

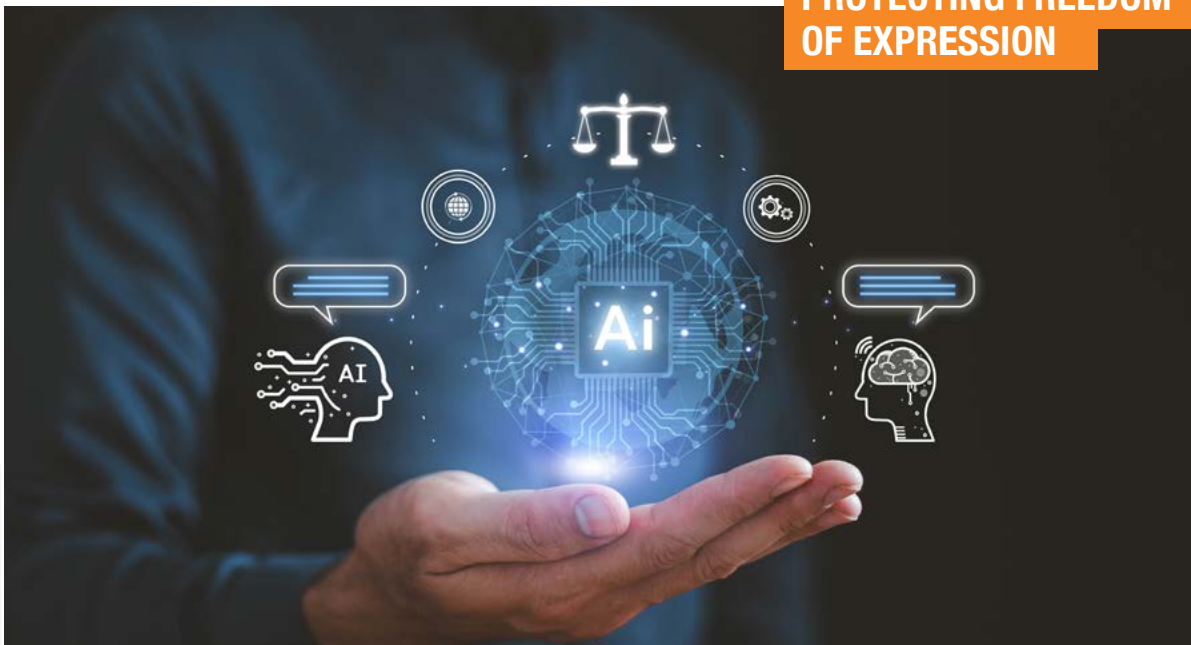
2024 ANNUAL REPORT



**IMPROVING HOW META TREATS PEOPLE
& COMMUNITIES AROUND THE WORLD**

Co-Chairs' Foreword

PROTECTING FREEDOM OF EXPRESSION



In 2024, the biggest election year in modern history, the Oversight Board published its first white paper. Drawing on analysis from our case work, the paper shared ways in which social media companies can counter challenges to the safe and reliable running of elections while protecting freedom of expression, guided by international human rights standards. This focus on elections informed the cases we selected last year on voter fraud, misinformation and political satire. We also issued an expedited decision for cases covering the post-elections violence in Venezuela.

Another theme the Board explored in 2024, through a second white paper, looked at content moderation in the era of AI and automation. Delving into the design of Meta's automated systems and the need for platforms to consider global human rights when deploying AI tools, the paper also highlighted recommendations we have made on AI-generated content. One of those key recommendations, which called on Meta to label AI-created or altered content, was adopted by the company in 2024. See page 5 of this report for the other changes made by Meta in response to our recommendations, which we issue as part of our case decisions.

In every quarter in 2024, Meta accepted and implemented many more of our recommendations than it turned down. Importantly, we made this assessment using independent methodology, which requires Meta to provide proof that implementation has taken place.

Since January 2021, we have made more than 300 recommendations to Meta. Implementation or progress on 74% of these has resulted in greater transparency, clear and accessible rules, improved fairness for users and greater consideration of Meta’s human rights responsibilities, including respect for freedom of expression.

The Board’s data team, which developed our approach to tracking implementation, has been working with Meta to understand more about the impact of this work on users. In 2024, the Board received the most detailed data to validate this impact, presented on pages 7-9. This data is a crucial part of the recommendation lifecycle, to understand how people are affected by the Board’s guidance and how we can continue to improve Meta’s platforms for the benefit of billions of users and their speech.



