Global Media Journal
Journal for International and Transcultural Communication
German Edition

BerlinUP Journals

Special Section: Norms, Power Relations and Injustices in Digitality
Peer-Reviewed Original Article

Genderwashing in Digital Platforms' Self-Regulation: A Case Study of Decisions by the Meta Oversight Board

Mariana Magalhães Avelar & Luana Mathias Souto

Abstract: Online gender-based violence curbs women's rights while generating profit for digital platforms due to the high engagement generated by such content. This dynamic exemplifies the gender-related legal issues raised by digital platform business models. These harmful practices are scrutinised by legal and social science scholarship and addressed within a broad regulatory discourse encompassing both soft and hard law measures and social media self-regulating initiatives. The latter is especially representative of the development of terms and conditions and the creation of "quasi-judicial" boards, which operate as regulatory intermediaries. One such body is Meta's Oversight Board—an independent board that reviews Meta's decisions about Facebook, Instagram, and Threads content. This research analyses two gender-related binding decisions by the Board to assess: (i) the legal values, norms, and principles applied in gender-sensitive rulings and (ii) whether these decisions led to effective protection, remediation, and infrastructural changes within Meta's services. The article argues that Meta's responses remain largely superficial, using human rights discourse to legitimise actions without addressing the core issue—its infrastructure. These measures often serve as symbolic gestures rather than substantive reforms and may amount to forms of *genderwashing*, ultimately exacerbating rather than mitigating harms experienced by women, girls, and LGBTQIA+ users online.

Keywords: *genderwashing*, digital platforms, self-regulation, Meta Oversight Board, gender equality.

Author information:
Mariana Magalhães Avelar holds a PhD and a master's degree in law from the Federal University of Minas Gerais (Brazil). She is a theoretical and interdisciplinary legal researcher, with a focus on administrative law, as well as business and human rights. Since June 2023, she has also been a visiting researcher at the Max Planck Institute for Comparative Public Law and International Law in Heidelberg, Germany.
Email: mmagalhaesavelar@gmail.com

Global Media Journal – German Edition
Vol. 15 No. 2 2025
ISSN 2196-4807
DOI: 10.60678/gmj-de.v15i2.335

Luana Mathias Souto is a Postdoctoral Researcher at the Gender and ICT Research Group, Universitat Oberta de Catalunya (Spain). She is the Principal Investigator of the project Reproductive Health under Algorithm Surveillance (THELMA), for which she was awarded a Marie Skłodowska-Curie (MSCA) Postdoctoral Fellowship by Horizon Europe in the 2023 call. She holds doctoral and master's degrees in Law from Pontifícia Universidade Católica de Minas Gerais – PUC Minas (Brazil). She is a theoretical and interdisciplinary legal researcher focusing on Legal Theory, Law and Literature, Human Rights, Democracy, and Gender Studies. She has also served as a visiting Postdoctoral Researcher at the Max Planck Institute in Frankfurt (MPILHLT in 2023) and Hamburg (MPIPriv in 2024), and as a Research Fellow at the Weizenbaum Institute in Berlin (2024).
ORCID: 0000-0002-6961-0187
Email: luana.mathias.souto@gmail.com

Funded by the European Union under the Agreement n# 101149321. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Horizon Europe 2021-2027/Marie Skłodowska-Curie Actions. Neither the European Union nor the granting authority can be held responsible for them.

The authors would like to thank Lucas Muniz Da Conceição for the thoughtful reading of the early version of this manuscript and for his bright suggestions.

To cite this article: Avelar, Mariana Magalhães & Souto, Luana Mathias (2025). *Genderwashing* in Digital Platforms' Self-Regulation: A Case Study of Decisions by the Meta Oversight Board *Global Media Journal – German Edition*, *15*(2), DOI: 10.60678/gmj-de.v15i2.335

The Meta Oversight Board (MOB) is an independent body that reviews content decisions made by Meta Platforms, Inc., across Facebook, Instagram, and Threads. It has the authority to overturn Meta's decisions with the stated goal of improving how the company treats people and communities globally. The Board can also apply warning screens to certain content, issue policy recommendations, and publish advisory opinions to guide Meta on specific issues. During the appeals process, it may receive public comments from any interested individuals and organisations[1], and decisions are made by a majority vote of a five-member panel.

Considering this background, this paper will examine two gender-related cases reviewed by the Meta Oversight Board: (1) "Violence against Women" (2023-002-IG-UA, 2023-005-IG-UA) and (2) "Image of Gender-Based Violence" (2023-006-FB-UA). The analysis aims to evaluate (i) the legal values, norms, and principles applied in gender-sensitive decisions and (ii) whether these decisions led to obligations for effective protection, remediation, and infrastructural changes within Meta's platforms. These cases were selected due to their procedural status as standard cases and their explicit focus on gender.

Based on this analysis, we argue that Meta's responses to gender-related risks remain largely superficial. While the Board invokes human rights language to legitimise its actions, Meta's implementation of the Board recommendation fails to address the root cause: the design of its infrastructure. This practice, referred to as *genderwashing*, allows platforms like Meta to claim progress in gender equality while avoiding meaningful accountability. Self-regulatory mechanisms such as the Oversight Board may serve more as mere smokescreens to draw the focus away from persistent gendered harmful practices that benefit the platform's economic interests.

The paper is organised into three parts. It first presents the MOB's decisions on gender by analysing two gender-related cases. It then examines how the Board's human rights rhetoric functions as a liberal legitimation mechanism that ultimately fails to strengthen the infrastructural dimensions of content moderation. Finally, it

[1] The MOB Rulebook for case review and policy guidance clarifies that "the Board may request public comment briefs. Calls for public comment briefs will be posted on the Board's website with requirements for form and substance, and a deadline for submission. Submissions will be shared with the panel" (Meta Oversight Board, 2024a, p.8). As clarified by the FAQ of the public comments' portal, public comments must respond to issues identified by the MOB and "as cases are assigned to panels, the Oversight Board will post a brief, anonymized description of the cases under review on the Oversight Board website. For up to 14 days after this posting, individuals and organization have an opportunity to share their insights. For expedited cases, this timeframe will be shortened(...) Anyone submitting comments will have an opportunity to provide consent to the Oversight Board to publish or attribute their comments publicly, as well as allow the Oversight Board to follow up with them regarding the content of their comments. The Oversight Board expects to publish comments in an appendix to each case decision, provided the comments meet the guidelines and the commenter has consented to the publication. (...)The panels reviewing cases will consider submissions at their sole discretion, and do not expect to be able to consider every submission in their deliberations" (Meta Oversight Board, 2025c).

introduces the concept of *genderwashing* and explains how this strategy aligns with Meta's business model concerning gender equality.

