



OVERTURNED

2022-013-FB-UA



Iran protest slogan

The Oversight Board has overturned Meta's original decision to remove a Facebook post protesting the Iranian government, containing the slogan "marg bar Khamenei".

Policies and topics

-  Protests, Governments, Sex and gender equality
-  Violence and incitement

Region and countries

-  Middle East and North Africa
-  Iran

Platform

-  Facebook

Attachments

[Iran protest slogan public comments](#)

Case summary

The Oversight Board has overturned Meta's original decision to remove a Facebook post protesting the Iranian government, which contains the slogan "marg bar... Khamenei". This literally translates as "death to Khamenei" but is often used as political rhetoric to mean "down with Khamenei". The Board has made recommendations to better protect political speech in critical situations, such as that in Iran, where historic, widespread protests are being violently suppressed. This includes permitting the general use of "marg bar Khamenei" during protests in Iran.

About the case

In July 2022, a Facebook user posted in a group that describes itself as supporting freedom for Iran. The post contains a cartoon of Iran's Supreme Leader, Ayatollah Khamenei, in which his beard forms a fist grasping a chained, blindfolded woman wearing a hijab. A caption below in Farsi states "marg bar" the "anti-women Islamic government" and "marg bar" its "filthy leader Khamenei".

The literal translation of "marg bar", is "death to". However, it is also used rhetorically to mean "down with". The slogan "marg bar Khamenei" has been used frequently during protests in Iran over the past five years, including the 2022 protests. The content in this case was posted days before Iran's "National Day of Hijab and Chastity", around which critics frequently organise protests against the government, including against Iran's compulsory hijab laws. In September 2022, Jina Mahsa Amini died in police custody in Iran, following her arrest for "improper hijab". Her death sparked widespread protests which have been violently suppressed by the state. This situation was ongoing as the Board deliberated this case.

After the post was reported by a user, a moderator found that it violated Meta's Violence and Incitement Community Standard, removed it, and applied a "strike" and two "feature limits" to its author's account. The feature limits imposed restrictions on creating content and engaging with groups for seven and 30 days respectively. The post's author appealed to Meta, but the company's automated systems closed the case without review. They then appealed to the Board.

After the Board selected the case, Meta reviewed its decision. It maintained that the content

After the Board selected the case, Meta reviewed its decision. It maintained that the content violated the Violence and Incitement Community Standard but applied a newsworthiness allowance and restored the post. A newsworthiness allowance permits otherwise violating content if the public interest outweighs the risk of harm.

Key findings

The Board finds that removing the post does not align with Meta's Community Standards, its values or its human rights responsibilities.

The Board finds that this post did not violate the Violence and Incitement Community Standard, which prohibits threats that could lead to death or high-severity violence. Applying a newsworthiness allowance was therefore unnecessary. In the context of the post, and the broader social, political and linguistic situation in Iran, "marg bar Khamenei" should be understood as "down with". It is a rhetorical, political slogan, not a credible threat.

The Board emphasises the importance of context in assessing slogans calling for "death to", and finds that it is impossible to adopt a universal rule on their use. For example, "marg bar Salman Rushdie", cannot be equated to "marg bar Khamenei", given the fatwa against Rushdie and recent attempts on his life. Nor would "death to" statements used during events such as the 6 January riots in Washington D.C. be comparable, as politicians were clearly at risk and "death to" statements are not generally used as political rhetoric in English, as they are in other languages.

The centrality of language and context should be reflected in Meta's policies and guidance for moderators. This is particularly important when assessing threats to heads of state, who are legitimately subject to criticism and opposition.

In the Iranian context, the Board finds that Meta must do more to respect freedom of expression, and permit the use of rhetorical threats. The Iranian government systematically represses freedom of expression and digital spaces have become a key forum for dissent. In such situations, it is vital that Meta supports users' voice. Given the "National Day of Hijab and Chastity" was approaching, Meta should have anticipated issues around the over-removal of Iranian protest content and prepared an adequate response. For example, by instructing "at-scale" reviewers not to remove content containing the "marg bar Khamenei" slogan.

As this case shows, its failure to do so led to the silencing of political speech aimed at protecting women's rights, including through feature limits, which can shut people out of social movements and political debate. Public comments submitted to the Board indicate that "marg bar Khamenei" has been used widely during the recent protests in Iran. This is supported by independent research commissioned by the Board. Many of these posts would have been removed without benefitting from the newsworthiness allowance, which Meta rarely applies (in the year to June 2022 it was used just 68 times globally).

The Board is concerned that Meta is automatically closing appeals and that the system it uses to do so fails to identify important cases. It recommends that the company takes action to improve its respect for freedom of expression during protests and in other critical political contexts.

The Oversight Board's decision

The Oversight Board overturns Meta's original decision to remove the post.

The Board also recommends that Meta:

- Amend the Violence and Incitement Community Standard so that it more accurately reflects its policies. This should include providing the criteria used to determine when rhetorical threats against heads of state are permitted. These criteria should protect clearly rhetorical political speech, used in protest contexts, that does not incite violence, and should take language and context into account.
- Pending changes to the Violence and Incitement Community Standard, issue guidance to its reviewers that, in the context of protests in Iran, "marg bar Khamenei" statements do not violate the policy.
- Err on the side of issuing scaled allowances when potentially violating content is used during protests, where this is in the public interest and is unlikely to lead to violence.
- Revise the indicators it uses to rank appeals for review and to automatically close appeals without review to help identify public interest expression, particularly that related to protest.
- Announce all "scaled" allowances, their duration and notice of their expiry.
- Explain the "newsworthiness allowance" in more detail in its transparency centre, including the criteria used to decide whether to "scale" an allowance.

- Publicly explain its process for automatically prioritising and closing appeals, including the criteria that it uses to do so.

* Case summaries provide an overview of the case and do not have precedential value.

Full case decision

1. Decision summary

The Oversight Board overturns Meta's original decision to remove a Facebook post protesting the Iranian government's human rights record and its laws on mandatory hijab (head covering). The post contains a caricature of the country's Supreme Leader, Ayatollah Ali Khamenei and the phrase, "marg bar [...] Khamenei", a protest chant that literally means "death to ... Khamenei", but has frequently been used as a form of political expression in Iran which can also be understood as "down with [...] Khamenei". The Board found that the content did not violate the Violence and Incitement policy. Nationwide protests in Iran, triggered by the killing of Jina Mahsa Amini were being violently suppressed by the Iranian government at the time of the Board's deliberation.

Meta reversed its decision after it was informed that the Board had selected this case. The company maintained that the content violated the Violence and Incitement Community Standard, but restored the content using the "newsworthiness allowance".

The case raises important concerns about Meta's Violence and Incitement policy and its "newsworthiness allowance". It also raises concerns about how Meta's policies may impact freedom of expression and women's rights in Iran and elsewhere. The Board finds that Meta did not meet its human rights responsibilities in this case, in particular to prevent errors adversely impacting freedom of expression in protest contexts. The Board recommends that Meta review its Violence and Incitement Community Standard, its internal implementation guidelines for moderators, and its approach to newsworthy content, in order to respect freedom of expression in the context of protests.