Evelyn Aswad, Paolo Carozza, Michael McConnell,
Pamela San Martín, Helle Thorning-Schmidt

CO-CHAIRS OF
THE OVERSIGHT BOARD

Foreword by the Chair of the Trust

TRANSPARENT & ACCOUNTABLE CONTENT GOVERNANCE



As a unique model for oversight of global content moderation, independent of companies and governments, the Board's continuing evolution aims to keep pace with industry changes and the regulatory landscape. Soon approaching five years since the Board opened its appeals process, it's remarkable to think how this experiment has grown into the institutional model for independent, principled and global content governance.

In 2024, Meta confirmed another round of funding, with a contribution of \$30 million topping up the irrevocable trust to ensure the Board's operations are funded through 2027. While Meta renewed its commitment to independent oversight, the Board also took action to optimize its operations, implementing targeted budget cuts to prioritize the most impactful aspects of its work.

With continuing optimization in mind, we have more recently made changes to the way we are governed. While ensuring the Board remains independent of Meta and the financial support the company provides, our structure has been adapted to bring Co-Chairs into the Oversight Board LLC's body of managers, creating one unified space for strategy, budget and operations to ensure the Board's impact is maximized over the long term.

Fostering accountability has never been more important for building trust among technology companies. The Board's model and its thought leadership on key governance issues demonstrate how this can be achieved.



Stephen Neal
CHAIR OF THE OVERSIGHT
BOARD TRUST

Executive Summary

The Board has made **317 recommendations to Meta since 2021, 74% of which are implemented, in progress or Meta reports as work it already does.**

In response to these recommendations, Meta made the following changes in 2024/early 2025:



Started labeling AI-created or altered content, providing additional information to users without unduly restricting speech



Improved the indicators it gives to moderators when they are reviewing long-form videos for potential violations, for more accurate enforcement of this content



Unified its policies so that they apply evenly to Facebook, Instagram and Threads, improving consistency across its platforms and clarity on the rules for users



Completed an audit of its slur lists in 22 countries with elections in 2024, as part of efforts to improve enforcement accuracy of this Hateful Conduct rule



Gave users the chance to avoid a strike on their accounts for their first violation of an eligible policy by completing an educational exercise, helping to protect expression and giving users an opportunity to learn why their posts contained a violation



Updated its Dangerous Organizations and Individuals policy to enable more speech when, in certain contexts, users refer to designated individuals as “shaheed,” effectively ending the blanket ban on the term



Allowed users to give additional context when appealing against removal of their content for hate speech, limiting overenforcement of satirical posts and other content that could benefit from other policy exceptions such as awareness raising



Aligned the public language of its Violent and Graphic Content policy with internal guidance to reviewers, clarifying to users the types of content not allowed under the policy and bringing transparency to how it is enforced

Implementation by-the-numbers

Breaking down the latest on implementation status by Meta, the table below shows a continuing upward trend of the company fully or partially implementing the Board's recommendations, as verified through published information.

The Board continues to scrutinize implementation through its own independent, data-driven approach.

Implementation Status*	Total	Difference Compared to 2023 Annual Report
Implementation demonstrated through published information	60	+15
Partial implementation demonstrated through published information	47	+17
Progress reported**	89	+8
Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation	47	+5
Recommendation declined after feasibility assessment	12	-3
Recommendation declined	34	+2
Recommendation omitted or reframed	26	+10

* Implementation status accurate as of August 18, 2025

** Progress reported is a category used for recommendations that Meta is working towards, sometimes with long timelines because of the complexities and coordination involved



**RECOMMENDATIONS TO META
BY THE BOARD**

Bringing About Change on Meta's Platforms

TO IMPROVE THE POLICIES APPLIED
TO BILLIONS OF USERS



In addition to the binding decisions we make on individual pieces of content, the Board also issues recommendations that Meta must respond to publicly within 60 days. These recommendations, which push for greater respect for human rights, transparency, consistency and fairness in Meta's content moderation, are impacting users and organizations around the world in the following ways.



How our recommendations impact users and organizations

Fewer Burdens on Speech

- ➔ In response to one of the Board's recommendations in our [Altered Video of Biden case](#), Meta started to label AI-created or altered content on its platforms in May 2024. The company is adding "AI info" labels to a range of video, audio and images across Facebook, Instagram and Threads, providing users with important context about how content has been created. Not all AI-manipulated content is harmful, and labeling offers an alternative to content removal that does not unduly restrict expression.