A Gender-Based Case Law Analysis of the Meta Oversight Board (MOB)

Social media platforms are considered an essential part of the digital public sphere, constituting an infrastructure resource which can shape public discourse through their self-regulatory measures and structures, such as terms of service, guidelines for algorithmic content moderation, and regulatory intermediaries (Kira, 2025), such as Social Media Councils (Fertmann & Kettemann, 2022). The social media business model produces a specific social order and structure, while the creation of regulatory intermediaries, such as MOB, is part of hybrid forms of governance that recognise and incorporate non-state and non-corporate actors into platform decision-making.

The Meta Oversight Board (MOB) emerged as a strategy to improve Facebook's content moderation governance following the controversies surrounding the 2016 United States presidential election (Muniz Da Conceição, 2024). Although it shares certain similarities with other platform governance councils—operating "below the threshold of illegality (or even below private rule violation), functioning as a firm-level self-regulatory body" (Fertmann et al., 2022, para. 4)—the MOB distinguishes itself through its institutional architecture. Notably, Kira (2025) notes that the Board is endowed with substantial resources necessary for executing its functions and is composed of high-profile, qualified members.

According to section 8 of the MOB Charter (Meta Oversight Board, 2025b), for the initial establishment of the board, "Meta selected a group of co-chairs" and "the co-chairs and Meta then jointly selected candidates for the remainder of the initial board seats." (p. 6) Thereafter, a committee of the board is tasked with "select[ing] candidates to serve as board members based on a review of the candidate's qualifications." (p. 6) In addition, both Meta and the public may propose candidates to the board. However, it remains unclear whether these nominations have a tangible effect on the selection of new board members. The selection process thus shapes the board's overall composition, which reflects both diversity goals and observable imbalances.

The MOB Rulebook for case review and policy guidance (Meta Oversight Board, 2024a) clarifies that the assignment of board members to specific cases varies according to the type of procedure (standard or expedited procedure[2]

[2] According to section 2.1 of the MOB Bylaws (Meta Oversight Board, 2024b), "Meta will have the ability to refer cases to the board, and may include a request that the board review a case on an expedited basis" (p.22). Regarding Meta-referred cases, they "will generally be subject to standard procedures, but in exceptional situations, Meta may require they be subject to expedited procedures" (Meta Oversight Board, 2024a, p.3). On standard cases, "the Oversight Board will aim to issue a decision within 90 days of receiving a user-generated appeal or a Meta-referred case" (Meta Oversight

or summary decisions[3]).

To standard procedures, such as the ones examined in this article, the MOB convene *ad hoc* panels composed of five board members. The members are constituted by the Case Management Tool (CMT), a Meta-developed platform that the MOB uses to receive and review case submissions and store case files. CMT could be seen as a functional equivalent to judicial case management software[4], but only along the ruler dimension, since the system is a Meta proprietary software, and its access is not open to the general public.

In any case, the MOB Rulebook clarifies that "[f]ollowing case selection, the CMT will constitute a panel with five Board Members, including at least one Member from the region the content primarily affects and ensuring gender diversity. Panelists will be notified of the assignment and invited to review the case file" (Meta Oversight Board, 2024a, p.8).

The decision drafted by the panel will be circulated to all Board Members for review, during which members may propose textual or substantive revisions; however, MOB plenary deliberation, for the decision drafting process itself, is exceptional and only occurs in case a “majority of the panel determine that a case requires plenary Board deliberation due to its significance or complexity” (Meta Oversight Board MOB, 2024a, p.10).

After incorporating board member suggestions, the panel will circulate the final decision draft to all Board Members, for review and approval. It is possible that MOB decides, by majority vote, to reassign the case to a new panel for reconsideration and, in such cases, a new panel will be convened promptly (Meta Oversight Board, 2024a, p.10).

Board, 2024a, p. 4), while for expedited cases, the process must be completed within 30 days of the case reaching the Board.
The expedited review, according to section 2.1.2 of the MOB Bylaws occurs (i) when Meta asks the board to act under expedited review due to exceptional circumstances, for example when content could result in urgent real-world consequences and the co-chairs will decide to accept Meta’s referral and (ii) independent of any Meta referral, when the co-chairs, with the consent of Meta, decide that a selected case will be reviewed under the expedited process (Meta Oversight Board, 2024b, p. 22).

[3]According to section 2.1.3 of the MOB Bylaws (Meta Oversight Board, 2024b), “Where Meta determines that the content in a particular case was incorrectly actioned by Meta and reverses its original decision, the case selection committee may choose to select the case for summary decision. The procedure for summary decisions will be determined by the co-chairs. Summary decisions will be voted on by the panel without a full-board vote” (p.24).

[4] Such software can be used for case initiation, structure docket workflows and automate the distribution of incoming matters as well as for document handling and publication of decisions. The Brazilian national electronic process platform (Processo Judicial Eletrônico – Pje), for example, allows auditable case tracking, while also implementing automated case assignment to promote impartiality and balance judicial workloads.
When auditable, these platforms can institutionalise standardised procedural milestones, maintain authoritative digital records, and employ configurable case-allocation and scheduling logic to reduce administrative error and reinforce procedural fairness.

As of December 2025, the MOB had 21 members. Board members represent a wide range of cultural and professional backgrounds, including academics, legal experts, policymakers, and journalists from around the world, speaking various languages. According to Meta, they are selected to reflect the diversity of users across Facebook, Instagram, and Threads and to bring their unique perspectives to content moderation on its platforms (Meta Oversight Board, n.d.-a). The MOB emphasises geographic balance among its members, as noted in item 1.4.1 of the MOB bylaws (Meta Oversight Board, 2024b, p. 10-11). Nonetheless, there is a predominance of members with a legal background (13 out of 21) and from the United States (6 out of 21). Countries with strong media law regulations and content moderation traditions, such as Germany, are not represented in the MOB, and only one member has a background in electoral matters, one of the most sensitive content moderation topics.