2. Case description and background

In mid-July 2022, a person posted in a public Facebook group that describes itself as supporting freedom for Iran, criticising the Iranian government and Iran's Supreme Leader,

Ayatollah Khamenei, particularly their treatment of women, including Iran's strict compulsory hijab laws. The post was made days before the "National Day of Hijab and Chastity" in Iran. The government intends this day to be a celebration of mandatory hijab, but critics have used it to protest against mandatory hijab and broader government abuses in Iran, including online.

The post contains a cartoon of Ayatollah Khamenei, in which his beard forms a fist grasping a woman wearing a hijab. The woman is blindfolded with a chain around her ankles. A text bubble next to the caricature says that being a woman is forbidden. A caption below in Farsi reads, "marg bar hukumat-e zed-e zan-e eslami va rahbar-e kasifesh Khamenei". The term "marg bar" translates literally as "death to". The caption literally calls for "death to" the "anti-women Islamic government" and its "filthy leader Khamenei". However, in some contexts, "marg bar" is understood to have a more rhetorical meaning equivalent to "down with". The post also calls the Islamic Republic "the worst dictatorship in history", in part due to restrictions on what women can wear. It also calls on women in Iran not to collaborate in the oppression of women.

On the day that the content was posted, another person on Facebook reported it as hate speech. One of Meta's at-scale reviewers assessed the post as violating the Violence and Incitement Community Standard, which prohibits threats that could lead to death or high-severity violence against others. Meta removed the content, resulting in the author of the post receiving a strike, which led to the automatic imposition of 30-day and seven-day account restrictions known as "feature limits". While feature limits vary in nature and duration, they can generally be understood as punitive and preventative measures denying individuals the regular use of the platform to express themselves. The 30-day feature limit prevented the content's author from posting or commenting in groups, inviting new members to groups, or creating new groups. The seven-day feature limit prevented them from creating any new content on any Facebook surface, excluding the Messenger app. When their content was removed, the author was informed of the seven-day feature limit through notifications, but did not receive notifications about the 30-day group-related feature limit.

Hours after Meta removed the content, the author of the post appealed the decision. Meta's automated systems did not prioritise the appeal and it was later closed without being reviewed. The user received a notification that their appeal was not reviewed because of a temporary reduction in review capacity as a result of COVID-19. At this point, they appealed Meta's removal decision to the Oversight Board.

After it was informed that the Board had selected this case, Meta determined that its previous decision to remove the content was incorrect. It found that, although the post violated the Violence and Incitement Community Standard, it would restore the content under the newsworthiness allowance. This permits content that would otherwise violate Meta's policies if the public interest in the content outweighs the risk of harm. The content was restored in August, more than a month after it was first posted, but after the "National Day of Hijab and Chastity" had already passed. Meta reversed the strike against the person's account, but the account restrictions that had been imposed could not be reversed, as they had already run their full duration.

In September, the Iranian government's morality police arrested 22-year old Jina Mahsa Amini for wearing an "improper" hijab. Amini fell into a coma shortly after collapsing at the detention centre and died three days later, while in custody. Her death at the hands of the state sparked widespread peaceful protests, which were met with extreme violence from the Iranian government. This situation was ongoing at the time the Board deliberated this case.

The United Nations has raised concerns about Iranian security forces using illegitimate force against peaceful protesters, killing and injuring many, including children, as well as arbitrarily detaining protesters and imposing Internet shutdowns. The United Nations has reiterated calls for the release of detained protesters, and the UN Human Rights Council convened a Special Session on 24 November to address the situation. The resolution adopted at that session (A/HRC/Res/S-35/1) expressed "deep concern" about "reports of restrictions on communications [...] including Internet shutdowns and blocking of social media platforms, which undermine the exercise of human rights". It called on the Iranian Government to end all forms of discrimination and violence against women and girls in public and private life, to uphold freedom of expression and to fully restore Internet access. The UN Human Rights Council also established an independent international fact-finding mission to investigate alleged human rights violations in Iran related to the protests that began on 16 September.

Public comments and experts the Board consulted confirmed that the "marg bar Khamanei" slogan was being widely used in these protests and online, and that it had been commonly used in protests in Iran in 2017, 2018, 2019 and 2021. Public comments often included perceptions that Meta over-enforces its policies against Farsi language content during protests, including in the most recent protests that have mostly been led by women and girls. These perceptions are also reflected in Memetica's research on platform data that the Board commissioned, which found that from 1 July to 31 October, 400 public Facebook posts and

commissioned, which found that from 1 July to 31 October, 400 public Facebook posts and 1,046 public Instagram posts used the hashtag #MetaBlocksIranProtests.

People in Iran have been protesting for gender equality and against compulsory hijab since at least the 1979 revolution. The Islamic Penal Code of Iran penalises women who appear in public without a "proper hijab" with imprisonment or fine. Women in Iran are banned from certain fields of study, many public places, and from singing and dancing, among other things. Men are considered the head of the household and women need the permission of their father or husband to work, marry or travel. A woman's court testimony is considered to have half the weight of a man's, limiting access to justice for women.

The Iranian government systematically represses freedom of expression. While online spaces have become a key forum for dissent, the government has taken extreme measures to silence debate there too. Human rights advocacy is a common target, in particular women's rights advocacy, political dissent, artistic expression and those calling for the government to be held to account for its human rights violations. Facebook, Twitter and Telegram have all been banned in Iran since 2009. The Iranian government also blocked access to Instagram and WhatsApp in September 2022 amid protests over Amini's death. The Open Observatory of Network Interference documented new forms of censorship and Internet shutdowns in various parts of Iran during the protests. Usage of virtual private networks (VPN) (tools that encrypt communications and can be used to circumvent censorship) reportedly increased more than 2000% during September 2022. Public comments that the Board received emphasised that social media platforms are one of the only tools for people to freely express themselves, given the Iranian state's tight control of traditional media. The state's advanced capabilities to restrict online expression make the lifeline of social media particularly precarious. Social media plays a crucial role in ensuring that people in Iran can exercise their rights, particularly in times of protest.

3. Oversight Board authority and scope

The Board has authority to review Meta's decision following an appeal from the person whose content was removed (Charter Article 2, Section 1; Bylaws Article 3, Section 1). The Board may uphold or overturn Meta's decision (Charter Article 3, Section 5), and this decision is binding on the company (Charter Article 4). Meta must also assess the feasibility of applying its decision in respect of identical content with parallel context (Charter Article 4). The Board's decisions may include policy advisory statements with non-binding recommendations that Meta must respond to (Charter Article 3, Section 4; Article 4). Where

Meta commits to act on recommendations, the Board monitors their implementation.

When the Board selects cases like this one, where Meta subsequently revises its initial decision, the Board focuses its review on the decision that is appealed to it. In this case, while Meta recognised the outcome of its initial decision was incorrect and reversed it, the Board notes that this reversal relied on the newsworthiness allowance, which is among the enforcement options that are only available to Meta's internal policy teams and not to content moderators working at scale. The case was not an "enforcement error", as the scaled content reviewer removed the content in accordance with the internal guidance they were given, though as noted below, these differ from the public-facing Community Standards in ways that are material to this case.

