Alleged Audio Call to Rig Elections in Iraqi Kurdistan

2025-019-FB-UA

Summary

The Oversight Board has overturned Meta's decision not to label a likely manipulated audio clip of two Iraqi Kurdish politicians discussing rigging parliamentary elections, less than two weeks before the polls opened, in a highly contested and polarized election. The Board requires Meta to label the content.

The Board is concerned that, despite the increasing prevalence of manipulated content across formats, Meta's enforcement of its manipulated media policy is inconsistent. It must prioritize investing in technology to identify and label manipulated audio and video at scale in order that users are properly informed.

As in this case, Meta's failure to automatically apply a label to all instances of the same manipulated media is incoherent and unjustifiable.

Additionally, Meta should make labels for manipulated media available in the local language already available on its platforms. This should, at the least, form part of Meta's electoral integrity efforts.

About the Case

Less than two weeks before Iraqi Kurdistan's parliamentary elections, in October 2024, a popular media outlet affiliated with one of the region's main political parties, the Kurdistan Democratic Party (KDP), shared a two-minute audio clip on its Facebook page. The post's caption in Sorani Kurdish alleges the audio is a "recorded conversation" between brothers Bafel and Qubad Talabani, members of the region's other main political party, Patriotic Union of Kurdistan (PUK), about their "sinister plans" to rig the October 2024 elections. In the audio, two men speak with an English

voiceover (with Sorani Kurdish and English subtitles). One man says a "minimum of 30 seats" have been guaranteed to the PUK but they must "get rid" of the "11 seats" the KDP allegedly "has always been using to their advantage." The other man agrees, emphasizing the need to make it appear that those seats have been legitimately won since people are aware – yet cannot prove – that the PUK is supported by Baghdad and their "neighbor." The media outlet's Facebook page has about 4,000,000 followers. The post has had about 200,000 views.

Two users reported the content for misinformation but Meta closed the reports without review. After one of those users appealed to Meta, the company upheld its decision based on a classifier score. The user then appealed to the Oversight Board.

Meta identified other posts containing the audio clip on the Facebook and Instagram pages of the same media outlet and the KDP Facebook page. After consulting with a news media outlet based outside of Iraqi Kurdistan and a Trusted Partner to review the possibility of the audio being digitally created, Meta labeled some of the posts, but not the content in this case. The label applied to other posts with the same audio states: "This content may have been digitally created or altered to seem real."

Key Findings

When it comes to identifying AI-created or manipulated content on its platforms, Meta told the Board that it is only able to automatically identify and label static images, not video or audio content. Given the company's expertise and resources and the wide usage of Meta's platforms, it must prioritize investing in technology to identify and label manipulated video and audio at scale.

Meta's failure to deploy the tools it has to automatically apply the "AI Info" label to all instances of the same manipulated media is incoherent and unjustifiable. In the Altered Video of President Biden case, Meta committed to implementing the Board's recommendation that for manipulated media, not violating other Community Standards, the company should apply a label to "all identical instances of that media on the platform." Meta's claim in this case that it does not automatically apply the "High

Risk” label to content containing the audio contradicts this recommendation. “AI Info” and “High Risk” labels are informative labels Meta has for manipulated media.

The Board notes there are reliable indicators, including technical signals, that the clip was digitally created. It meets the requirements of manipulated media under Meta’s Misinformation policy. Placing a “High Risk” label on it is consistent with Meta’s policies and human rights responsibilities. The audio was posted during a highly contested electoral period in a region with a history of irregular elections. This increases the audio’s ability to influence electoral choices and harm electoral integrity. Placing an informative label on the case content, instead of removing the content altogether, satisfies the requirements of necessity and proportionality.

The Board is concerned that Meta’s manipulated media labels are not available in Sorani Kurdish. This is despite Sorani Kurdish being one of the in-app languages available to Facebook users. To ensure users are informed when content is digitally created or altered, making the label available in the local language already available on Meta’s platforms should, at the least, form part of its electoral integrity efforts.