Section 3.1.4 of the MOB Bylaws states that MOB panels may, at their discretion and prior to deliberation, "request and receive information from a global pool of outside subject-matter experts, including academics, linguists, and researchers on a specific issue (e.g., region, cultural norm, or phrase)" (Meta Oversight Board, 2024b, p. 16). The Bylaws emphasise that "[t]his pool of experts will be populated at the discretion and/or direction of the board" (p. 16). In addition to this pool of experts, the Board can decide to procure research and related services and also "request issue briefs from advocacy or public interest organizations that reflect a range of perspectives" (p. 16). No public information exists concerning the processes for procurement or expert selection, including whether they undergo eligibility and integrity assessments, and there is no evidence of any structure to prevent or address conflicts of interest. Despite its quasi-judicial function, MOB's governance demonstrates a lack of fundamental transparency concerning its deliberative frameworks.

Also, despite the MOB's globalist intentions, platform councils such as the MOB do not function as neutral institutions; their decision-making processes are susceptible to particular biases stemming from the composition and perspectives of their members. While these bodies can enhance the accountability of content moderation policies, they also present notable limitations. Kettemann and Schulz (2023) highlight potential trade-offs of social media councils such as the MOB, noting that they can result in the "weakening of state regulators, confusion of responsibility, ethics-washing, a normative fig-leaf effect, and a globalist approach to speech rules that is non-responsive to local and regional practices" (Kettemann & Schulz, 2023, para. 6)—considerations that are critical to the gender issues assessed herein. Recognising these potential shortcomings is integral to evaluating the overall efficacy of such oversight mechanisms within the digital regulatory landscape.

Operating as a "quasi-judicial" body, the MOB primarily focuses on adjudicating individual cases of content moderation. Its case-by-case decision-making process, characterised largely by binary determinations—either removing or retaining content—raises critical questions about its capacity to address systemic issues or implement comprehensive policy reforms (Douek, 2022). Howard and Kira (2024) argue

that the reasoning underpinning MOB decisions could be refined to allow for more proportionate measures, such as content demotion.

Given that the MOB's funding derives from a trust established by Meta, some authors criticise its lack of financial independence (Douek, 2024) and note that the Board "lacks the power to make significant and lasting change where it matters most: in the actual content moderation guidelines" (Morar, 2019, para. 13).

The authority and transformative impact of the MOB were directly challenged by Meta's CEO's decision to remove moderation of harmful gender-related speech from the platform's policies in January 2025—just days before Donald Trump assumed office for a second term as President of the United States (Duffy, 2025). On one hand, gender issues remained one of the Board's seven strategic priorities. This focus stems from MOB's consideration of user appeals and includes concerns raised by women, non-binary, and trans individuals who face barriers to exercising their right to freedom of expression online, particularly in the context of gender-based violence and harassment (Meta Oversight Board, n.d.-b). On the other hand, Joel Kaplan (2025), Meta's Chief Global Affairs Officer, announced that the company is "getting rid of a number of restrictions on topics like immigration, gender identity and gender that are the subject of frequent political discourse and debate".

Until 20 December 2024, the MOB had decided on 18 cases classified under the "sex and gender equality" category, divided into two different procedures: standard decisions and summary decisions (10 standard decisions and 8 summary decisions). For the purposes of this study, however, we focus on two gender-related cases decided by the Oversight Board: (1) *Violence against Women* (2023-002-IG-UA, 2023-005-IG-UA) and (2) *Image of Gender-Based Violence* (2023-006-FB-UA). These cases were selected based on their procedure status as standard decisions and their classification under the "sex and gender equality" category in the MOB's decision repository.

Cases requiring an intersectional analysis—such as those involving queer and transgender rights—were excluded from this study due to the need for more extensive research. Similarly, summary decisions were not considered for the case analysis, as they are approved by a panel rather than by a full-board vote, do not involve public comments, and lack precedential value.

The two selected cases approach gender issues from distinct perspectives but ultimately highlight a shared concern: Meta's difficulty in providing appropriate gender-sensitive interpretations and responses within its digital platform infrastructure. In the first case, Meta misinterpreted a woman's post about domestic violence as a violation of its hate speech policy, leading to the unjustified removal of the content. In contrast, the second case involved a post that explicitly depicted gender-based violence, which Meta failed to recognise as a policy violation. Despite receiving three requests for removal, the company did not take action.

In summary, the first case illustrates how Meta wrongfully removed content that did not violate its policies, while the second case reveals Meta's repeated failure to enforce its own rules by allowing harmful content to remain online. This inconsistency in the application of its policies across different gender-sensitive contexts is precisely why these two cases were selected for analysis.

In addition to the two selected cases, at least three other decisions relevant to women's rights were submitted to the Board by December 2024[5] and merit future analysis. However, due to methodological constraints, this paper limits its scope to the two aforementioned standard cases.

The Board's focus on gender issues stems from an analysis of user appeals and includes concerns raised by women, non-binary, and trans individuals who face barriers to exercising their right to freedom of expression online, particularly in the context of gender-based violence and harassment. Nevertheless, as this paper will demonstrate, there is a lack of consistency in how Meta labels cases under the "sex and gender equality" policies and topics category[6]. The conflation of sex and gender reflects a lack of gender-sensitive criteria in the evaluation of MOB precedents, undermining the clarity and effectiveness of their classification and subsequent decisions.

Case 1: "Violence against Women" (2023-002-IG-UA, 2023-005-IG-UA)

In 2023, the Oversight Board decided to overturn two of Meta's removal decisions on Instagram posts made by Swedish users who reported gender-based violence. Meta's decisions were based on the company's Hate Speech Community Standard (Meta Transparency Centre, n.d.-a), which considers as hate speech the videos shared by a "woman describing her experience in a violent intimate relationship, including how she felt unable to discuss the situation with her family" (Meta Oversight Board, n.d.-c) (first post), and her reflections as a man-hater, noting that she does "not hate all men. She also states that she is a man-hater for condemning misogyny and that hating men is rooted in fear of violence" (Meta Oversight Board, n.d.-c) (second post).

After Meta's removal, the user submitted an appeal to the Oversight Board claiming:

[5] The other pertinent cases are (1) "Call for Women's Protest in Cuba" (2023-014-IG-UA); (2) "United States Posts Discussing Abortion" (2023-011-IG-UA, 2023-012-FB-UA, 2023-013-FB-UA); (3) "Explicit AI Images of Female Public Figures" (2024-007-IG-UA, 2024-008-FB-UA).

