https://www.lgbtqnation.com/2025/04/advocates-slam-metas-new-ai-for-recommending-conversion-therapy/> - Reiterates concerns about Meta's recent policy changes favoring conservative perspectives, including reduced content moderation.
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