4. Sources of authority and guidance

The following standards and precedents informed the Board's analysis in this case:

I. Oversight Board decisions

The most relevant previous decisions of the Oversight Board include:

- "Russian poem" case ([2022-008-FB-UA](#)): In this case, the Board recommended that the details from internal enforcement guidelines should be reflected in the public-facing rules in the Community Standards.
- "Mention of the Taliban in news reporting" case ([2022-005-FB-UA](#)): In this case, the Board recommended that Meta release more information on its strikes system and addressed concerns about the effects of "feature limits" on people during times of crisis.
- "Wampum belt" case ([2021-012-FB-UA](#)): In this case, the Board emphasised the importance of examining the whole content for contextual cues, and not removing posts based on a decontextualised phrase in isolation.
- "Colombia protests" case ([2021-010-FB-UA](#)). In this case, which concerned anti-government protests, the Board reiterated that Meta notify users when their content is restored under the newsworthiness allowance and further recommended that Meta develop and publicise clear escalation criteria and a process for reviewing potentially newsworthy content.

II. Meta's Content Policies

The policy rationale for the Violence and Incitement Community Standard explains that it intends to "prevent potential offline harm that may be related to content on Facebook" while acknowledging that "people commonly express disdain or disagreement by threatening or calling for violence in non-serious ways". Under this policy, Meta does not allow "threats that could lead to death (and other forms of high-severity violence)", where "threats" are defined as including "calls for high-severity violence" and "statements advocating for high-severity violence".

The Board's analysis of the content policies was informed by Meta's commitment to voice, which the company describes as "paramount", and its values of safety and dignity. In explaining its commitment to voice, Meta explains that "in some cases, we allow content – which would otherwise go against our standards – if it's newsworthy and in the public interest". This is known as the newsworthiness allowance, which is linked from Meta's commitment to voice. The newsworthiness allowance is a general policy exception applicable to all Community Standards. To issue the allowance, Meta conducts a balancing test, assessing the public interest in the content against the risk of harm. Meta says that it assesses whether content "surfaces an imminent threat to public health or safety, or gives voice to perspectives currently being debated as part of a political process". Both the assessment of public interest and harm take into account country circumstances such as whether an election or conflict is under way, whether there is a free press and whether Meta's products are banned. Meta states that there is no presumption that content is inherently in the public interest solely on the basis of the speaker's identity, for example their identity as a politician. Meta says that it removes content "even if it has some degree of newsworthiness, when leaving it up presents a risk of harm, such as physical, emotional and financial harm, or a direct threat to public safety".

III. Meta's human rights responsibilities

The UN Guiding Principles on Business and Human Rights (UNGPs), endorsed by the UN Human Rights Council in 2011, establish a voluntary framework for the human rights responsibilities of private businesses. In 2021, Meta announced its Corporate Human Rights policy, where it reaffirmed its commitment to respecting human rights in accordance with the UNGPs. The Board's analysis of Meta's human rights responsibilities in this case was informed by the following international standards:

- Freedom of expression: Article 19. International Covenant on Civil and Political

Rights ([ICCPR](#)); [General Comment No. 34](#), Human Rights Committee, 2011; UN Human Rights Council resolution [A/HRC/Res/23/2](#) on freedom of expression and women's empowerment (2013), and [A/HRC/Res/S-35/1](#) (2022) on the deteriorating situation of human rights in Iran; UN Special Rapporteur on freedom of opinion and expression, reports: [A/76/258](#) (2021), [A/74/486](#) (2019), [A/HRC/38/35](#) (2018) and [A/HRC/44/49/Add.2](#) (2020).

- [The prohibition on incitement](#): Article 20, para. 2, ICCPR; Rabat Plan of Action, UN High Commissioner for Human Rights report: [A/HRC/22/17/Add. 4](#) (2013).
- [Equality and non-discrimination based on sex and gender for participation in political and public life](#): Article 2, para. 1, Articles 25 and 26, ICCPR; Articles 2, 7 and 15, Convention on the Elimination of All Forms of Discrimination Against Women ([CEDAW](#)); [General Recommendation No. 23](#) on political and public life, CEDAW Committee, 1997.
- [Freedom of peaceful assembly](#): Article 21, ICCPR; [General Comment No. 37](#), Human Rights Committee, 2020.
- [Freedom of religion or belief](#): Article 18, ICCPR; UN Special Rapporteur on freedom of religion or belief, report: [A/68/290](#), 2013.
- [Right to security of person](#): Article 9, ICCPR.
- [Right to life](#): Article 6, ICCPR.
- [Right to a remedy](#): Article 2, para. 3, ICCPR.

5. User submissions

In their appeal to the Board, the person who authored the post shared that they intended to raise awareness of how people in Iran are being "abused" by the Iranian "dictatorship" and that people "need to know about this abuse". For them, the "Facebook decision is unfair and against human rights".

6. Meta's submissions

Meta explained that assessing whether the phrase "death to" a head of state constitutes rhetorical speech as opposed to a credible threat is challenging, particularly at scale. Meta said that there has been much internal and external debate on this point and welcomed the Board's input on where to draw the line. Meta also said that it would welcome guidance on drafting a policy it can apply at scale.

Meta explained to the Board that the phrase "death to Khamenei" violated the Violence and

Incitement Community Standard, and this was the reason why the content was initially removed. The Community Standards have been available in Farsi since February 2022.

According to Meta, the policy prohibits "calls for death targeting a head of state". It currently distinguishes calls for lethal violence where the speaker expresses intent to act (e.g. "I am going to kill X"), which is violating, from content expressing a wish or hope that someone dies without expressing intent to act (e.g. "I hope X dies" or "death to X"). The latter is generally non-violating, because Meta considers that the word "death" "is not itself a method of violence". Meta generally considers this to be "hyperbolic language", where the speaker does not intend to incite violence.

However, internal guidance instructs moderators to remove "death to" statements where the target is a "high-risk person". The guidance is called "known questions" and includes a confidential list of categories of person (rather than named individuals) who Meta considers high risk. Essentially, Meta's removals at scale are formulaic: the combination of ["death to"] plus [target is a high-risk person] will result in removal, even if other context indicates that the expression is hyperbolic or rhetorical, and therefore similar to speech permitted against other targets. "Heads of state" are listed as high-risk persons, and Meta explained that this is because of "the potential safety risk" against them.

Meta further said that it has developed an evolving list of high-risk persons based on feedback from its policy teams, as well as external experts. Meta provided the full list to the Board. In addition to heads of state, other examples of "high risk" persons include: former heads of state; candidates and former candidates for head of state; candidates in national and supranational elections for up to 30 days after election if not elected; people with a history of assassination attempts; activists and journalists.

After it was informed that the Board selected the case, Meta revisited its decision and decided to restore the post under the "newsworthiness allowance". While Meta maintained the content violated its policies, restoring it was the right thing to do because "the public interest value outweighed any risk of contributing to offline harm". Meta has previously informed the Board that the kind of contextual analysis its policy teams can conduct to reach decisions on-escalation is not available to moderators at scale, who must follow internal guidance.