The Board is also concerned by the company’s reliance on third parties for technical assessment of likely manipulated content. Meta should reconsider having this expertise available internally.

The Board notes that the issue in this case concerns whether the audio is real or fake, rather than if what is said in the audio is true. Given that labeling the audio as likely digitally created or altered would similarly alert users to the accuracy of its content, the Board finds the application of the Misinformation policy on manipulated media to be sufficient. However, the Board is concerned that Meta did not have Kurdish language fact-checkers available to review content during the election as part of its election integrity measures.

The Oversight Board’s Decision

The Oversight Board overturns Meta’s decision not to label the content, requiring the post to be labeled.

The Board also recommends that Meta:

- Apply a relevant label to all content with the same manipulated media, including all posts containing the manipulated audio in this case.
- Ensure that the informative labels for manipulated media on Facebook, Instagram and Threads are displayed in the same language that the user has elected for its platform.

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

Full Case Decision

1. Case Description and Background

Less than two weeks before Iraqi Kurdistan's parliamentary elections, in October 2024, a popular media outlet affiliated with the political party the Kurdistan Democratic Party (KDP), shared a two-minute audio clip on its Facebook page. The post's caption in Sorani Kurdish alleges the audio is a "recorded conversation" between brothers Bafel and Qubad Talabani, members of the political party Patriotic Union of Kurdistan (PUK), about their "sinister plans" to rig the October 2024 elections. In the audio, two men speak with an English voiceover (with Sorani Kurdish and English subtitles). One of them reassures the other that a "minimum of 30 seats" have been guaranteed to them, but they must "get rid" of the "11 seats" the KDP allegedly "has always been using to their advantage." This would give the KDP "a taste of their own medicine," according to the speaker. The other speaker agrees, emphasizing the need to make it appear that those seats have been legitimately won since people are aware – yet cannot prove – that they are supported by Baghdad and their "neighbor." The media outlet's Facebook page has about 4,000,000 followers. The post has had about 200,000 views and has been shared under 100 times.

A user reported the post the same day it was posted, under Meta's Misinformation Community Standard. However, the report was not prioritized for human review and

Meta closed the report automatically, keeping up the post on Facebook without further action. Meta's High Risk Early Review Operations (HERO) system also identified the content in this case for further review based on indications it was likely to go viral. The report was later closed without human review because, according to Meta, the "virality was not high enough for it to proceed to review stage." Shortly after, another user reported the content for misinformation, but Meta closed this report as well. The second user appealed this with Meta and the company upheld its decision based on a classifier score. The user then appealed Meta's decision to the Oversight Board. A day later, Meta identified other posts containing the audio clip shared on the Facebook and Instagram pages of the same media outlet that shared the post in this case, as well as the Facebook page of the KDP. Meta's Content Policy team then sent these posts to a news outlet for their assessment of the likelihood that the content in the posts was AI-generated, pursuant to a 2024 partnership. A Trusted Partner also provided input on technical and contextual signals to help determine whether the content was AI-generated. After these consultations, Meta decided to label some posts containing the audio clip, but not the content in this case. The label applied to other posts with the same audio states: "This content may have been digitally created or altered to seem real." The Trusted Partner Program is a network of NGOs, humanitarian agencies and human rights researchers from 113 countries that report content and provide feedback to Meta about its content policies and enforcement.

The Board notes the following context in reaching its decision.

The KDP and PUK are the main political parties in Iraqi Kurdistan. Each has its own security forces. Longstanding rivalry existed between the parties, culminating in a civil war from 1994 to 1998. Eventually, the parties reached a fragile power-sharing arrangement that temporarily mitigated their feud. However, in recent years, tensions have resurfaced and intensified, leading to years-long delays of elections that were finally held in October 2024.

Most media outlets in Iraqi Kurdistan are "directly affiliated" with political parties, with parties or party leaders providing funding. This leads to partisan news reporting. Experts consulted by the Board as well as public reports state that the media outlet that

posted the case content is affiliated with and funded by the KDP and lacks editorial independence.