Over 29 days in October 2024, users viewed more than 360 million pieces of content with AI labels on Facebook and 330 million on Instagram. Of these, users on Facebook clicked on 6 million posts with these labels and 13 million pieces on Instagram, to learn more about how the content had been created.*

* All information is aggregated and de-identified to protect user privacy. All metrics are estimates, based on best information currently available for a specific point in time.

How our recommendations impact users and organizations



Preventing Strikes and Account Restrictions

- ➔ Users are now given a chance to prevent a strike being applied to their account when they commit their first violation of an eligible policy* by completing an educational exercise. Launched at the start of 2025, this system reflects a recommendation made in the Board's first policy advisory opinion on the [Sharing of Private Residential Information](#). Now when users commit a violation for the first time, Meta sends them an “eligible violation notice,” which includes details about the policy they breached, along with the option of either appealing the decision or doing the exercise. Preventing penalties in this way provides users with more information on why their content violated a non-severe Community Standard and supports their free expression by avoiding account restrictions.

More than 7.1 million Facebook and 730,000 Instagram users opted to view the “eligible violation notice” during a three-month period starting in January 2025. Among these users, nearly 3 million then embarked on the educational exercise, with the majority (80%+ on Facebook, 85%+ on Instagram) going on to complete the steps, and avoid a strike and resulting account restrictions.*

* This feature excludes the most severe Community Standards violations, such as sexual exploitation, high-risk drugs and glorifying dangerous organizations.



Improving Response Times to Trusted Partners

- ➔ Meta's work with Trusted Partners, a network of NGOs, humanitarian agencies and human rights researchers from 113 countries, helps to flag emerging harms and complex cases on Facebook, Instagram and Threads that could otherwise be missed. In a decision about [gang violence in Haiti](#), the Board noted our concerns over the different response times to reports escalated by Trusted Partners, often during times of crisis. Responding to our related recommendation:

Meta increased the number of cases resolved within five days of escalation through the program from 69% in the second quarter of 2022 to 81% in the second quarter of 2024. This increase was achieved alongside Meta receiving four times the amount of content via the program.



Supporting Freedom of Expression

- ➔ A [policy advisory opinion](#) by the Board provided detailed analysis of how Meta’s approach to the Arabic term “shaheed” (loosely translated as “martyr”) was impacting the free expression of millions of users. At the time, the term, when used to refer to individuals designated as dangerous by Meta, was the reason for more content takedowns under the Community Standards than any other single word. The Board made recommendations to update Meta’s Dangerous Organizations and Individuals policy to allow people to use “shaheed” in their post when it does not contain signals of violence and does not praise designated individuals or organizations.

Using the Meta Content Library, the Board’s own data team identified a 19.5% increase in daily posts, with more than 50,000 views, containing the word “shaheed” following implementation.*

* Measured during a pre-implementation period of March-August 2024 and post-implementation period of October 2024-April 2025, based on 31,498 posts collected.



More Awareness Around Sexual Abuse

- ➔ In a 2022 case about content showing the [sexual assault of a woman in India](#), the Board considered how to distinguish posts shared to raise awareness about sexual harassment from content intended to perpetuate violence or discrimination. A Board recommendation called on Meta to make an exception to its Adult Sexual Exploitation policy for depictions of non-consensual sexual touching without nudity, only to be applied on escalation and if the content met other criteria. Qualifying content that met the specific purpose of raising awareness of such abuse would then remain on Meta’s platforms with a warning screen.

During a three-month period starting in December 2024, more than 15,000 pieces of content across Facebook and Instagram were identified in which users raised awareness of sexual harassment or abuse in line with the policy’s new criteria. This includes content that previously would have been taken down but now remains on Meta’s platforms with a warning screen.

IN 2024, THE OVERSIGHT BOARD...



