ANTISEMITISM POLICY TRUST

Response to Meta's Oversight Board Consultation Regarding Posts Supporting UK Riots

The Antisemitism Policy Trust is a UK-based charity that works to educate and empower parliamentarians and policy makers to address antisemitism. For more than two decades, the Trust has provided the secretariat to the All-Party Parliamentary Group (APPG) Against Antisemitism. We have published research papers on online antisemitism and have advised Government, Parliament and the UK regulator, Ofcom, on matters relating to online safety. We have also worked with technology companies and social media platforms, including meta, on reducing harms caused by online antisemitism.

We are submitting this advice based on our expertise in the fields of hate speech, disinformation and conspiracy theories, radicalisation, and incitement to violence against minorities. Specifically, our expertise relates to anti-Jewish racism, but much of this work also applies to many other ethnic and religious minorities.

Before we address the specific posts under consideration by the Oversight Board, there are some general themes that should be understood about the harms caused by disinformation and rhetoric that dehumanises groups and incites against them.

Demonising and dehumanising different groups, including Muslims, Jews, and immigrants, has profound and damaging effects. Throughout history, such rhetoric has fostered an environment of intolerance and hatred, leading to discrimination, social exclusion, and even violence against these groups. When individuals are portrayed as less than human, it becomes easier for others to justify mistreatment and abuse. This dehumanisation erodes our social fabric, undermining the principles of equality and respect that are essential for a cohesive society.

One of the most notorious examples is the dehumanisation of Jews by the Nazis, which was a central component of their genocidal campaign during the Holocaust. Nazi propaganda portrayed Jews as subhuman. This dehumanisation was achieved through various means, including depicting Jews as vermin, parasites, and a threat to society. This portrayal was designed to disengage moral concern and to justify the atrocities committed against Jews. It facilitated the view of Jews as 'the other,' and removed them from society, inculcating a permissive environment for anti-Jewish violence.

The dehumanisation of Jews throughout history, up to and including the present day, has led some to have become indifferent to – or sometimes even support – violence directed at Jews. After Hamas's massacre of 7 October for example, there has been shocking denial of, or indifference to, heinous crimes committed by terrorists, including the torture and rape of Jewish women, children and men. Years of stereotyping Jews as 'hyper-white,' privileged, and powerful, rather than an ethnic minority with a long history of oppression and persecution, and the spread of conspiracy theories and tropes that demonise Jews, may explain a real and perceived lack of empathy with Jewish people.

Much of the same language that has been used against Jews, is also being used against Muslims and migrant communities. This includes a depictions of them as vermin, dangerous and subhuman. On social media, and even in media outlets, these communities have been described as a threat to white women and children, as sexual predators, and as criminals in crass, gross and untrue stereotypes and generalisations. Both Muslim and immigrant communities have been described as 'swarming' or 'flooding' the West – like pests. This rhetoric drives up hatred against them, promotes harmful stereotypes, and contributes to discrimination and even violence against these groups.

Online content that spreads disinformation and perpetuates harmful stereotypes exacerbates the dehumanisation and demonisation of groups. Disinformation is often inflammatory and divisive, and as such, it can quickly go viral, reaching a wide audience and reinforcing negative biases. It not only misinforms the public, but also creates a climate of fear and mistrust. Communities targeted by such disinformation may experience heightened anxiety and a sense of vulnerability, while the broader society becomes more polarised and divided.

The risks to communities are significant. Hate speech and disinformation can incite violence, as seen in numerous incidents where online rhetoric has translated into physical attacks (Utoya, Finsbury Park Mosque, The Tree of Life Synagogue in Pittsburg, Buffalo and so on). Moreover, the spread of false information can undermine democratic processes by influencing public opinion and policy decisions based on lies and prejudice.

The rioting following the rapid spread of disinformation about Southport attacker is only one example of the correlation between hateful disinformation and dehumanising rhetoric, to violence. Another, more tragic, example is the 1994 genocide in Rwanda. The Hutu majority used a popular radio station to refer to Tutsi tribal members as "cockroaches," stripping away their humanity and making it easier to seek to justify the extermination of some 800,000 Tutsi.