Facebook is the third most widely used platform in Iraq. The use of coordinated disinformation campaigns is common during Kurdish political processes, such as the 2018 parliamentary elections. "Shadow media," a network of social media pages affiliated with political parties and influential political figures, allegedly attempted to influence public opinion ahead of the October 2024 elections in favor of their patrons, who include key figures of the two parties.

On October 20, 2024, the KDP won the most parliamentary seats in the election. The PUK came second. The KDP's Masrour Barzani is the current prime minister. The PUK's Qubad Talabani, alleged to be one of the speakers in the audio, is the current deputy prime minister.

2. User Submissions

The user who reported the content to the Board stated the audio was AI-generated, was shared during a "sensitive" election campaign and was being used to "damage a party's reputation."

The media outlet that posted the content also submitted a statement to the Board explaining it shared the audio as part of its news coverage, "without offering any opinions or commentary on the content itself" but rather "simply as a local news item." It said its goal is to publish "accurate" news. The media outlet stated a political party released the audio and that "many political parties regularly share such videos."

3. Meta's Content Policies and Submissions

I. Meta's Content Policies

Misinformation Community Standard

Meta's Misinformation Community Standard governs the moderation of manipulated media on its platforms. For content that does not violate the "do not post" portion of

the Community Standard, the policy focuses on "reducing its prevalence" or "creating an environment that fosters productive dialogue." For this purpose, Meta may place an informative label on the face of the content – or reject content submitted as an advertisement – when the content is a photorealistic image or video, or realistic-sounding audio, that was digitally created or altered and creates a "particularly high risk of materially deceiving the public on a matter of public importance."

II. Meta's Submissions

Meta has three different informative labels for manipulated media: (i) "AI Info" label, (ii) "High Risk" label and (iii) "High Risk AI" label.

The "AI Info" label applies automatically to content that Meta detects through "industry standard AI image indicators or when people disclosed that they were uploading AI-generated content." Meta previously informed the Board that to detect industry standard image indicators the company relies on "metadata that GenAI creation tools embed in the content." However, presently, the automatic detection and application of the "AI Info" label does not apply to video or audio. In such cases, Meta relies on users to disclose that they are uploading AI-generated content.

The "High Risk" label applies to digitally created or altered content, while the "High Risk AI" label applies to AI-generated content, according to internal enforcement guidelines. Both labels apply to image, audio and video content meeting all the following conditions: (i) it creates a particularly high risk of materially deceiving the public on a matter of public importance; (ii) there are reliable indicators that the content was digitally created or altered; (iii) the "High Risk AI" label additionally requires the content to have reliable indicators of being created or altered with AI. The label affixed to a post varies depending on whether it is a "High Risk" label or a "High Risk AI" label. A "High Risk" label states, "This content *may* have been digitally created or altered to seem real," while a "High Risk AI" label states "This content *was* digitally created or altered with AI to seem real" (emphasis added). Both "High Risk" labels come with a link to "learn more," directing the user to this article. According to Meta, the company can provide more precise information to users by having two different labels, about how likely it is that the content is manipulated, and the method used to alter or create the

media. When the company is less certain that the content is digitally created, it affixes the "High Risk" label, alerting users that the content *may* be digitally created or altered. The label affixed to a post normally defaults to the language set by the user on the platform, but Meta stated it does not translate the label to Kurdish (even though users can set their default language to two different variants of Kurdish, including Sorani Kurdish).

Meta said that placing an informative label – whether an "AI Info" label or either one of the "High Risk" labels – does not result in the demotion of the content or its removal from recommendations. Meta may show a pop-up to users who click to reshare content with the "High Risk" label. When users attempt to reshare a "High Risk"-labeled image or video on Facebook, Instagram or Threads, they will receive a notice alerting them that the content may have been digitally created or altered. However, this is not available when posts are shared as Instagram stories or Facebook reels, as in this case.