In this case, Meta determined that the public interest was high, as the post related to public

discourse on compulsory hijab laws and criticised the government's treatment of women. Meta found the cartoon to be political in nature, and given the religious significance of beards to some who practice Islam, that its imagery could be criticism of the use of religion to control and oppress women. The political context and timing of the post were important, in the run-up to the mid-July "National Day of Hijab and Chastity", when Meta understood that many people were using social media hashtags to organise protests. Meta cited the Board's "Colombia protests" case in support of its public interest assessment, and pointed to the Iranian government's history of suppressing freedom of expression and Internet shutdowns.

Meta determined that the public interest outweighed the risk of offline harm, which was low. It was clear to Meta that the author of the content did not intend to call for violent action against Ayatollah Khamenei, but rather to criticise the government's "anti-women" policies. In this situation, Meta gave more weight to the rhetorical meaning of "marg bar" as "down with", noting its frequent use as a form of political expression in Iran. Restoring the content was, for Meta, consistent with its values of voice and safety.

Meta explained to the Board that it has two categories of newsworthy allowances: "narrow" and "scaled". In this case, Meta applied a narrow allowance, which only restores the individual piece of content, and has no effect on other content, even if it is identical. A "scaled" allowance, by contrast, applies to all uses of a phrase that would otherwise violate policy, regardless of the identity of the speaker. Scaled allowances are normally limited in duration. Both types of allowances can only be issued by Meta's internal policy teams; a content moderator reviewing posts at scale cannot issue such allowances, but they do have options for escalating content to Meta's internal teams.

Meta explained that it has three times granted scaled newsworthiness allowances for the "death to Khamenei" phrase, first in connection with the 2019 fuel price protests in Iran, second in the context of the 2021 Iranian election, and third, related to the 2021 water shortage protests. However, no scaled allowance has been issued to allow these statements since the beginning of the protests against compulsory hijabs in 2022.

Meta disclosed to the Board that it has become more hesitant to grant "scaled" allowances and favours considering "narrow" allowances on a case-by-case basis. This is due to "public criticism" of Meta for temporarily allowing "death to" statements in a prior crisis situation. In response to the Board's questions, Meta clarified that it does not publish the newsworthiness allowances that it issues. Meta also clarified the number of times that it issued scaled

newsworthiness allowances globally for content that would otherwise violate the Violence and Incitement policy in the 12 months up to 5 October 2022, but requested that this data be kept confidential as it could not be validated for release in the time available.

The Board asked Meta how much content would have been impacted if the company had issued a scaled newsworthiness allowance to permit "death to Khamenei" statements. Meta said that this cannot accurately be determined without assessing each post for other violations. While Meta provided the Board with data on the usage of "death to Khamenei" hashtags between mid-July and early October 2022, it requested that data be kept confidential as it could not be validated for release in the time available.

While Meta did not issue a scaled newsworthiness allowance for "marg bar Khamenei" statements, the company disclosed that on 23 September 2022, ten days after the killing of Jina Mahsa Amini, it issued a "spirit of the policy" allowance for the phrase "I will kill whoever kills my sister/brother". This scaled allowance was still in effect when the Board was deliberating the present case.

In this case, the author of the post received a "strike" as a result of their content being assessed as violating. Meta disclosed that in May 2022, it issued guidance that "marg bar Khamenei" slogans should be removed for violating the Violence and Incitement Community Standard, but should not result in a strike. Meta explained that this was intended to mitigate the impact of removing content with some public interest value, though not enough to outweigh the risk of harm and warrant a newsworthiness allowance. It was still in effect as the Board finalised its decision. Meta explained that this guidance is distinct from a newsworthiness allowance, as it does not affect the decision to remove the content and only applies to the penalty imposed. In response to the Board's questions, Meta explained that the author of the content in this case did not benefit from this penalty exemption because it is only available for content decisions made by internal teams "at escalation". As the post in this case was assessed as violating by a content moderator at scale, a strike was automatically issued, and, taking into account the accrual of prior strikes, corresponding "feature limits" were imposed. In response to the Board's questions, Meta disclosed that the user was notified about the seven-day feature limit but not the 30-day group-related feature limit. As such, the user would only find out about the 30-day group-related feature limit if they were to access the status section of their account or if they attempted to perform a restricted action related to a group. In response to prior Oversight Board recommendations, Meta has [provided more information publicly](#) on the operation of its strikes system and resulting account penalties. During the finalisation of this decision, Meta also informed the Board that

account penalties. During the finalisation of this decision, Meta also informed the Board that, in response to recommendations in the "Mention of the Taliban in news reporting" case, it would increase the strike threshold for the imposition of "read-only" penalties and update its transparency centre with this information.

Meta explained that when the author of the content appealed the initial removal decision, that appeal did not meet prioritisation criteria and was automatically closed without review. In response to the Board's questions, Meta provided further explanation of its review capacities for Farsi content. Meta explained that the Community Standards have been available on its website in Farsi since February 2022. Meta shared that for "higher risk" markets, such as the Persian market, which are characterised for example by recurring volume spikes due to real-world events, or markets with "long lead times required to increase capacity", it over-allocates content moderation resources so that it can deal with any crisis situations that arise.

Meta cited the human rights to freedom of expression (Article 19, ICCPR), freedom of assembly (Article 21, ICCPR) and the right to participate in public affairs (Article 25, ICCPR) in support of its revised decision. At the same time, Meta acknowledged that it needs "bright-line" rules to accomplish, at scale, the legitimate aim of its Violence and Incitement policy of protecting the rights of others from threats of violence. Meta told the Board that the "application of these bright-line rules sometimes results in removal of speech that, on escalation, we may conclude (as we did in this case) does not contain a credible threat". Meta explained that it continuously monitors trends on its platforms to protect political speech that might otherwise violate policies, and in the past has made more use of scaled allowances. The company invited the Board's input as to when it should grant these types of allowances and the criteria that it should consider when doing so.

The Board asked Meta 29 questions in writing. Questions related to: the criteria and process for issuing at-scale policy exceptions; automatic closure of appeals; measures taken by Meta to protect user's rights during protests; the company's content review capacity in countries where its products are banned; and alternative processes or criteria that Meta has considered to effectively permit rhetorical non-threatening political expression at scale. 26 questions were answered fully and three were answered partially. The partial responses were to questions on: data comparing auto-closure of appeals for content in Farsi and English languages; the prevalence of several variations of the "death to Khamenei" slogan on Meta's platforms, and the accuracy rates on the enforcement of the Violence and Incitement policy in Farsi.

7. Public comments

The Oversight Board received 162 public comments related to this case. 13 comments were submitted from Asia Pacific and Oceania, six from Central and South Asia, 42 from Europe and 36 from the Middle East and North Africa. 65 comments were submitted from the United States and Canada. The submissions covered the following themes: the distinction between political rhetoric and incitement; the importance of context, in particular for language and imagery, when moderating content; the limitations of the newsworthiness allowance in dealing with human rights violations and overreliance on automated decision-making; and freedom of expression, human rights, women's rights, government repression and social media bans in Iran.

To read public comments submitted for this case, please click [here](#).

8. Oversight Board analysis

The Board examined whether this content should be restored by analysing Meta's content policies, human rights responsibilities and values.