[6] The categorisation is inherently contentious and excludes significant gender-related cases, such as The Breast Cancer Symptoms and Nudity (2020-004-IG-UA), in which the Board reversed Facebook's decision to remove a breast cancer awareness post. The Board provided recommendations about Meta's use and regulation of algorithmic content moderation. In a more recent summary case (Breast Cancer Awareness -FB-HG46TXVV, recommendation no. 5), the MOB recommended the implementation of "an internal audit procedure to continuously analyse a statistically representative sample of automated content removal decisions to reverse and learn from enforcement mistakes" (Meta Oversight Board, 2025a).

> … they wanted to show women who face domestic violence that they are not alone. They also stated that removing the post stops an important discussion and keeps people from learning, and possibly sharing the post. In their second appeal, they explained that it was clear that they do not hate all men but want to discuss the problem of men committing violence against women (Meta Oversight Board, 2023c).

According to the user, the first post mentioned the International Day for the Elimination of Violence against Women. In Meta's submission to the Board, they recognised, regarding the first post, that it had been removed in error and did not violate the hate speech policy. However, they interpreted the second post as a violation of the company's hate speech policy, which defines hate speech as a direct attack against people based on protected characteristics, including sex and gender. They understood that her "statement that she does not hate all men does not impact the assessment of other parts of the post" (Meta Oversight Board, 2023c).

During the debates, the majority of Board members identified, based on Meta's understanding of non-violation of their Hate Speech policy, that the first post's "statement is also better understood as assurance to other victims of domestic violence that they are not alone." It is therefore a non-violating qualified statement" (Meta Oversight Board, 2023c). However, for other Board members, "the user posted a clearly unqualified behavioural statement that men 'murder, rape and abuse' women 'all the time, every day'" (Meta Oversight Board, 2023c).

Regarding the second post, the Board recognises its complexity and acknowledges that it does not violate Meta's hate speech policy. According to the Board's majority, the post must be assessed entirely, including the "user's analogy to the fear of venomous snakes" (Meta Oversight Board, 2023c), in which she points out that "the fear of venomous snakes brushes off onto all snakes, causing many or most humans to be frightened of snakes as a class" (Meta Oversight Board, 2023c). Therefore, "this does not mean she hates all men and describes man-hating as being defined by discussing fear and condemning violence against women" (Meta Oversight Board, 2023c). For this reason, the second post is not an expression of contempt that violates Meta's policy. Again, some Board members dissented, arguing that the second post's language "could lead to negative unintended consequences for both men and women" (Meta Oversight Board, 2023c); for this reason, it violates Meta's policy and should not be restored.

In conclusion, the majority of the Oversight Board decided that the posts did not violate Meta's Hate Speech policy and should be restored. Additionally, they stated a recommendation for Meta to include in their policy "a clearer exception to allow content that condemns or raises awareness of gender-based violence in the public language … as well as update its internal guidance so that moderators can effectively implement this exception" (Meta Oversight Board, 2023c).

In order to address Meta's human rights responsibilities in this case, the Board's decision was informed by the different international standards regarding the

freedom of opinion and expression[7], the prohibition of incitement to discrimination, hostility or violence[8] and, to a very limited extent, the right to non-discrimination[9]. In this case, besides 13 public comments submitted by the general public to the Board for consideration, the five Board members designated to form the panel[10] received the assistance by three independent experts: researchers from an institute at the University of Gothenburg, an advisory firm called *Duco Advisors*, and an organisation called *Memetica*. Meta's transparency website states that this was the company's "first bundled case about violence against women" (Meta Transparency Centre, 2023). The site also provides information regarding the status of the Board's recommendations. The first recommendation was to "include the exception for allowing content that condemns or raises awareness of gender-based violence in the public language of the Hate Speech policy" (Meta Transparency Centre, 2023) until June 2024, which had been partially implemented. The process of "refining and clarifying our Community Standards as part of holistically reviewing the overlaps and differences between our policies on organic and ads content …" (Meta Transparency Centre, 2023) is ongoing.

Regarding the second recommendation, which aimed "to ensure that content condemning and raising awareness of gender-based violence is not removed in error" (Meta Transparency Centre, 2023), Meta reported partial implementation or described it as work already undertaken, but did not publish evidence demonstrating full implementation. In this context, the Board awaited updates to the Hate Speech Community Standards, which aimed to provide "additional nuance to clarify internal guidance around our approach to behavioural statements, generalisations, and qualified behavioural statements" (Meta Transparency Centre, 2023).

The expected changes were never fully realised. Instead, the Hate Speech policy was dismantled and replaced with a "Hateful Conduct" policy, which expressly allows the use of "sex- or gender-exclusive language when discussing access to spaces often

[7] The decision explicitly references: Article 19, International Covenant on Civil and Political Rights (ICCPR), General Comment No. 34, Human Rights Committee, 2011; UN Human Rights Council resolution on freedom of expression and women's empowerment A/HRC/Res/23/2 (2013); Special Rapporteur on freedom of opinion and expression, reports: A/76/258 (2021), A/74/486 (2019), A/HRC/38/35 (2018), A/68/362 (2013); and Joint Declaration on Freedom of Expression and Gender Justice, Special Rapporteurs on freedom of opinion and expression of The United Nations (UN), the Organization for Security and Co-operation in Europe (OSCE), the Organization of American States (OAS), the African Commission on Human and Peoples' Rights (ACHPR) (2022).

[8] Article 20, para. 2, ICCPR; Rabat Plan of Action, UN High Commissioner for Human Rights report: A/HRC/22/17/Add.4 (2013).

[9] The decision cites solely the right to non-discrimination as articulated in Article 2, paragraph 1, and Article 26 of the ICCPR. The Convention on the Elimination of All Forms of Discrimination against Women (CEDAW) and United Nations General Assembly Resolution A/RES/54/134, which established the International Day for the Elimination of Violence against Women (date that was on the background of the moderated content), are not referenced in the Board decision.

[10] Standard procedures are always assigned to *ad hoc* panels. Meta developed the Case Management Tool (CMT), which the Meta Oversight Board (MOB) uses to receive and review case submissions, as well as to collect and store case files. According to the MOB Rulebook, once a case is selected, the CMT will form a panel of five Board Members. This panel must include at least one Member from the region most affected by the content and must ensure gender diversity. Panelists are then notified of their assignment and invited to review the case file (Meta Oversight Board, 2024a, p. 8).

limited by sex or gender, such as access to bathrooms, specific schools, specific military, law enforcement, or teaching roles, and health or support groups," or even permits calls "for exclusion or use of insulting language in the context of discussing political or religious topics, such as when discussing transgender rights, immigration, or homosexuality," as well as "cursing at a gender in the context of a romantic break-up" (Meta Transparency Centre, n.d.).
So far, the only recommendation fully implemented is number three: "to improve the accuracy of decisions made upon secondary review … by sending [them] to different reviewers than those who previously assessed the content" (Meta Transparency Centre, 2023). According to Meta:

> [We] have an internal system called Dynamic Multi Review (DMR) which enables us to review certain content multiple times by different reviewers before making a final decision. This ensures that the quality and accuracy of human review are carefully considered upon secondary review, taking into account factors such as virality and potential for harm (Meta Transparency Centre, 2023).