After Meta's internal teams identified three posts containing the same audio clip, they sent the posts to a news outlet to review the possibility of the audio being digitally created. In addition, Meta consulted a Trusted Partner to assess the audio's authenticity. Meta also considered other factors specific to Iraqi Kurdistan, such as the brief interval between the date the audio was shared and the date of the upcoming elections, the lack of Kurdish language fact-checkers in Iraqi Kurdistan, comments indicating some users did not see the audio as manipulated and Meta's internal classifiers detecting other instances of the audio potentially going viral. Based on these considerations, Meta affixed a "High Risk" label on the posts with the same audio the company had detected and assessed, but not the content in this case. The company explained that it did not apply the label to all posts containing the audio to avoid mislabeling content that did not meet the criteria for the label, such as posts debunking or condemning the audio. For Meta, labeling all instances of the audio could confuse users.

As part of its integrity efforts for the October 2024 elections, Meta informed the Board that it conducted daily targeted searches for hostile speech and voter interference violations, political ads monitoring, and monitoring for impersonation and harassment

of political candidates. Meta also noted that it organized a cross-functional team with language and contextual expertise to assess on-platform risks and implement appropriate mitigation measures before and during the October 20, 2024, elections. This was how the three viral posts containing the audio, separate from the case content, were identified. An internal classifier detected these three posts as potentially going viral, with over 1.5 million combined views.

The Board asked questions on: the labels Meta applies for manipulated media; the process for identifying such media and when Meta automatically applies a label to similar or identical content; whether the content was fact-checked; and Meta's election integrity efforts in Iraqi Kurdistan. Meta responded to all questions.

4. Public Comments

The Oversight Board received two public comments that met the terms for submission. One comment was submitted from the Middle East and North Africa and the other from Central and South Asia. To read public comments submitted with consent to publish, click here.

The submissions covered the following themes: the authenticity of the audio clip; the prevalence of propaganda during elections in Iraqi Kurdistan; and press freedom in the region.

5. Oversight Board Analysis

The Board selected this case to address how Meta ensures freedom of expression while tackling manipulated media shared in the context of an election. This case falls within the Board's strategic priority of Elections and Civic Space.

The Board analyzed Meta's decision in this case against Meta's content policies, values and human rights responsibilities. The Board also assessed the implications of this case for Meta's broader approach to content governance.

5.1 Compliance With Meta's Content Policies

Content Rules

Misinformation Community Standard

The Board finds that the case content meets the requirements of manipulated media under Meta's Misinformation policy. Therefore, a "High Risk" label should have been applied to it.

The post concerns a matter of public importance in the region, i.e. the Iraqi Kurdistan elections. The content creates a high risk of deceiving the public, both domestic and international audiences. The English voiceover may have been aimed at the diplomatic personnel in the region or international observers, and the subtitles in Sorani Kurdish made it accessible to the electorate. The content had around 200,000 views and was shared right before the election. There are reliable indicators that the content was digitally created or altered. Sources relied upon by Meta confirmed this when assessing the audio's authenticity. These sources included one Trusted Partner and one news outlet outside of Iraqi Kurdistan that Meta relied on to assess content that may have been digitally created or altered. These sources highlighted the stilted nature of the conversation and the lack of conversational beats. They also noted technical signals indicating the audio was likely digitally created. These signals included disjointed background audio, as well as ratings by AI detector tools. Approval was obtained from Meta's internal teams to place a "High Risk" label on the underlying audio in other posts containing it, not the case content. Therefore, a "High Risk" label, indicating the content *may* be digitally created or altered, should have been applied to the case content. The Board finds Meta's explanation for why the company did not apply the label to all content with the same audio unconvincing. Analysis of Meta's arguments and the Board's conclusions about applying the label to all content with the same manipulated media is included in the human rights analysis section below.

5.2 Compliance With Meta's Human Rights Responsibilities

The Board finds placing a "High Risk" label on the content is required by a proper interpretation of Meta's policies. Pursuant to the International Covenant on Civil and Political Rights (ICCPR) Article 19 analysis below, applying the label to the content in

this case as well as to all instances of the audio clip is also consistent with Meta's human rights responsibilities.

