

Tell MAMA works across the United Kingdom providing holistic and confidential support to Muslim communities impacted by anti-Muslim hatred, discrimination, and Islamophobia and in over a decade of work, publishes groundbreaking reports that seek to change public debate on anti-Muslim hate whilst working to encourage mutuality between communities.

We continue to warn about the prevalence of AI-generated falsehoods and the dangers they pose to Muslim and other minoritised communities, in our investigative reports and in subsequent annual reports.¹ In October 2025, we exposed a TikTok account which used generative software to depict violence and murder towards politicians, refugees and migrants, including a racist assault on London's Muslim Mayor Sir Sadiq Khan. As we stressed in the initial article and in subsequent research, it not only spoke to the "visual realisation of their desires for racist violence," but also audiences most receptive to it, expressed delight (or related emotions) amongst further coded calls for violence (or expressions of it). For example, 'One day this will come true,' gained hundreds of likes.² Due to the ease with which the average person can access (and indeed abuse such platforms to get around existing safeguards against violence), we have since uncovered various accounts pushing such content on platforms like TikTok that resurface on Meta-based platforms. The TikTok account we identified operated several similar accounts, producing a "dizzying" number of AI-generated videos that gained thousands of followers, until we passed our investigation on to TikTok. On Facebook, we revealed how a self-styled "comedy" page took this violent AI-generated content, notably depicting the shooting of refugees and migrants in the English Channel, with one such video gaining over one million views, which Meta declined to remove.

Subsequent Tell MAMA research identified how such violent, racialised AI-generated content appeared alongside the gamification of terrorism online, including content on TikTok replicating the horrors of the Christchurch terrorism, turning real-life murder and white supremacy into video game content.

More broadly, on platforms like Telegram, Tell MAMA found content promoting and encouraging terrorism, racist memes, and AI-generated materials, as well as content encouraging the vandalism of mosques. In a recent article on the dangers of AI-generated materials normalising racialised tropes and conspiracies about Muslims, refugees and migrant communities, we stressed that: "the extremism flourished [on Telegram], with terrorist documents uploaded and encouragement from those who operated the channel to encourage individuals within the UK to vandalise mosques, with or without financial incentives."³ Consequently, we published a wide-ranging report online on this incentivised form of far-right extremism and how its origins may pertain to an individual (or individuals) of Russian background.⁴ Moreover, we saw examples of the racist, violent and destructive materials, AI-generated (or otherwise) appear on Meta-based platforms, including Threads and Facebook. Regarding the latter, an individual from the East Midlands of England who included the text in an AI-generated image, "Should we dismantle the infrastructure of the caliphate? All of the following has been achieved by just 4 million Muslims out of a population of 66 million: Over 3,000 Muslim mosques. Over 130 Muslim Sharia courts. Over 50 Muslim Sharia councils."

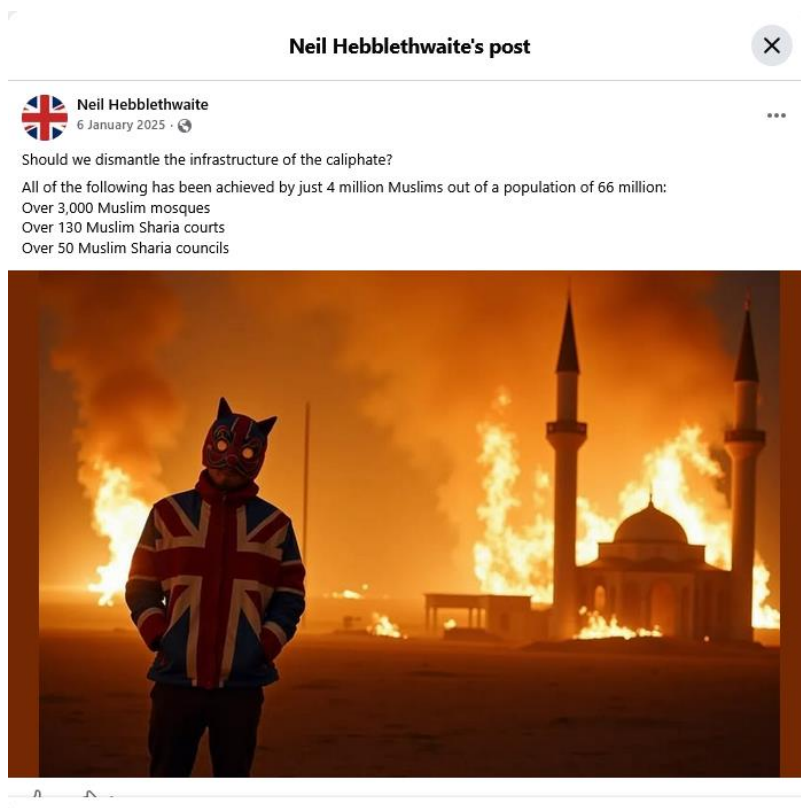
¹ Tell MAMA (2025). "The New Norm of anti-Muslim Hate". Available at: <https://tellmamauk.org/tell-mama-records-the-highest-number-of-anti-muslim-hate-cases-in-2024-since-its-founding/>