Finally, recommendation four, "to provide greater transparency to users and allow them to understand the consequences of their actions" (Meta Transparency Centre, 2023), was omitted or reframed by Meta without further context.

Case 2: "Image of Gender-Based Violence" (2023-006-FB-UA)

The case analyses a Facebook post joking about gender-based violence (GBV) in Iraq although the MOB wrongly labels it as taking place in Eritrea (Meta Oversight Board, 2023d). The post depicted a woman who suffered domestic violence as a result of a typographical error. According to the post's caption, the husband physically beat his wife, though the woman had asked him to bring a "donkey" or called him a donkey, when in fact she was requesting a "veil." Understanding this mockery requires knowledge of Arabic, as the words for "donkey" and "veil" are visually similar ("حمار" and "خمار"). The post implies that the husband physically assaulted her due to the misunderstanding caused by the typographical error. The woman depicted is identifiable: she is a well-known Syrian activist whose image has previously circulated widely on social media. (Meta Oversight Board, 2023d).
A user reported the content three times for violating the Violence and Incitement Community Standard. Yet, the report was not subject to human review, and Meta automatically closed the denouncement. Later, the case was considered relevant "to explore Meta's policies and practices in moderating content describing and joking about gender-based violence and its impact on the rights of users on and off Meta's platforms" (Meta Oversight Board, 2023b).

In its call for public comments, the Board highlighted as the main point of discussion: (i) Meta's policy and enforcement choices about content joking about or mocking gender-based violence; (ii) The relationship between Facebook and Instagram content that jokes about or mocks gender-based violence and its effect on people who may be impacted by this content and their ability to use these platforms; (iii)

The relationship between Facebook and Instagram content that jokes about or mocks gender-based violence and its effect on off-platform gender-based violence; (iv) How depictions of gender-based violence may be used to target public figures, human rights defenders, and activists; (v) Insights into the socio-political context in Iraq (and the region), regarding gender-based violence and its depiction on social media (Meta Oversight Board, 2023b).

As in the previously described case, the five Board members received assistance from independent experts: researchers from an institute[11] at the University of Gothenburg, Sweden; an advisory firm called *Duco Advisors*; an organization named *Memetica*; and linguistic expertise provided by *Lionbridge Technologies, LLC*. Additionally, 19 public comments were submitted by the participants to the Board during the public participation process. In summary, the Digital Rights Foundation's contribution stated that Meta should consider the content of the satire—especially when it has artistic content—as well as the potential harm it may cause. Similarly, *Fundacion Karisma* emphasised that freedom of expression should be the default principle and suggested analysing the context, creating resources for victims of harassment and bullying, and establishing appropriate moderation techniques. The public comment from Lawyers for Justice in Libya, on the other hand, highlighted the broader context of violence against women in the region and argued for the need to assess the vulnerabilities involved in order to understand the significance of linguistic and cultural abuses. Furthermore, their commentary affirmed the importance of placing the victim at the centre of any considerations.

The Oversight Board found that the post should be removed because it violated Meta's policy on bullying and harassment by mocking the injuries of a woman who had suffered gender-based violence and recommended that a more precise moderation mechanism should be created by the establishment of "a policy aimed at addressing content that normalises gender-based violence through praise, justification, celebration or mocking of gender-based violence" (Meta Oversight Board, 2023b). In addition, the Board recommended improving the Bullying and Harassment Community Standard to clarify that the term "medical condition" includes posts which mock serious physical injury.

The Board made two different recommendations on its decisions, regarding opportunities to improve Meta's content policy. The first recommendation was that Meta should improve the clarity of its policies for users, by explaining that the term "medical condition," as used in the Bullying and Harassment Community Standard, includes "serious physical injury. Regarding this point, Meta indicated that the company was "in the process of refining our guidance around the term "medical condition" within our Bullying and Harassment Community Standards and will share our definition for "medical condition" in a planned update to our Community Standards (Meta Oversight Board, 2023d). The MOB recommendation tracker (Meta

[11] Meta does not disclose or specify information about the institute.

Oversight Board, n.d.-c) signalled that “progress [was] reported”, without further clarification.

The second recommendation involves the development of a policy aimed at addressing content that normalizes gender-based violence through praise, justification, celebration or mocking of gender-based violence. Meta affirmed, on the past, that the company was “making final changes to our policy as a result of research, external engagement, and internal discussions, which include refinements to our approach to closing unintentional gaps in our policies covering content that may praise gender-based violence” (Meta Oversight Board, 2023d)). The MOB recommendation tracker (Meta Oversight Board, n.d.-c), nonetheless, indicates that Meta reported implementation or described as work Meta already does but did not publish information to demonstrate implementation.

Besides, the 2023 Oversight Board Annual Report hinted at a potential change by stating that in 2024, the Board will “explore how our recommendations can identify and mitigate systemic risks created by upstream product design choices and the automated treatment of content online.” (Meta Oversight Board, 2024c). Notwithstanding the Board's efforts and multiple recommendations, Meta's leadership has shown a contrary stance, as previously noted in the "Violence against Women" case.

MOB Gender Narratives and the Application of International Human Rights as a Legitimation Strategy

The aforementioned cases exemplify the presence of misogynistic public discourse and a variant of unlawful freedom of expression restriction, as it imposes “very high costs to speaking out and produces concrete subordination and exclusion from accessing and controlling the means of production and socio-economic participation in techno-capitalism” (Valente, 2022, p.111-112).

In light of the systemic nature of online abuse directed at women and the push for gender-sensitive governance on platforms, a common approach is to integrate gender considerations through the frameworks of human and fundamental rights. Some authors, such as Suzor (2020), claim that “human rights is [sic] probably the most powerful tool we have to encourage intermediaries and governments to make their governance processes more legitimate” (p. 3).

This position views human rights as framework that tends to safeguard a universally oriented set of values that can be understood and endorsed by both the private and public sectors and can be applied to social media platforms through the influence of international human rights law. This includes international treaties, soft law norms, and industry benchmarks, such as the *United Nations Guiding Principles on*

Business and Human Rights (UNGP), which are expressly mentioned on Meta's Corporate Human Rights Policy[12].