Freedom of Expression (Article 19 ICCPR)

Article 19 of the ICCPR provides for broad protection of expression, including political expression (General Comment No. 34, paras. 11-12). The United Nations (UN) Human Rights Committee has stated that the value of expression is particularly high when discussing political issues (General Comment No. 34, paras. 11 and 13). When restrictions on expression are imposed by a state, they must meet the requirements of legality, legitimate aim, and necessity and proportionality (Article 19, para. 3, ICCPR). These requirements are often referred to as the "three-part test." The Board uses this framework to interpret Meta's human rights responsibilities in line with the UN Guiding Principles on Business and Human Rights, which Meta itself has committed to in its Corporate Human Rights Policy. The Board does this both in relation to the individual content decision under review 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, para. 41).

The UN Human Rights Committee has emphasized that freedom of expression is essential for the conduct of public affairs and the effective exercise of the right to vote (General Comment No. 34, para. 20; see also General Comment No. 25, paras. 12 and 25). At the same time, several UN Special Rapporteurs and UN Working Groups made a joint statement in 2024 that: "While political speech enjoys strong protection under international law, sometimes politicians have abused the freedom, using it as a license to engage in toxic discourse and to spread disinformation, including in relation to electoral outcomes and the integrity of elections. The coordinated smear campaigns and use of deepfakes through social media are particularly alarming as they have become powerful tools to manipulate elections, including by foreign actors who seek to interfere in elections from across borders" (see also UN Special Rapporteur report, A/HRC/47/25, para. 18, on how disinformation is spread by politicians and traditional

media; UN Special Rapporteur report on the realization of the right to freedom of opinion and expression in electoral contexts, A/HRC/26/30).

The Board notes that the audio was posted during a highly contested electoral period in a region with a history of elections being marred by irregularities. This increases the ability of the audio to influence electoral choices and harm the integrity of the election. As the UN Special Rapporteur's report indicates, social media platforms have a particularly powerful influence on electoral integrity (A/HRC/47/25, para. 16). The Board finds that given the UN independent expert finding of the pervasiveness of electoral disinformation around the world, Meta's human rights responsibilities necessitate taking appropriate measures to mitigate its adverse effects. When such mitigation measures limit the user's right to freedom of expression, they must meet the requirements of the three-part test under Article 19(3) of the ICCPR.

I. Legality (Clarity and Accessibility of the Rules)

The principle of legality requires rules limiting expression to be accessible and clear, formulated with sufficient precision to enable an individual to regulate their conduct accordingly (General Comment No. 34, para. 25). Additionally, these rules "may not confer unfettered discretion for the restriction of freedom of expression on those charged with [their] execution" and must "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.*). The UN Special Rapporteur on freedom of expression has stated that when applied to private actors' governance of online speech, rules should be clear and specific (A/HRC/38/35, para. 46). People using Meta's platforms should be able to access and understand the rules and content reviewers should have clear guidance regarding their enforcement.

The Board finds that the Misinformation policy on manipulated media is sufficiently clear as applied to the content in this case. The public-facing language of the policy clearly apprises users of the applicable rules and the consequences of posting such content on Meta's platforms (i.e., placing an informative label), when it does not violate other Community Standards. However, Meta should consider integrating the

information on all the different manipulated media labels on one page in the Transparency Center so that users can easily learn more about them.

II. Legitimate Aim

Any restriction on freedom of expression should pursue one or more of the legitimate aims listed in the ICCPR, which includes protecting the rights of others (Article 19, para. 3, ICCPR). In the Altered Video of President Biden case, the Board held that protecting the right to participate in public affairs (Article 25, ICCPR) is a legitimate aim.

III. Necessity and Proportionality

Under ICCPR Article 19(3), necessity and proportionality requires that restrictions on expression "must be appropriate to achieve their protective function; they must be the least intrusive instrument amongst those which might achieve their protective function; they must be proportionate to the interest to be protected" (General Comment No. 34, para. 34).