The Board selected this case because it offered the potential to explore how Meta assesses criticisms of government authorities, and whether heads of state in certain countries receive special protection or treatment, as well as important matters around advocacy for women's rights and the participation of women in public life. Additionally, this content raises issues around criticism of political figures through rhetorical speech that may also be interpreted as threatening and the use of the newsworthiness allowance. The case provides the Board with the opportunity to discuss Meta's internal procedures, which determine when and why policy exceptions should be granted, as well as how policies and their exceptions should be applied. The case primarily falls into the Board's elections and civic space priority, but also touches on the Board's priorities of gender, Meta and governments, crisis and conflict situations, and treating users fairly.

8.1 Compliance with Meta's Content Policies

I. Content rules

a. Violence and incitement

The Board finds that the content in this case does not violate the Violence and Incitement Community Standard. Therefore, it was not necessary for Meta to apply the newsworthiness allowance to the post.

This conclusion is supported by the analysis that Meta conducted when it revisited its decision after the Board selected the case. The policy rationale explains that Meta intends to "prevent potential offline harm" and that it "removes language that incites or facilitates serious violence". Under the heading "do not post", the Community Standard prohibits threats "that could lead to death or high-severity violence". However, internal guidance indicates that, generally, "death to" statements against any targets, including named individuals, are permitted on Meta's platforms and do not constitute a violation of this rule, except when the target of the "death to" statement is a "high-risk person". The internal guidance instructs reviewers to treat "death to" statements against "high-risk persons" as violating regardless of other contextual cues. Further, the public-facing "do not post" rules in the Violence and Incitement Community Standard do not reflect this internal guidance and accordingly do not expressly prohibit "death to" statements targeting high-risk individuals, including heads of state.

Meta's analysis of the content found that it presented a low risk of offline harm; that it did not intend to call for Ayatollah Khamenei's death; and that the "death to Khamenei" slogan has frequently been used as a form of political expression in Iran which is better understood as "down with Khamenei". This should have been sufficient for Meta to find the content non-violating and allow the post and other similar content to remain on its platform. The Board is concerned that Meta has not taken action to allow use of "marg bar Khamenei" at scale during the current protests in Iran, despite its assessment in this case that the slogan did not pose a risk of harm.

Linguistic experts consulted by the Board confirmed that the "marg bar Khamenei" slogan is commonly used in Iran, in particular during protests, as a criticism of the political regime and Iran's Supreme Leader, rather than as a threat to Ayatollah Khamenei's safety. The post preceded by several days the "National Day of Hijab and Chastity", during which Meta noted an increase in use of social media in Iran to organise protests.

In this context, the slogan should have been interpreted as a rhetorical expression, meaning "down with" Khamenei and the Iranian government. It did not therefore fall within the rule on "threats that could lead to death" and it did not advocate or intend to cause high-severity

threats that could lead to death , and it did not advocate or intend to cause high severity violence against the target. The Board notes that "down with" statements against a target are permissible under Meta's policies, regardless of the target's identity. This is consistent with Meta's commitment to voice and the importance of protecting political discontent. There is no "genuine risk of physical harm" or "direct threats to public safety", which the policy aims to avoid. Rather, the content falls squarely in the category of statements through which people "commonly express disdain or disagreement by threatening or calling for violence in non-serious ways".

Meta's internal guidance for moderators contains a presumption in favour of removing "death to" statements directed at "high-risk persons". This would apply to Ayatollah Khamenei, as a head of state. This rule, while not public, is consistent with the policy rationale for content that places these persons at heightened risk. However, its enforcement in this case is not. The policy rationale of the Violence and Incitement Community Standard states that the language and context of a particular statement ought to be considered in determining whether a threat is credible. This did not occur in the present case, as the stated presumption was applied regardless of language and context, though the Board notes that the reviewer acted in a manner consistent with the internal guidance. As Meta later acknowledged, various elements of the post, and the broader context in which it was posted, make clear that it was not making a credible threat but employing political rhetoric. The Board's decision in the "Wampum belt" case is relevant here. In that decision, the Board held that a seemingly violating phrase "kill the Indian", should not be read in isolation but in the context of the full post, which made clear that it was not threatening violence but opposing it. Similarly, in the "Russian poem" case, the Board found that various excerpts of a poem (e.g. "kill a fascist") were not violating, as the post was using rhetorical speech to call attention to a cycle of violence, not urging violence.

With the internal guidance drafted as it is, the Board understands why the content moderator made the decision that they did in this instance. However, Meta should update this guidance so that it is more consistent with the stated policy rationale. The Board agrees that "death to" or similar threatening statements directed at high-risk persons should be removed due to the potential risk to their safety. Though the Board also agrees that heads of state may be considered high-risk persons, this presumption should be nuanced in Meta's internal guidance. For these reasons, the Board also finds that removing the content was not consistent with Meta's commitment to voice and was not necessary to advance safety. Meta should have issued scaled guidance that instructed moderators not to remove this protest slogan by default, and, accordingly, not removed this post during at-scale review.

b. Newsworthiness allowance

It follows from the Board's assessment that the Violence and Incitement Community Standard was not violated, that the newsworthiness allowance was not required. Notwithstanding this conclusion, the Board finds that, when Meta chose to apply a newsworthiness allowance to this post, it should have been scaled to apply to all "marg bar Khamenei" slogans, regardless of the speaker. This was done in response to several similar previous widespread protests in Iran. In the Board's view, those actions were more consistent with Meta's commitment to voice than the action that Meta took when revisiting its decision in this case. Many other people are in the same situation as the user in this case, so the allowance should not have been limited to an individual post. Scaling the decision was necessary given the importance of social media to protest in Iran, the human rights situation in the country, and the fact that Meta should reasonably have anticipated that the same issue would recur many times. The failure to apply a scaled allowance has had the effect of silencing political speech aimed at protecting women's rights, by removing what the Board has concluded was non-violating speech.

Criticisms of Meta's use of allowances to permit otherwise violating "death to" statements in relation to Russia's invasion of Ukraine should not, in the Board's view, have led to Meta reducing the use of scaled allowances in protest contexts. The situation in Iran concerns a government violating the human rights of its own citizens, while repressing protests and severely limiting the free flow of information. There would have been many thousands of usages of the "marg bar Khamenei" slogan in recent months on Meta's platforms. Very few posts, if any besides the content in this case, would have benefited from a newsworthiness allowance. While the newsworthiness allowance is often framed as a mechanism for safeguarding public interest speech, its use is relatively rare considering the global scope of Meta's operations. According to Meta, only 68 newsworthiness allowances were issued across all policies globally between 1 June 2021 to 1 June 2022.

In the Board's view, in contexts of widespread protests, Meta should be less reluctant to scale allowances. This would help to protect voice where there are minimal risks to safety. This is particularly important where systematic violations of human rights have been documented, and avenues for exercising the right to freedom of expression are limited, as in Iran.