Nonetheless, as asserted in its last annual report, the Board acknowledges the challenges of adapting this logic to the specificities of online harm:

> As a board, we now have more than three years of practical experience of applying international human rights standards to a private company moderating the content of billions of people around the world. These standards have many benefits. They place freedom of expression and human dignity at the center of our analysis, offer a cross cultural reference point and foster transparency, making our work part of an ecosystem of human rights stakeholders. However, there are also challenges. For example, the kinds of harms found online are also different to in the real world. In our decisions, we have expressed concern at how the scale and speed of online content can create cumulative harms that would not exist offline. Our experience as a board shows that an approach based on international human rights standards can be useful. Moving forward, we will continue to adapt our approach given the challenges above. This will be difficult, but immensely valuable, and we will rely on feedback from academics and civil society as we continue down this path (Meta Oversight Board, 2024c, p. 13).

The Board argued, in both cases hereby analysed, that Meta holds human rights responsibilities and allegedly assessed this responsibility through the lens of international standards. Yet, the analysis of the “Image of Gender-Based Violence” and “Violence Against Women” cases shows some points for improvement in the Board’s decision-making process and, above all, on the enforcement of its decisions by Meta, including (i) the need for in-depth local consideration to assess Meta’s human rights responsibilities; (ii) more detailed guidance on the decision’s enforcement; and (iii) an impact assessment of the infrastructural responses that are needed to achieve responsive and gender-informed content moderation activity.

The first two points (the need for in-depth local consideration and more detailed enforcement guidance) are illustrated in the "Image of Gender-Based Violence" case. In this decision, the Oversight Board inaccurately tagged Eritrea as the affected region, even though the posts originated from Iraq and the woman depicted was Syrian. This apparent geographical error or negligence complicates a normative analysis of the Board's decision. Had the case in fact involved women in Eritrea, the

[12] Meta’s Human Rights Policy clarifies that the company is committed “to respecting human rights as set out in the United Nations Guiding Principles on Business and Human Rights (UNGPs). This commitment encompasses internationally recognized human rights as defined by the International Bill of Human Rights—which consists of the Universal Declaration of Human Rights; the International Covenant on Civil and Political Rights; and the International Covenant on Economic, Social and Cultural Rights—as well as the International Labour Organization Declaration on Fundamental Principles and Rights at Work. Depending on circumstances, we also utilize other widely accepted international human rights instruments, including the International Convention on the Elimination of All Forms of Racial Discrimination; the Convention on the Elimination of All Forms of Discrimination Against Women; the Convention on the Rights of the Child; the Convention on the Rights of Persons with Disabilities; the Charter of Fundamental Rights of the European Union; and the American Convention on Human Rights. We specifically recognize that the universal obligation of non-discrimination is a necessary—but not sufficient—condition for real, lived, equality. We are committed to implementing the Global Network Initiative (GNI) Principles on Freedom of Expression and Privacy, and their associated Implementation Guidelines” (Meta, n.d.).

Board would have been required to consider the applicability of the Protocol to the African Charter on Human and Peoples' Rights on the Rights of Women in Africa (the "Maputo Protocol", see African Union, 2003), given Eritrea's status as an African country, which signed but did not ratify the Protocol. In such scenario, Article 5(d) of the Maputo Protocol, which explicitly includes the "protection of women at risk of harmful practices or other forms of violence, abuse, and intolerance" (African Union, 2003) would have constituted a central interpretive reference point for the MOB's assessment.

This episode reveals a broader structural issue: the Oversight Board does not consistently demonstrate a substantive commitment to human rights protection when adjudicating cases involving Global South contexts. Its failure to engage with the geopolitical dimensions of the case, or to examine the relevant social, cultural, political, and economic specificities, weakens the legitimacy of its reasoning. Moreover, the Board's inattention to the regional human rights instruments that govern particular jurisdictions further compromises the analytical integrity of its decisions. Confusing Eritrea with Iraq is a major error. It is a consequential misclassification with the potential to materially affect the Board's legal and normative conclusions in this and further cases.

Furthermore, there are issues concerning the application of article 8 of United Nations Declaration on the Right and Responsibility of Individuals, Groups and Organs of Society to Promote and Protect Universally Recognized Human Rights and Fundamental Freedoms General Assembly Resolution adopted by the General Assembly A/RES/53/144. (1999, March 9), which encompasses the right to effective access, on a non-discriminatory basis, to participation in the conduct of public affairs. The Board expressly mentioned the use of this standard to assess Meta's human right responsibilities, highlighting that the case "offers the opportunity to explore how Meta's policies and enforcement address content that targets women human rights defenders and content that mocks gender-based violence, issues the Board is focusing on through its strategic priority of gender" (Meta Oversight Board, 2023d).

Nonetheless, the discussions about the specific gender violence suffered by a woman acting as human rights defender are not even mentioned in the recommendations made by the Board on its decision of the case. A "victim-centred" approach, that considered prevention and effective redress, would require a more nuanced decision-making process, by the MOB, and increase accountability of Meta on such matters. On one hand, the challenge is not easy to tackle since the horizontal application of fundamental rights (the *Drittwirkung* doctrine[13]), is, itself, deeply contested. For

[13] The horizontal application of fundamental rights, also known as *Drittwirkung* doctrine is a creation of the *Bundesverfassungsgericht,* the German Federal Constitutional Court (FCC). As synthetised by Matthias Kettemann and Torben Klausa (2020), "In this doctrine, the FCC argues that because Germany's constitution is not value-free, these values—embedded for example in fundamental rights—radiate into nonpublic fields of law and private law relationships, like contracts. Therefore, courts have to consider rights in their reading, potentially resulting in an indirect application of fundamental rights to private actors. There are, of course, limits to this. I cannot invoke my right to

some authors, the application of fundamental rights in the context of the platforms requires that "content-related standards need to be (and by now usually are) published, enshrined in terms of service that meet fundamental rights standards, and formulated as general rules that are applied non-arbitrarily and allow for effective recourse against deletions and suspensions" (Kettemann & Tiedeke, 2020, p. 12).