² Ibid.

³ Tell MAMA. "How AI-generated Images Fuel Anti-Muslim, Anti-migrant Sentiment Online." TELL MAMA. Last modified September 17, 2025. <https://tellmamauk.org/how-ai-generated-images-fuel-anti-muslim-anti-migrant-sentiment-online/>.

⁴ Tell MAMA (2025). "The Risk of Foreign Influence on the UK Far-Right and Anti-Muslim Hate." Available at: <https://tellmamauk.org/the-risk-of-foreign-influence-on-the-uk-far-right-and-anti-muslim-hate/>

That AI-generated image, however, proved highly inflammatory and risked inspiring criminality towards Muslim communities as it showed a self-styled “patriot” in Union Jack clothing standing over a mosque they had set alight. Whilst this example appeared to generate many interactions or reach when it appeared online on 6 January 2025, questions remain about Meta’s ability to autodetect hateful images; could the software not detect either the flames or the mosque? Or, does it give AI-generated content more leeway? Clarification is needed to help ensure that, moving forward, Meta-based platforms can detect such harmful content, even if the text of the image falls within the bounds of legitimate political expression and debate.



Scholarly work on the nature of dehumanisation and how it can encourage violence, as Rai, Valdesolo and Graham (2017) argued (within the bounds of a stricter definition of dehumanisation) that a failure to see the targets of dehumanisation as human fosters apathy to their suffering, when committing violence for instrumental purposes, as it would require those partaking feeling no guilt or empathy towards those target alongside feelings of hostility and a “moral impetus” to cause violence and harm.⁵

On the topic of incentivisation and reach, The Bureau for Investigative Journalism has published several wide-ranging articles, including those of individuals based abroad who, with the ease of AI-generated content, produce materials at unprecedented levels that, crucially, they argue, create “compelling images and videos.”⁶ The BBC has found similar concerns regarding foreign

⁵ Rai, Tage S., Piercarlo Valdesolo, and Jesse Graham. “Dehumanization increases instrumental violence, but not moral violence.” *Proceedings of the National Academy of Sciences* 114, no. 32 (2017): 8511-8516.

⁶ McIntyre, Niamh. “Racist AI-generated Videos on Facebook Are a Booming Industry.” TBIJ. Last modified May 19, 2026. <https://www.thebureauinvestigates.com/stories/2026-05-19/why-the-racist-ai-slop-industry-is-booming>.

accounts attempting to influence public opinion on immigration in the UK, highlighting the role of inauthentic and interconnected accounts that abuse Meta-based platforms.⁷

For all other misinformation, we focus on reducing its prevalence or creating an environment that fosters a productive dialogue. We know that people often use misinformation in harmless ways, such as to exaggerate a point ("This team has the worst record in the history of the sport!") or in humour or satire ("My husband just won Husband of the Year.") They also may share their experience through stories that contain inaccuracies. In some cases, people share deeply held personal opinions that others consider false or share information that they believe to be true, but others consider incomplete or misleading.