The UN Special Rapporteur on freedom of expression has noted (A/74/486, para 51) that "companies have tools to deal with content in human rights-compliant ways, in some respects a broader range of tools than that enjoyed by States." In this vein, the Board finds that placing an informative label on the case content, such as a "High Risk" label, instead of removing the content altogether, satisfies the requirements of necessity and proportionality. In the determination of the necessity and proportionality of this measure, the Board considered the following factors: a) that the content was posted close to the election; b) a history of disinformation and misinformation during previous elections; c) the polarized nature of the political environment; d) political control over news media in the region, i.e. the source of the information; e) the likelihood that the media contained in the post is altered, as indicated by two independent evaluations; and f) the likelihood that digitally altered media will mislead and influence the electorate.

The right to access information and the right to vote are intrinsically linked, mutually reinforcing and constitutive elements of democracy. The 2009 Joint Statement of the

UN Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe Representative on Freedom of the Media, the Organization of American States Special Rapporteur on Freedom of Expression and the African Commission on Human and People's Rights Special Rapporteur on Freedom of Expression and Access to Information stressed that "free and fair elections are possible only where the electorate is well informed and has access to pluralistic and sufficient information." Placing an informative label on the content is the least intrusive measure that would inform users of the likelihood that the audio clip was digitally created, which would mitigate the risks of misleading the public ahead of a highly polarized and contested election. By applying the "High Risk" label, Meta indicates to users that the content *may* be digitally created and provides guidance for how to assess the content without deciding on the veracity of the information that could influence the election. The Board further notes the partisan media environment in Iraqi Kurdistan, both in traditional media and on social media (see PC-31199, Centre for Advanced Studies in Cyber Law and Artificial Intelligence). Keeping the content up without a label adversely affects the ability of users to be informed and participate in political discourse in an already weak ecosystem for freedom of expression.

The internal consistency and coherence of the interference on expression (i.e., labeling) requires that the label be applied to all instances of the audio. In the Altered Video of President Biden case, the Board recommended that for manipulated media not violating other Community Standards, the company should apply a label to "all identical instances of that media on the platform." The Board acknowledges Meta's commitment to implement this recommendation. Applying the label to all instances of the audio that is the subject of this case would be consistent with that commitment. Meta's claim in this case that it does not automatically apply the "High Risk" label to content containing the audio contradicts this recommendation.

Importantly, labeling does not demote the content or prevent it from being recommended. Additionally, Meta's concern over confusing users by mislabeling content that debunks or condemns the audio is neither compelling nor supported by evidence. To the contrary, not applying the label to all instances of the audio increases the chances of misleading and confusing the public. The Board notes that some users

commented on the media outlet's post thinking the audio was real. According to Meta the company had applied a label to other posts from the same media outlet that also contained the same audio. Labeling some but not all posts from the same media outlet is likely to confuse, or worse, mislead users even more. As the Board noted in the Altered Video of President Biden decision, applying a label to a small portion of content "could create the false impression that non-labeled content is inherently trustworthy." The Board is concerned over Meta's selective application of the label in the lead-up to a highly contested and polarized election. The Board, however, is not convinced of the need for two separate "High Risk" labels, one indicating content *may* be manipulated and the other stating that it *was*. Meta should reassess whether this added complexity serves any useful purpose.

The Board is also concerned that Meta's manipulated media labels are not available in Sorani Kurdish, which is one of the languages in the case content (i.e., the subtitles are in Sorani Kurdish). The Board notes that Sorani Kurdish is one of the in-app languages available to Facebook users. However, users whose language settings default to Sorani Kurdish will not be able to see the AI labels, such as the "High Risk" label, in this dialect. In several decisions, the Board has recommended the translation of Meta's Community Standards and aspects of Meta's internal enforcement guidance into languages spoken by its users. This is to ensure users are informed of the rules and so that there is accurate enforcement (See Punjabi Concern over the RSS in India, Myanmar Bot, Reclaiming Arabic Words). Similarly, the "High Risk" label is meant to inform the user that the content may be digitally manipulated. At the least, making the label available in the local language already available on Meta's platforms should form part of its electoral integrity efforts.