On the other hand, some argue that human-rights-based legal reasoning alone is insufficient to address the Board's lack of democratic legitimacy. In this context, Brenda Dvoskin (2023) observes that the Board approaches international human rights law as a set of exogenous and universal principles that can be translated and implemented on social media by human rights experts—an idea that is deeply contested. Dvoskin (2023) argues that, rather than concealing its authority behind a purported adherence to higher principles, the Board could enhance its legitimacy by fostering greater involvement of civil society actors. Even though Board decisions usually involve a process of public participation through the submission of comments, it remains unclear to what extent these contributions are incorporated into the Board's reasoning (Muniz Da Conceição, 2024), and how meaningful stakeholder involvement can be achieved through all the content moderation process. The follow-up to the implementation of MOB recommendations is, so far, solely made by the MOB itself, lacking independent oversight and accountability structures which could pinpoint systemic accountability issues.

Lucas Muniz Da Conceição's scholarship corroborates this conclusion: in recent publications, the author argues that the MOB fails to adequately connect with the political and social realities of Meta's users, leading to legitimacy issues by not fostering meaningful engagement with the diverse contexts in which the platform operates (Muniz Da Conceição, 2024; 2025) research highlights the tension between the MOB's "quasi-judicial" structure and its ability to address the complex, context-dependent nature of content moderation across diverse global communities. He suggests that the Board's current operational model may inadvertently perpetuate existing power imbalances in digital governance, prioritising certain regions or user groups over others. Through a combination of quantitative and qualitative analysis of MOB decisions and public comments, the author further posits that case selection may reflect regional disparities, favouring users from more lucrative markets or those more sensitive to regulation. This selectivity, as pointed out by Muniz Da Conceição (2024) potentially undermines the Board's claim to global representation and impartiality. The superficial preoccupation with human rights standards in content moderation discourse hinders the enhancement of regulatory intermediaries' reflexive capacity—a capacity that, according to Muniz Da Conceição (2024), can only be advanced through the "identification of community within the understanding of what constitutes it, thus politicising it beyond normative determinations" (p. 575).

physical integrity (Art. 2(2)(1) Basic Law) against my violin-practicing neighbor, even if they butcher Mozart. However, especially in systems, or "constellations," with (extremely) disparate power relationships, the FCC and other German courts use the *Drittwirkung* doctrine to balance opposing interests that would otherwise play out one-sidedly" (para.5).

These aspects were brought to the MOB's attention in the public comments of the "Violence against Women" case. On that occasion, at the Centre for Digital Citizens - Northumbria University, an important reflection could be made on a well-known issue in technology and gender: women's under-representation in STEM areas:

> The combination of lack of context in automated moderation, a technical workforce made largely of men, and a set of broad and opaque policies enables mistakes like the one this case addresses. More efforts towards gender diversity in hiring technical workers and investment initiatives to bring more women to the table are therefore greatly needed (Meta Oversight Board, 2023a, p.9).

A UNESCO study underscores this concern, stating: "Women are outnumbered in higher education and in career and leadership roles in STEM. Underrepresentation moves them to the margins, including among the decision-makers shaping STEM today and into the future." (Straza, 2024, p. 8). This reflects a structural gender inequality that undermines women's rights across the entire digital platform infrastructure. Scholars such as O'Neil (2016), Fry (2018) and Zou & Schiebinger (2018) have emphasized this gender gap as a key factor contributing to biased technological data, which can lead to misogynistic and sexist outcomes. For example, "when Google Translate converts news articles written in Spanish into English, phrases referring to women often become 'he said' or 'he wrote'." (Zou & Schiebinger, 2018).

Furthermore, when women are excluded from platform design, senior management roles in tech, and gender-related analyses—such as those relevant to the cases discussed here—the resulting decisions often overlook or misinterpret gender nuances (Rubio-Marín, 2022). For example, MOB overturned Meta's decision to remove a post in which women spoke about domestic violence ("Violence Against Women" case) determining that it was not a violation of Meta's hate speech policy and, therefore, should be reinstated. In another instance, in the "Image of Gender-Based Violence" case, Meta failed three times to acknowledge a user's report of a violation of the Violence and Incitement Community Standard, in a case of explicitly gender-based violence. These cases illustrate the urgent need for participative practices that ensure gender-sensitive interpretations and responses within the digital platform infrastructure, going beyond the mere rhetoric reference to human rights frameworks.

Lucas Muniz Da Conceição draws attention to the side effects of applying strictly normative frameworks—such as human rights discourse, the Rule of Law, and gender equality principles—to digital governance instruments like the MOB: for the author, the use of these paradigms may inadvertently constrain their capacity to imagine and implement innovative governance models by replicating the institutional logics of real-world courts within digital infrastructures (Muniz Da Conceição, 2024). The first constraint is based on the limited powers of the Board: the core of the transformative power of its decisions is concentrated in its recommendations, which are, according to item 2.3 of the MOB's bylaw, non-binding (Meta Oversight Board, 2024b).

In both cases hereby analysed, not only were the Board recommendations not fully implemented[14], but structural changes were made in the opposite direction, following Meta's recent review of policies and standards. As shown by a report of the European Observatory of Online Hate (2025), "after Meta changed its moderation policies in January 2025, average daily toxicity remained significantly elevated, stabilising at 30-40% higher than levels observed in November 2024".

As Valente (2022) argues, "the conditions for developing technologies and making them accountable—and sensitive to local contexts—are also connected to feminist ethics, and legislative efforts must incorporate that" (p. 118). The perspective developed by Valente (2022) underscores the importance of embedding feminist principles into the design and oversight of digital systems.

Following the "Violence Against Women" case, Meta could have implemented measures to prevent the erroneous removal of content that condemns and raises awareness of gender-based violence by updating the guidelines for its at-scale moderators, with particular emphasis on the regulations concerning permissible qualifications and speech. However, Meta failed to provide data showing the adoption of the MOB recommendation, which might significantly influence women's freedom of expression in a manner that the mere invocation of human rights terminology in Board decisions cannot secure.

To achieve legitimacy, digital governance systems must move beyond the mere transplantation of legal norms. They require a regulatory architecture specifically designed for the sociotechnical dynamics of online platforms. Addressing human rights violations in these spaces requires not only symbolic recognition but also effective mechanisms for remediation and prevention. Without these, legitimacy remains superficial, as confirmed by the cases analysed here.