II. Enforcement action

The impact of "feature limits" on individuals during times of protest is especially grave. Such limits hamper people's ability to use the platform to express voice almost entirely. They can shut people out of social movements and political discourse in critical moments, potentially undermining calls for action gaining momentum through Meta's products. Meta appeared to partly recognise this in May 2022, issuing directions to its escalation teams not to impose strikes for "marg bar Khamenei" statements. However, this measure was not intended to apply to decisions made by moderators at scale. When the same content is assessed as violating at scale (i.e. by Meta's outsourced moderators), strikes result against the post author's account automatically. A reviewer at scale ordinarily has no discretion to withhold a strike or resulting penalties; this did not change when Meta issued the limited exception in May. Only content decisions that reached its escalation teams would have the option of withholding a strike for violating content. This may have applied to content or accounts that went through programmes such as cross-check, which enable users' content to be reviewed on escalation prior to removal. Whereas high-profile accounts may have benefited from that exception, the author of the post in this case did not.

The Board is also concerned that the author of the post was not sufficiently notified that feature limits were imposed, particularly on group-related features. The first time they would be aware of some of them was when they attempted to use the relevant features. This compounds users' frustrations at being placed in "Facebook jail", when the nature of the punishment and reasons for it are often unknown. In the "Mention of the Taliban in news reporting" case, the Board expressed similar concerns when a user was blocked from full use of their account at a crucial political moment. According to Meta, this issue is already known to them and, consistent with the Board's prior recommendations, relevant internal teams are working to improve its user-communication infrastructure on this issue.

The Board welcomes that Meta has progressed with implementing other recommendations in the "Mention of the Taliban in news reporting" case, and that the strike threshold for imposing some feature limits will increase. Reflecting those changes in the transparency centre explanation of Meta's enforcement practices is good practice.

The Board notes that the author of the post in this case did not have their appeal reviewed, and received misleading notifications on the reasons for this. This is troubling. Meta has publicly announced that the company is shifting towards using automation to prioritise content for review. Appeals are automatically closed without review if they do not meet a set

threshold concerning various signals including: the type and virality of the content, if the violation is not an extremely high-severity violation, such as suicide content, and the time elapsed since the content was posted. The Board is concerned about Meta's automatic closure of appeals, and that the prioritisation signals that the company is applying may not sufficiently account for public interest expression, particularly when it relates to protests. The signals should include features such as topic sensitivity and false-positive probability to help identify content of this nature, to avoid appeals against erroneous decisions being automatically closed. This is especially important where, as a result of incorrect enforcement of its policies, users are locked out of using key features on Meta's products at crucial political moments.

III. Transparency

The Board welcomes the increase in data and examples that Meta is providing of narrow newsworthiness exceptions in the Transparency Centre. These disclosures would be further enhanced by distinguishing "scaled" and "narrow" allowances that the company grants annually in its transparency reports. Providing examples of "scaled" newsworthiness allowances would also advance understanding of the steps that Meta is taking to protect user voice. In key moments, such as the Iran protests, Meta should publicise when scaled newsworthiness allowances are issued at the time, so that people understand that their speech will be protected.

8.2 Compliance with Meta's human rights responsibilities

The Board finds that Meta's initial decision to remove the content is inconsistent with Meta's human rights responsibilities as a business.

Freedom of expression (Article 19, ICCPR)

Article 19 of the ICCPR provides "particularly high" protection for "public debate concerning public figures in the political domain and public institutions" (General Comment No. 34, para. 38). Extreme restrictions on freedom of expression and assembly in Iran make it especially crucial that Meta respects these rights, in particular at times of protest ("Colombia protests" case decision; General Comment No. 37, at para. 31). The expression in this case was artistic and a political protest. It related to discourse on the rights of women and their participation in political and public life, and freedom of religion or belief. The Board has recognised the importance of protest speech against a head of state. even where it is offensive. as they are

...legitimate subject to criticism and political opposition" ("Colombia protests" case; General Comment No. 34, at paras 11 and 38). Freedom of expression in the form of art protects "cartoons that clarify political positions" and "memes that mock public figures" (A/HRC/44/49/Add.2, at para. 5).

Where a State restricts expression, it must meet the requirements of legality, legitimate aim, and necessity and proportionality (Article 19, para. 3, ICCPR). The Board uses this three-part test to interpret Meta's voluntary human rights commitments, both for the individual content decision and what this says about Meta's broader approach to content governance. As the UN Special Rapporteur on freedom of expression has stated, although "companies do not have the obligations of Governments, their impact is of a sort that requires them to assess the same kind of questions about protecting their users' right to freedom of expression" (A/74/486, at para. 41).

1. Legality (clarity and accessibility of the rules)

The principle of legality under international human rights law requires rules that limit expression to be clear and publicly accessible (General Comment No.34, at para. 25). It further requires that rules restricting expression "may not confer unfettered discretion for the restriction of freedom of expression on those charged with [their] execution" and "provide sufficient guidance to those charged with their execution to enable them to ascertain what sorts of expression are properly restricted and what sorts are not" (*Ibid*). Applied to the Community Standards, the UN Special Rapporteur on freedom of expression has said that they should be clear and specific (A/HRC/38/35, at para. 46). People using Meta's platforms should be able to access and understand the rules, and content reviewers should have clear guidance on their enforcement.

It is welcome that, consistent with prior Board decisions, Meta has ensured the translation of its content policies into more languages, including Farsi.

Meta's internal guidelines (the "known questions") on its Violence and Incitement policy contain presumptions of risk that are not currently in the public-facing policy. The Community Standard does not reflect the explanation in the internal guidance that a "death to X" statement is generally permitted except when the "target" is a "high-risk person". It is a serious concern that this hidden presumption also has a non-public exception, in particular as it relates to expression that may be legitimate political criticism of state actors. The Board

further notes that there are no examples of "high-risk persons" in the public-facing Community Standard, so it is not known that heads of state receive this particular protection. Indeed, the rationale for including some high-ranking public officials on the internal list and not others, such as members of the legislature and judiciary, is unclear. At the same time, the Board acknowledges that there may be good reasons for not disclosing the full list of high-risk targets publicly, in particular for individuals who are not afforded the protection of the State's security apparatus. In the policy rationale for the Violence and Incitement Community Standard, which is public, Meta states that it considers "language and context" to differentiate content that contains a "credible threat to public or personal safety" and "casual statements". However, the "do not post" section of the policy does not explain how language and context figure in the assessment of threats and calls for death or high-severity violence. Whereas the policy rationale appears to accommodate rhetorical speech of the kind that might be expected in protest contexts, the written rules and corresponding guidance to reviewers do not. Indeed, enforcement in practice, in particular at scale, is more formulaic than the rules imply, and this may create misperceptions to users of how rules are likely to be enforced. The guidance to reviewers, as currently drafted, excludes the possibility of contextual analysis, even when there are clear cues within the content itself that threatening language is rhetorical.

The Violence and Incitement policy requires revision, as do Meta's internal guidelines. The policy should include an explanation of how Meta moderates rhetorical threats, including "death to" statements against "high-risk persons", and how language and context are factored into assessments of whether a threat against a head of state is credible under the Violence and Incitement policy. Internal guidance should be especially sensitive to protest contexts where the protection of political speech is crucial. The related presumptions in its internal guidance should be nuanced and brought into alignment with this.

While Meta has provided further public explanations of the newsworthiness allowance, the lack of public explanation of "scaled" allowances is a source of confusion. Meta should make a public announcement when it issues a scaled allowance in relation to events like the protests in Iran, and either specify their duration or announce when those exceptions are lifted. This would help people using its platforms to understand what expression is permissible. Such announcements are opportune moments to remind people who use Meta's platforms of the existence of the rules, to raise awareness and understanding of them. This is especially important when those changes have material impacts on users' ability to express themselves on the platform.