The Board notes that the issue in this case concerns the authenticity of the audio, independent of the accuracy of the allegations made therein. Given that labeling the audio as likely digitally created or altered would similarly alert users to the accuracy of its content, the Board finds the application of the Misinformation policy on manipulated media to be sufficient. However, the Board is concerned that Meta did not have Kurdish language fact-checkers available to review content during the election as part of its election integrity measures. This is especially so in the context of a highly polarized

political environment, limited independent media, and a history of electoral misinformation and coordinated disinformation campaigns.

Enforcement

The Board is concerned both with the inconsistent and limited enforcement of Meta's manipulated media policy and the company's reliance on third parties for technical assessment of likely manipulated content. In the Altered Video of President Biden decision, Meta told the Board that "videos involving speech were considered the most misleading and easiest to reliably detect." In that case, the Board highlighted to Meta that "audio-only content can include fewer cues of inauthenticity and therefore be as or more misleading than video content." In that case, the Board suggested that Meta not focus on photographs, or static images, identifying manipulated video and audio as the priority. However, when it comes to identifying AI-created or manipulated content on its platforms, Meta reported that it is only able to automatically identify and label static images, not video or audio content. In applying the "AI Info" label to audio and video, the company relies on users to self-disclose the use of AI. Given that Meta is one of the leading technology and AI companies in the world, with its resources and the wide usage of Meta's platforms, the Board reiterates that Meta should prioritize investing in technology to identify and label manipulated video and audio at scale. Additionally, for content that the company has already identified and assessed as likely manipulated and that "creates a particularly high risk of materially deceiving the public on a matter of public importance," Meta's failure to deploy the tools it has to automatically apply the label to all content with the manipulated media is incoherent and unjustifiable.

Finally, Meta should reconsider the viability of having the expertise internally to assess whether content has been manipulated. It is not clear to the Board why a company of this technical expertise and resources outsources identifying likely manipulated media in high-risk situations to media outlets or Trusted Partners.

6. The Oversight Board's Decision

The Oversight Board overturns Meta's decision not to label the content, requiring the post to be labeled.

7. Recommendations

Enforcement

1. To ensure "High Risk" labels are applied consistently to all identical or similar manipulated content, Meta should apply the relevant label to all content with the same manipulated media on its platforms, including all posts containing the manipulated audio in this case.

   The Board will consider this recommendation implemented when Meta provides the Board with a clear process for consistently identifying and applying the appropriate "High Risk" label to all instances of manipulated media across the platform.

2. As part of its electoral integrity efforts and to ensure users are informed of manipulated media on Meta's platforms in the lead-up to an election, Meta should ensure that the informative labels for manipulated media on Facebook, Instagram and Threads are displayed in the same language that the user has elected for its platform.

   The Board will consider this recommendation implemented when Meta provides information in its Transparency Center about the languages in which manipulated media labels are available to users on its platforms.

***Procedural Note:**

- The Oversight Board's decisions are made by panels of five Members and approved by a majority vote of the full Board. Board decisions do not necessarily represent the views of all Members.

- Under its Charter, the Oversight Board may review appeals from users whose content Meta removed, appeals from users who reported content that Meta left up, and decisions that Meta refers to it (Charter Article 2, Section 1). The Board has binding authority to uphold or overturn Meta's content decisions (Charter Article 3, Section 5; Charter Article 4). The Board may issue non-binding recommendations that Meta is required to respond to (Charter Article 3, Section 4; Article 4). Where Meta commits to act on recommendations, the Board monitors their implementation.

- For this case decision, independent research was commissioned on behalf of the Board. The Board was assisted by Duco Advisors, an advisory firm focusing on the intersection of geopolitics, trust and safety, and technology.

- Linguistic expertise was provided by Lionbridge Technologies, LLC, whose specialists are fluent in more than 350 languages and work from 5,000 cities across the world.