Another example is the violence against Rohingya refugees in Myanmar. In 2017, a surge of online disinformation and inflammatory language on social media platforms, including Facebook, played a significant role in inciting violence against the Rohingya Muslim minority. False claims and hate speech spread rapidly, portraying the Rohingya as terrorists and a threat to national security. This online rhetoric fuelled widespread violence, leading to mass killings, sexual violence, and the displacement of hundreds of thousands of Rohingya people.

A more recent example is when, during the COVID-19 pandemic, misinformation and hate speech against Asians surged online, leading to over 11,000 incidents of physical violence and online aggression. This shows how harmful online rhetoric can translate into physical attacks, creating a climate of fear and hostility.

These examples highlight the dangerous impact of dehumanising language and online racism, emphasising the need for heightened vigilance and proactive measures to combat such harmful behavior, especially when an incident or event happen that may trigger violence, such as the attack in Southport or a war between Israel and Hamas.

To combat these risks online, it is crucial for social media platforms to act quickly and consistently to remove illegal content and material that violates terms and conditions. As a platform used by many millions across the world, Meta has a particular responsibility, but it also has the power to create a positive change. Posts containing illegal hate speech and disinformation that can incite

racial hatred and to violence, should be identified and removed quickly – and indeed this is a legal requirement now. There should also be safety by design mechanisms in place to ensure that posts containing content that is harmful but that the platform considers within the bounds of acceptability, are not actively promoted and do not become viral – specifically damaging and radicalising conspiracism. Directing users away from false and inflammatory information about minority groups and towards reliable sources of information is an imperative, and in line with Meta's stated approach to fighting disinformation about Covid-19 vaccines during the pandemic.

In the case presented by Meta's Oversight Board, it seems that the company has failed to protect the public by not removing posts quickly enough, and by allowing posts that encourage racism against Muslims to remain on the platform. We therefore welcome the Board's consultation on this. We would like to emphasise that our analysis of the posts in question is based solely on the description provided by the Board, as we have not seen these ourselves.

We agree that it was the right decision to remove the post that was taken offline. However, judging by your description of the other posts, we believe that those too should have been removed for promoting hatred and possible incitement to harm Muslim. The image in which a giant man wearing a union jack shirt is chasing Muslim men could be understood as encouraging white English people to target Muslim men because 'EnoughIsEnough' – alluding to a narrative that depicts Muslims as dangerous invaders that need to be driven out.

The AI-generated image in which four Muslim men are chasing a white toddler with a knife helps embolden a stereotype that all Muslims are terrorists, and plays into the disinformation about the Southport killer. We disagree with the argument that this image did not violate T&C because it referred to specific men. According to your description, the men in the image were 'generic' Muslims in traditional clothing, and not specific, known, individuals. Similarly, the white toddler represents a general white victim (specifically, a child) of Muslim aggression and terrorism. This was not in reference to a specific toddler whose identity is known. This was also not in relation to a particular incident that has occurred, but a claim about a general threat posed by the Muslim community and a call for people to 'wake up, which could be seen as a call for action against this community. The fact that the men are chasing after a toddler is meant to encourage a particularly strong emotional reaction, namely anger and fear, and to strengthen dangerous stereotypes.

Additionally, the two images were also AI-generated. Artificial Intelligence has added to the barrage of disinformation about the Southport killing. Even when these images are clearly AI-generated, they still have the ability to influence public opinion. Indeed, AI-generated images and deep fake videos are becoming increasingly realistic and can be used to manipulate public opinion on a large scale and in a very convincing way. We would like to use this opportunity to encourage Meta to adopt policies that will minimise the spread of disinformation by AI. One suggestion is adding transparency by labelling AI-generated content as such, so that users are aware that what they are seeing is manufactured.

At a time of civil unrest, there is a need for platforms to be particularly mindful about how quickly and effectively they act online in order to reduce to chances of violence and bloodshed offline. Race and religious-based violence has long term consequences for our community cohesion, security and democracy. It also has a profound effect on minorities' feeling of safety. As a charity that combats antisemitism, we have seen the effects of widespread online anti-Jewish hatred and disinformation since 7 October had on Jewish communities. These communities have been subjected to unprecedented levels of discrimination, abuse and violent attacks. Synagogues, Jewish schools and other Jewish institutions have been targeted by shooters and arsonists. Jewish children have suffered threats and violence. Not all of this is due to online antisemitism, but the prevalence of this content online is a major driver of offline antisemitism, and we stand by all minority groups who suffer similar racism.

Kind regards,
The Antisemitism Policy Trust