Published our **first** white papers, offering thought leadership drawn from our casework on two content moderation themes: safeguarding the integrity of elections and adapting to a new era of AI and automation



Issued our **first** decision for content on Threads regarding statements about the Japanese Prime Minister



Issued **expedited** decisions on the post-election crisis in Venezuela

ISSUED 65 DECISIONS:

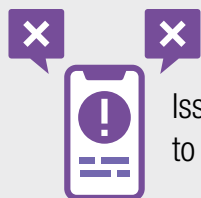
32 STANDARD
DECISIONS

31 SUMMARY
DECISIONS

2 EXPEDITED
DECISIONS

On wide-ranging issues including Holocaust denial, non-consensual intimate images, illegal voting, child marriage, political speech and homophobic violence

Published 1 policy advisory opinion on Meta's approach to moderating the Arabic term "Shaheed" when referring to dangerous organizations or individuals



Issued **48** recommendations to Meta



Received **3,250** comments

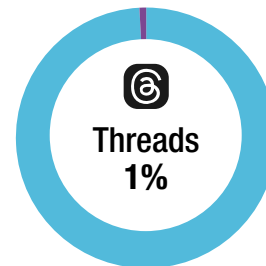
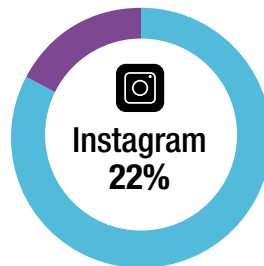
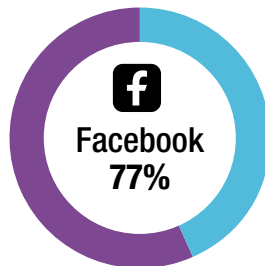
Appeals to the Board in 2024

TOTAL NUMBER: 558,235

INCLUDES 8 CASES REFERRED BY META

33% INCREASE ON 2023

BY PLATFORM

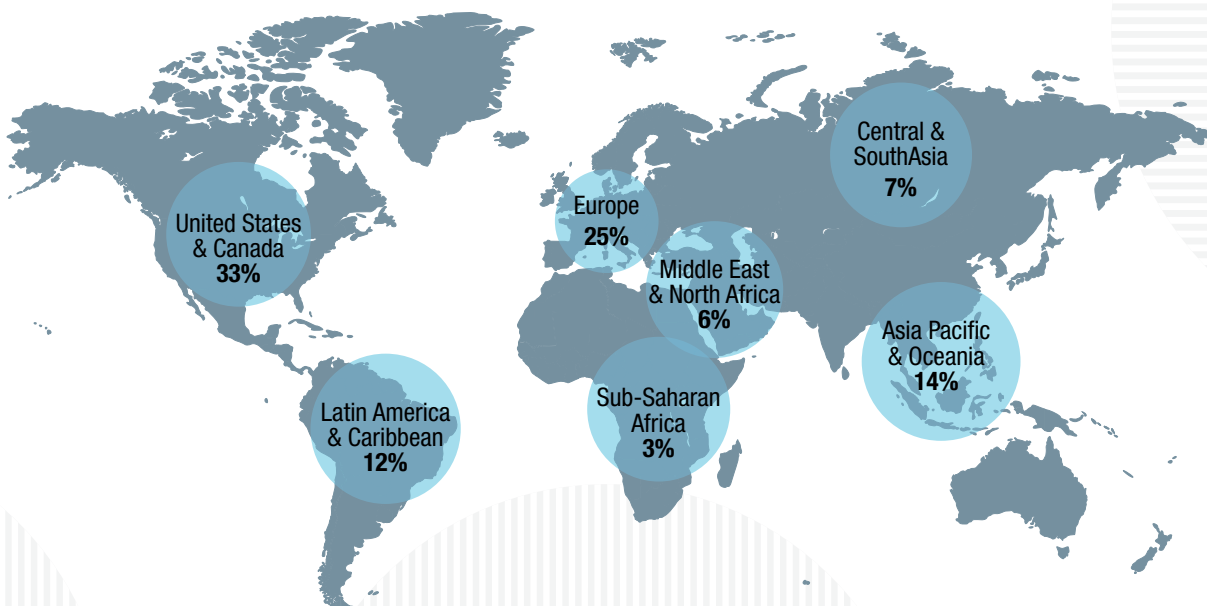


BY POLICY (TOP 5)*

Bullying & Harassment	22.3%
Adult Nudity & Sexual Activity	17.7%
Violence & Incitement	17.4%
Hateful Conduct (previously Hate Speech)	16.2%
Restricted Goods & Services	7.4%

*Based only on user-generated appeals to restore content.

Share of Cases Submitted to the Board by User-Selected Region in 2024





www.oversightboard.com

© 2025 Oversight Board LLC