Due consideration must, therefore, fall on Meta's definition of satire, which gives this example: "We know that people often use misinformation in harmless ways, such as to exaggerate a point ("This team has the worst record in the history of the sport!") or in humour or satire ("My husband just won Husband of the Year.")"⁸ It is self-evident that such a statement is a joke, but there is a danger of conflating two very different things. Misinformation for *political* purposes serves a more *insidious, caustic* function when it pertains to content or news stories involving Muslim or other minoritised communities, especially when it uses emotional (or manipulative) language to generate such reactions from those of partisan views (McLoughlin and Brady, 2024).⁹ The average Meta user could discern that a flippant remark like "My husband just Husband of the Year" is *inherently* light-hearted and comedic. However, there is no inherent satirical value to falsified, misleading statements like "Refugees are welcome here, even if they rape our women, because white people do that too," more so when one of the individuals who raised concerns about the failure to deal with this harmful content referenced the safety of female political figures. We cannot stress enough how serious the rise in threats, abuse and intimidation of political figures has grown¹⁰ in recent years,¹¹ much of it has a gendered dynamic towards women, especially women political figures from minoritised backgrounds, demonstrating too painfully, the overlap of racism and misogyny, an issue Tell MAMA continues to document and highlight. A survey of elected officials on a local level last year revealed that "Almost three-quarters of respondents (73%) reported feeling personally at risk when fulfilling their role as councillor, which remains consistent with last year's [2024's] survey (74%). This was higher among women (84%), ethnic minorities (84%), the LGBT+ community (85%), and disabled respondents (85%)."¹² Abuse and harassment have forced some women to step away from politics,¹³ and it remains of great importance that Meta views this issue beyond the lens of electoral periods (even when abuse did also increase during the last General Election, it does not emerge in a vacuum). Furthermore, it is deeply unfair to elected officials to receive no protection from "unwanted manipulated imagery," whereas a private individual would, this

⁷ Spring, Marianna. "Anti-immigration AI Videos Traced to Overseas Fakers, BBC Finds." BBC News. Last modified May 15, 2026. <https://www.bbc.co.uk/news/articles/ckgpyn30dp3o>.

⁸ Meta. "Misinformation." Transparency Center. Accessed June 3, 2026. <https://transparency.meta.com/en-gb/policies/community-standards/misinformation/>.

⁹ McLoughlin, Killian L., and William J. Brady. "Human-algorithm interactions help explain the spread of misinformation." *Current opinion in psychology* 56 (2024): 101770.

¹⁰ Local Government Association. "The online abuse of elected women representatives, House of Commons, 20 April 2021." Local Government Association. Last modified April 16, 2021. <https://www.local.gov.uk/parliament/briefings-and-responses/online-abuse-elected-women-representatives-house-commons-20>.

¹¹ Al-Othman, Hannah. "Female UK Election Candidates Report Increased Abuse." The Guardian. Last modified June 14, 2024. <https://www.theguardian.com/politics/article/2024/jun/14/abuse-female-election-candidates-getting-worse-uk-targeted>.

¹² Brown, Thomas. "Local councillors: Impact of increasing abuse and intimidation." House of Lords Library. Last modified May 20, 2026. <https://lordslibrary.parliament.uk/local-councillors-impact-of-increasing-abuse-and-intimidation/>.

¹³ King's College London. "Women Politicians Face 'double Burden' of Online Toxicity." King's College London. Last modified January 22, 2025. <https://www.kcl.ac.uk/news/women-politicians-face-double-burden-of-online-toxicity>.

serves to only incentivise such acts of disinformation and misinformation, and could have tangible impacts on the safety of elected officials if such content is viewed by those who may act on the false belief of what the content misinforms them about. For example, in 2014, fake audio attributed to London Mayor Sir Sadiq Khan used AI software to say, "I control the Met Police, they will do as the Mayor of London tells them", and saying, "the British public need to get a grip". It said the prime minister meeting with Met Police Commissioner Sir Mark Rowley was "a waste of time" because "the buck stops with me".¹⁴ It resulted in more online abuse, and according to Khan himself, risked genuine public disorder offline.¹⁵

We call on the Oversight Board to push for revisions to guidelines that help the platforms better detect manipulated political content, especially speech on issues that are likely to inflame tensions and allow for political figures to better report content that poses a risk to them, alongside private individuals.

¹⁴ BBC News. "Sadiq Khan Says Fake AI Audio of Him Nearly Led to Serious Disorder." BBC News. Last modified February 13, 2024. <https://www.bbc.co.uk/news/uk-68146053>.

¹⁵ Ibid.