

The Institute for Strategic Dialogue's Recommendations and Comments on Oversight Board Cases regarding anti-immigrant hate speech

Thank you to the Oversight Board for the opportunity to comment on two cases regarding anti-immigrant hate speech.

The [Institute for Strategic Dialogue \(ISD\)](#) is an independent, non-profit organization dedicated to safeguarding human rights and reversing the rising tide of polarization, extremism, and disinformation worldwide. Our work includes in-depth research and analysis identifying and tracking online manipulation, mis- and disinformation, hate, and extremism in real time. We also formulate, advocate and deliver evidence-based policy approaches and programming.

Our submission seeks to address the Oversight Board's request for comments regarding to cases that fall under the Board's Hate Speech Against Marginalized Groups strategic priority:

1. Whether "murzyn" should be considered a slur term for the purposes of Meta's Hate Speech policy.

The term should be considered a slur for the purpose of Meta's Hate Speech policy. It is connected to negative stereotypes and expressions of inferiority regarding a protected group and thus violates Meta's guideline regarding hate speech. The term in question is rejected by members of the black community in Poland¹ which shows that it is seen by members of this group to be offensive and derogatory. Furthermore, the term in question is sometimes used in idioms, where it is meant to indicate backwardness and primitiveness.² A linguist study points out, that the "word actively reproduces anti-black stereotypes and racist meanings"³. The Council for the Polish Language at the Presidium of the Polish Academy of Sciences has recommended not to use the term given its pejorative meaning.⁴ While the status of the word might have shifted, and while not everybody using the term does so with the intent to offend, or even with the knowledge about its pejorative meaning, the term is regularly weaponized to insult black people in Poland and is used to indicate inferiority, as can be seen in its use in various idioms.

2. The socio-political context in Poland, Germany and Europe more broadly, in particular attitudes to immigration and how it is discussed politically.

Recent surveys from Germany point to a broad support for a tightening of laws granting asylum to refugees.⁵ In Germany and throughout Europe, far-right parties are both profiting from anti-migrant

¹ See e.g. <https://time.com/5874185/poland-racism-women-murzyn/>

² Some examples for such phrases can be found in the following article: Balogun, B., & Pędziwiatr, K. (2023). 'Stop calling me Murzyn' – how Black Lives Matter in Poland. *Journal of Ethnic and Migration Studies*, 49(6), 1552–1569. <https://doi.org/10.1080/1369183X.2022.2154914> (p. 1561-1562).

³ Ohia-Nowak, M. (2020). Słowo „Murzyn” jako perlokucyjny akt mowy. *Przegląd Kulturoznawczy*, 2020, Numer 3 (45), s. 195 – 212. <https://doi.org/10.4467/20843860PK.20.023.12583>

⁴ <https://www.rp.pl/spoleczenstwo/art246621-murzyn-oficjalnie-odradzany-przez-rade-jezyka-polskiego>

⁵ <https://de.statista.com/statistik/daten/studie/1477240/umfrage/umfrage-zur-verschaerfung-der-fluechtlingspolitik/>

sentiment and are fuelling anti-migrant attitudes.⁶ The negative attitudes against migrants also result in a rise of crimes against refugees. Crimes against refugees in Germany have almost doubled from 2022 (1,248 recorded crimes) to 2023 (2,378 crimes, among them 313 violent crimes), according to police statistics.⁷ In the first half year of 2024, 519 attacks on refugees and asylum seekers were recorded in Germany.⁸ In Germany, negative sentiments towards migrants predominate both the [policy debate, as well discourse on social media platforms](#). German respondents questioned about migration in a survey in 2023, among other things, voiced worries about problems in schools, additional burdens for the welfare state and the potential for conflicts with migrants.⁹ Aside from these concerns, anti-migrant sentiment is driven by [disinformation campaigns](#).

3. Any documented links between anti-immigrant speech and violence or discrimination in Poland, Germany and Europe more broadly.

Correlations between online hate speech and offline hate crimes are a well-documented phenomenon.¹⁰ An article published in October 2024, for example, shows, that Facebook posts containing inflammatory language were a predictor of hate crimes following shortly after.¹¹ The researchers do not claim that this implies a causal relationship. Several incidents in Europe do suggest that online platforms do play a role in coordinating and fuelling offline violence. In the wake of the 29 July stabbing in Southport, UK, Telegram and X [played a significant role in fuelling misinformation, disseminating riot tactics, and encouraging violence in the UK](#). Research by ISD [has shown](#), how online misinformation across platforms was used to fuel riots. Similarly, misinformation helped to [incite violence](#) in Dublin in November 2023.