II. Legitimate aim

Restrictions on freedom of expression must pursue a legitimate aim. The Violence and Incitement Community Standard aims to "prevent potential offline harm" by removing content that poses "a genuine risk of physical harm or direct threats to public safety". This policy therefore serves the legitimate aim of protecting the right to life and the right to security of person (Article 6, ICCPR; Article 9 ICCPR).

III. Necessity and proportionality

The principle of necessity and proportionality provides that any restrictions on freedom of expression "must be appropriate to achieve their protective function; they must be the least intrusive instrument amongst those which might achieve their protective function; [and] they must be proportionate to the interest to be protected" (General Comment No. 34 , para. 34).

The Board finds that the removal of this content was not necessary to protect Ayatollah Khamenei from violence or threats thereof. Meta is right to be concerned about threats of violence, including those targeting high-ranking public officials in many contexts. However, its decision not to interpret the Violence and Incitement policy to permit this rhetorical content, and other content like it, is a serious concern. Not issuing a scaled allowance for "marg bar Khamenei" statements compounded the problem of Meta's policy not protecting this speech, failing to mitigate the harm to freedom of expression.

Meta knew in mid-July, when this post was made, that the "National Day of Hijab and Chastity" was approaching. In May 2022, Meta already issued guidance to remove the "marg bar Khamenei" statement for decisions made at escalation without imposing a strike. Meta also knew that its platforms have been crucial in similar moments in the recent past for organising protests in Iran, having previously issued scaled allowances for "marg bar Khamenei" statements around protests. The company should therefore have anticipated issues around the over-removal of protest content in Iran, and it should have developed a response beyond the very limited strike exemption for escalated decisions. As this case shows, its failure to do more for users' voice led to protestors' freedom of expression being unnecessarily restricted. When feature limits were imposed, the impacts of its wrongful decisions were made more severe, as they prevented users from organising on Meta's platforms. The fact that Meta scaled a spirit of the policy allowance on 23 September 2022 for the phrase "I will kill whoever kills my sister/brother" in Iran indicates that Meta should have

THE PHRASE "I WILL KILL WHOEVER KILLS MY SISTER/BROTHER" IN IRAN INDICATES THAT META SHOULD HAVE ALSO PERMITTED KNOWN PROTEST SLOGANS AT THIS CRITICAL TIME. "DEATH TO" STATEMENTS ARE NOT AS DIRECTLY THREATENING AS "I WILL KILL" STATEMENTS. ALTHOUGH THE "DEATH TO" PHRASE IN THIS CASE TARGETS A SPECIFIC INDIVIDUAL, THAT TARGET IS AYATOLLAH KHAMENEI, A HEAD OF STATE, WHO ROUTINELY USES THE FULL COERCIVE FORCE OF THE STATE, BOTH THROUGH JUDICIAL AND EXTRAJUDICIAL MEANS, TO REPRESS DISSENT. IT IS CRUCIAL THAT META PRIORITISE ITS VALUE OF VOICE IN SUPPORT OF INDIVIDUALS' FREEDOM OF EXPRESSION RIGHTS IN SITUATIONS SUCH AS THIS.

THE FACTORS IDENTIFIED ABOVE WEIGH HEAVILY IN FAVOUR OF PRESUMING "MARG BAR KHAMENEI" STATEMENTS MADE IN THE CONTEXT OF PROTESTS ARE POLITICAL SLOGANS AND ARE NOT CREDIBLE THREATS. THE SIX-FACTOR TEST DESCRIBED IN THE RABAT PLAN OF ACTION SUPPORTS THIS CONCLUSION. THE SPEAKER WAS NOT A PUBLIC FIGURE AND THEIR RHETORIC DID NOT APPEAR TO BE INTENDED, AND WOULD NOT HAVE BEEN INTERPRETED BY OTHERS, AS A CALL FOR VIOLENT ACTION. AS META ITSELF DETERMINED, THE PROTEST CONTEXT IN IRAN SPECIFICALLY MADE CLEAR THAT RHETORICAL STATEMENTS OF THIS KIND WERE EXPECTED, AND THE LIKELIHOOD OF VIOLENCE RESULTING FROM THEM WAS LOW. THE BOARD FINDS THE CONTENT TO BE UNAMBIGUOUSLY POLITICAL AND NON-VIOLENT IN ITS INTENT, DIRECTLY CRITICISING A GOVERNMENT AND ITS LEADERS OF SERIOUS HUMAN RIGHTS VIOLATIONS AND DRAWING ATTENTION TO THE ABUSE OF RELIGION TO JUSTIFY DISCRIMINATION. IN THE BOARD'S VIEW, THIS CONTENT POSED VERY LITTLE RISK OF INCITING VIOLENCE. THEREFORE, BOTH THE REMOVAL AND THE ADDITIONAL PENALTIES THAT RESULTED FROM THIS DECISION WERE NOT NECESSARY OR PROPORTIONATE.

IN OTHER CONTEXTS, "DEATH TO" STATEMENTS AGAINST PUBLIC FIGURES AND GOVERNMENT OFFICIALS SHOULD BE TAKEN SERIOUSLY, AS THE INTERNAL GUIDANCE CURRENTLY IN PLACE INDICATES. FOR EXAMPLE, CONTENT WITH THE SLOGAN "MARG BAR SALMAN RUSHDIE", WOULD POSE A MUCH MORE SIGNIFICANT RISK. THE FATWA AGAINST RUSHDIE, THE RECENT ATTEMPT ON HIS LIFE AND ONGOING CONCERNS FOR HIS SAFETY, ALL PUT HIM IN A DIFFERENT POSITION TO AYATOLLAH KHAMENEI. IN OTHER LINGUISTIC AND CULTURAL CONTEXTS, "DEATH TO" STATEMENTS MAY ALSO NOT CARRY THE SAME RHETORICAL MEANING AS THE TERM "MARG BAR" CAN CARRY AND SHOULD NOT BE TREATED THE SAME AS THE CONTENT IN THIS CASE. FOR EXAMPLE, DURING EVENTS SIMILAR TO THE 6 JANUARY RIOTS IN WASHINGTON D.C., "DEATH TO" STATEMENTS AGAINST POLITICIANS WOULD NEED TO BE SWIFTLY REMOVED UNDER THE VIOLENCE AND INCITEMENT POLICY. IN SUCH A SITUATION, POLITICIANS WERE CLEARLY AT RISK AND "DEATH TO" STATEMENTS ARE LESS LIKELY TO BE UNDERSTOOD AS RHETORICAL OR NON-THREATENING IN ENGLISH.

MOREOVER, THE BOARD IS CONCERNED THAT THE RATIONALE FOR THE LIST OF "HIGH-RISK" PERSONS APPEARS IN SOME RESPECTS OVERLY BROAD IN TERMS OF THE PRESUMPTION FOR REMOVAL IT CREATES, BUT THEN INEXPLICABLY NARROW IN OTHER RESPECTS. IN THE CASE OF HEADS OF STATE, THOUGH THE BOARD AGREES THAT THEY MAY BE CONSIDERED HIGH-RISK PERSONS, INTERNAL GUIDANCE SHOULD REFLECT THAT PROTEST-

related rhetorical political speech that does not incite violence and is aimed at criticising them, their governments, the political regime or their policies, must be protected. This is the case even if it contains threatening statements that would be considered violating towards other high-risk individuals.

When rhetorical threats against heads of state are used in the context of ongoing protests, reviewers should be required to consider language and context, bringing the guidance for moderators in line with the policy rationale. This would have the effect of permitting rhetorical threats targeted at heads of state, including "death to" a head of state, where, for example: historical and present usage of phrase across platforms evidences rhetorical political speech that is not intended to, and is not likely to, incite violence; the content as a whole is engaged in criticising governments, political regimes, their policies and/or their leaders; the statement is used in protest contexts or other crisis situations where the role of government is a topic of political debate; or it is used in contexts where systematic restrictions on freedom of expression and assembly are imposed, or where dissent is being repressed.

The Board acknowledges that this issue is not as straightforward as it may first appear, and it is not possible to adopt a global rule on the use of certain terms that excludes the need for consideration of contextual factors, including signals in the content itself that are possible to consider at scale (see the Board's decision in the "Wampum Belt" case) . Meta's current position is leading to over-removal of political expression in Iran at a historic moment and potentially creates more risks to human rights than it mitigates. In the Board's view, the frequency with which Meta has needed to apply allowances in this situation indicates that a more permanent solution to this problem is required. The reliance on allowances is too ad hoc and does not provide certainty that people's expression rights will be respected. Meta needs to protect voice at scale in relation to Iran, and other critical political contexts and situations.

The proportionality concerns with this content removal increase where "feature limits" are imposed as a result of an incorrect decision. The nature and duration of the penalties were disproportionate. Meta's approach to penalties should take into greater consideration the potential for them to deter people from future engagement on political issues on the platform. It is positive that Meta has introduced further transparency and coherence in this area as a result of implementing prior Oversight Board recommendations, moving towards what should be a more proportionate and transparent approach with higher strike-to-penalty thresholds. Meta's plans to issue more comprehensive penalty notifications should ensure that users are better placed to understand the reasons for the consequences of strikes and the reasons for

feature limits in the future.

Access to remedy

Access to effective remedy is a core component of the UN Guiding Principles on the Business and Human Rights (UNGPs). In August 2020, Meta publicly announced that it would rely more on automated content review and "teams will be less likely to review lower severity reports that aren't being widely seen or shared on our platforms". The Board is concerned that Meta's automatic closure of appeals means that users are not provided with appropriate access to remedy. Additionally, the fact that the current automated system does not take into account signals such as topic sensitivity and the likelihood of enforcement error makes it very likely that the most important complaints will not be reviewed. The Board finds that this may particularly affect online protesters' right to remedy because content wrongfully removed is restored belatedly or not at all, shutting them out of social movements and political discourse in critical political moments.

8.3 Identical content with parallel context

The Board expresses concern about the likely number of wrongful removals of Iran protest content including the phrase "marg bar Khamanei". It is important that Meta take action to restore identical content with parallel context that it has incorrectly removed where possible, and reverse any strikes or account-level penalties that it has imposed as a result.

9. Oversight Board decision

The Oversight Board overturns Meta's original decision to remove the content for violating the Violence and Incitement Community Standard.

10. Policy advisory statement

Content policy

1. Meta's Community Standards should accurately reflect its policies. To better inform users of the types of statements that are prohibited, Meta should amend the Violence and Incitement Community Standard to (i) explain that rhetorical threats such as "death to X" statements are generally permitted, except when the target of the threat is a high-risk person; (ii) include an illustrative list of high-risk persons, explaining that they may include

person; (ii) include an illustrative list of high-risk persons, explaining that they may include heads of state; (iii) provide criteria for when threatening statements directed at heads of state are permitted to protect clearly rhetorical political speech in protest contexts that does not incite to violence, taking language and context into account, in accordance with the principles outlined in this decision. The Board will consider this recommendation implemented when the public-facing language of the Violence and Incitement Community Standard reflects the proposed change, and when Meta shares internal guidelines with the Board that are consistent with the public-facing policy.

Enforcement

2. Meta should err on the side of issuing scaled allowances where (i) this is not likely to lead to violence; (ii) when potentially violating content is used in protest contexts; and (iii) where public interest is high. Meta should ensure that their internal process to identify and review content trends around protests that may require context-specific guidance to mitigate harm to freedom of expression, such as allowances or exceptions, are effective. The Board will consider this recommendation implemented when Meta shares the internal process with the Board and demonstrates through sharing data with the Board that it has minimised incorrect removals of protest slogans.

3. Pending changes to the Violence and Incitement policy, Meta should issue guidance to its reviewers that "marg bar Khamenei" statements in the context of protests in Iran do not violate the Violence and Incitement Community Standard. Meta should reverse any strikes and feature limits for wrongfully removed content that used the "marg bar Khamenei" slogan. The Board will consider this recommendation implemented when Meta discloses data on the volume of content restored and number of accounts impacted.

4. Meta should revise the indicators that it uses to rank appeals in its review queues and to automatically close appeals without review. The appeals prioritisation formula should include, as it does for the cross-check ranker, the factors of topic sensitivity and false-positive probability. The Board will consider this implemented when Meta shares with the Board their appeals prioritisation formula and data that shows that it is ensuring review of appeals against the incorrect removal of political expression in protest contexts.

Transparency

5. Meta should announce all scaled allowances that it issues, their duration and notice of their

expiry, in order to give people who use its platforms notice of policy changes allowing certain expression, alongside comprehensive data on the number of "scaled" and "narrow" allowances granted. The Board will consider this recommendation implemented when Meta demonstrates regular and comprehensive disclosures to the Board.

6. The public explanation of the newsworthiness allowance in the Transparency Centre should (i) explain that newsworthiness allowances can either be scaled or narrow; and (ii) provide the criteria that Meta uses to determine when to scale newsworthiness allowances. The Board will consider this recommendation implemented when Meta updates the publicly available explanation of newsworthiness and issues Transparency Reports that include sufficiently detailed information about all applied allowances.

7. Meta should provide a public explanation of the automatic prioritisation and closure of appeals, including the criteria for both prioritisation and closure. The Board will consider this recommendation implemented when Meta publishes this information in the Transparency Centre.

***Procedural note:**

The Oversight Board's decisions are prepared by panels of five Members and approved by a majority of the Board. Board decisions do not necessarily represent the personal views of all members.

For this case decision, independent research was commissioned on behalf of the Board. An independent research institute headquartered at the University of Gothenburg and drawing on a team of over 50 social scientists on six continents, as well as more than 3,200 country experts from around the world. The Board was also assisted by Duco Advisors, an advisory firm focusing on the intersection of geopolitics, trust and safety, and technology. Memetica, a digital investigations group providing risk advisory and threat intelligence services to mitigate online harms, also provided research. The company Lionbridge Technologies, LLC, whose specialists are fluent in more than 350 languages and work from 5,000 cities across the world, provided linguistic expertise.