4. How Meta's Hate Speech policy treats migrant status and whether it adequately protects both the rights of migrants and freedom of expression.

Meta's decision to only protect refugees, migrants, immigrants and asylum seekers from the most severe attacks under its Hate Speech policy is inadequate with regards to the company's human rights responsibilities. Refugees, migrants, immigrants and asylum seekers deserve the same protection against hate speech as other vulnerable groups. The problems that immigrants face as victims of hate crimes are reflected in immigration status being a protected characteristic in the hate crime legislation of California and Illinois. While the equal protection of the rights of refugees, migrants, immigrants and asylum seekers is an end in and of itself, expanding protections for these groups against hate speech is also necessary to effectively enforce Metas stated policy on hate speech regarding other

⁶ See e.g. <https://foreignpolicy.com/2023/11/01/the-far-right-is-winning-europes-immigration-debate/>

⁷ <https://www.tagesschau.de/inland/straftaten-gegen-gefluechtete-100.html>

⁸ <https://www.zdf.de/nachrichten/politik/deutschland/uebergrieffe-fluechtlinge-deutschland-100.html>

⁹ See <https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/willkommenskultur-in-krisenzeiten>

¹⁰ See e.g. Williams, M.L., Burnap, P. Javed, A., Liu, H., Ozalp, S. (2020). Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime. The British Journal of Criminology. 60 (1) p. 93 - 117 <https://doi.org/10.1093/bjc/azz049>; Jiménez D., Rafael and Müller, K and Schwarz, C. (2024): The Effect of Content Moderation on Online and Offline Hate: Evidence from Germany's NetzDG . Available at SSRN: <https://ssrn.com/abstract=4230296> or <http://dx.doi.org/10.2139/ssrn.4230296>

¹¹ Arcila Calderón, C., Sánchez Holgado, P., Gómez, J. et al. From online hate speech to offline hate crime: the role of inflammatory language in forecasting violence against migrant and LGBT communities. Humanit Soc Sci Commun 11, 1369 (2024). <https://doi.org/10.1057/s41599-024-03899-1>

groups. This is not only based on the fact that migrant status is sometimes used as a code for race or ethnicity (in one of the described cases, this is stated very openly). It also stems from the declared goals of Meta's Hate Speech policy, namely, to counter hate speech based on protected characteristics like race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity, and serious disease, and to prevent an environment of intimidation and exclusion as well as violence in the offline world. If refugees, migrants, immigrants and asylum seekers are not protected against all forms of attacks listed in Meta's Hate Speech policy, Meta creates a loophole for racist abuse and incitement to violence. Incitements to violence do not always use outright dehumanisation, they sometimes use rhetoric of disgust and inferiority, against which, under the current policy, refugees, migrants, immigrants and asylum seekers would not be protected. Therefore, we recommend extending the protection of people based on their migration status to ensure that they possess the same protection against hate speech as other vulnerable groups. Failing to do so would also mean a failure to reach the goal of ensuring freedom of expression, given that victims of online hate speech might increase feelings of insecurity for the victims,¹² and thus intimidate them from joining online discussions. We also recommend that Meta and the Oversight Board consistently enforce the platform's guidelines. As ISD [has pointed out](#) in the past, Meta should improve enforcement of guidelines regarding hate speech.

5. *Views on how Meta should distinguish commentary and criticism of immigration policies from direct attacks on people based on protected characteristics, such as race, especially during elections.*

Criticism of immigration policies that is based on specific issues, or policy challenges and problems associated with migration is different from attacks on migrants. The former discusses issues and challenges, the latter discusses people. The former might point to problems arising in connection with the presence of people, the latter rejects the presence of people themselves based on characteristics that should be protected characteristics. As soon as an alleged inherent trait of migrants is targeted (e.g. general suspicions of criminality or other claims of collective negative traits), the positions voiced are likely direct attacks on people rather than criticism of policies.

¹² Dreißigacker, A., Müller, P., Isenhardt, A. et al. Online hate speech victimization: consequences for victims' feelings of insecurity. *Crime Sci* 13, 4 (2024). <https://doi.org/10.1186/s40163-024-00204-y>