Genderwashing: Creating an Artificially Pro-Gender Business Model

The term *genderwashing* is inspired by a similar business model strategy known as *greenwashing*. This term, which relates to environmental and climate change debates, dates back to 1986, when the activist Jay Westerveld "described the hypocrisy

[14] The issue is also identified in broader Board recommendations, such as the policy for cross-checking high-profile users who frequently posted harmful content that was not removed by automated content moderation procedures due to their accounts being assigned privileges of enhanced human oversight.
In a policy advisory opinion regarding the case, the Board addressed the issue through several recommendations, including that "Meta should publicly mark the pages and accounts of entities receiving list-based protection in the following categories: all state actors and political candidates, all business partners, all media actors, and all other public figures included because of the commercial benefit to the company in avoiding false positives. Other categories of users may opt to be identified" (Meta Oversight Board, n.d.-c). The application of these procedures might have produced an infrastructural impact on the protection of human rights on its platforms if Meta had not declined to embrace this Board recommendation.

of hotels portraying money-making towel reuse programs as examples of environmental stewardship while failing to make improvements in areas of greater environmental impact, such as waste recycling" (Pearson, 2010, p. 39). Similarly, *genderwashing* is "a hegemonic process that operates through discourses and texts to reinforce unequal power relations. This reinforcement is partially achieved through the co-optation of diversity and the use of organisational texts to create the myth of equality" (Fox-Kirk et al., 2020, p. 5), a process that can intersect with the so-called *bluewashing*, a practice that involves "the malpractice of making unsubstantiated or misleading claims about, or implementing superficial measures in favour of, the ethical values and benefits of digital processes, products, services, or other solutions in order to appear more digitally ethical than one is" (Floridi, 2019, p. 187).

As demonstrated, Meta's initiative to establish a Board to decide on content removal under the "sex and gender equality" category without substantively engaging in gender-sensitive debates is problematic. The issues are manifold and range from the inaccuracy of the "sex and gender" categorization itself to sophisticated practices that, together, form a business strategy aimed at giving the appearance of a progressive agenda on gender equality while avoiding meaningful debate about the company's actual commitment to these issues. By undermining MOB recommendations, Meta diverts consumer attention from important discussions about structural gender inequalities—such as the under-representation of women in the tech sector or the uncritical use of the label "sex and gender equality".

The cases examined in this paper are not isolated incidents. Media tech companies often make arbitrary decisions to silence specific debates in order to "create and enforce their own norm systems with limited or no accountability or, alternatively, [that] could be misused by the state" (Khosla et al., 2023, p. 2). These companies use human rights discourses—including women's rights —only to render them ineffective. By keeping their practices regarding gender issues opaque, they avoid public scrutiny and shield themselves from criticism about how they address structural gender inequalities. The calculated result of this strategy, however, extends beyond the company's or MOB's decisions. As Fox-Kirk et al. (2020) warn, "if genderwashing practices continue unchallenged, the danger is that the myth of equality will be perceived as the 'truth'. As a result, it may become more difficult for inequalities to surface and to be questioned" (p. 588).

Connecting this practice to Meta's recent decisions on diversity, equity, and inclusion[15] reveals the superficial nature of its gender equality efforts. Meta's deconstruction of its gender-informed policy framework, which aligns with illiberal and "autocratic genderwashing" practices adopted by governments worldwide. As Bjarnegård & Zetterberg (2022) note, "by taking credit for advances in gender equality, autocratic governments put the spotlight on an area that is widely seen as linked with democracy, while drawing the focus away from persistent authoritarian practices"

[15] With Trump's second term election, Meta terminated its diversity, equity and inclusion (DEI) programs (Walker, 2025).

(p. 61). This demonstrates the collaboration between governments and Big Tech in eroding gender rights for profit within an unregulated business environment (DiMeco, 2023).

The absence of meaningful due diligence and precautionary approaches has exposed wide accountability deficits in legal regimes to address these situations. Furthermore, recent evidence points to a thriving digital economy based on gender trolling and violations of human rights, disincentivising meaningful action on these issues by tech companies (Khosla et al., 2023, p. 3).

By diverting attention from structural debates about gender inequalities through *genderwashing* practices, Big Tech "shapes gender norms, unpicks how the technological design, profit models, and organisational hierarchy all give way to patriarchal norms, and in doing so, perpetuate sexist, heteronormative and racist stereotypes" (Khosla et al., 2023, p. 3). These practices undermine women's rights and set the stage for gender backlash campaigns that reverberate among governments, citizens, and private companies.

Conclusion

In conclusion, this analysis of the Meta Oversight Board's (MOB) handling of gender-related cases reveals significant shortcomings in Meta's approach to addressing gender-based violence and discrimination. Despite the Board's stated commitment to human rights and gender equality, its decisions often reflect a superficial engagement with these issues, and its recommendations were frequently neglected by Meta, resulting in symbolic gestures rather than substantive reforms. The concept of *genderwashing* provides a critical lens through which to interpret Meta's practices—where the company utilizes human rights rhetoric to project an image of progress while failing to address the underlying structural inequalities embedded within its platform and governance mechanisms.

The cases examined highlight the need for a more nuanced understanding of gender issues in content moderation, as well as the importance of incorporating diverse perspectives—particularly those of underrepresented groups in technology. While the MOB's reliance on a "quasi-judicial" model can remediate concrete problems, it does not adequately address the systemic biases inherent in Meta's infrastructure, nor the broader socio-political contexts in which these issues emerge.

To move forward, it is essential for Meta to engage in genuine dialogue with civil society, enhance the transparency of its decision-making processes, and implement comprehensive policy reforms that prioritise the protection of marginalized voices. Only through meaningful accountability and a sincere commitment to addressing the root causes of gender inequality can Meta hope to foster a safer and more equitable online environment. As the digital landscape continues to evolve, the

responsibility lies with platforms like Meta to ensure that their governance structures reflect a genuine dedication to gender equality—rather than merely functioning as a façade for profit-driven motives.

References

African Union. (2003, July 11). *Protocol to the African Charter on Human and Peoples' Rights on the Rights of Women in Africa* (Maputo Protocol) (UNTS Reg. No. 26363). https://treaties.un.org/doc/Publication/UNTS/No%20Volume/26363/A-26363-0800002805265c4.pdf

Bjarnegård, E., & Zetterberg, P. (2022). How Autocrats Weaponize Women's Rights. *Journal of Democracy, 33*(2), 60-75. https://www.journalofdemocracy.org/articles/how-autocrats-weaponize-womens-rights/

DiMeco, L. (2023, February 17). 'Gender trolling' is curbing women's rights – and making money for digital platforms. *The Guardian*. https://www.theguardian.com/global-development/2023/feb/17/gender-trolling-women-rights-money-digital-platforms-social-media-hate-politics

Douek E, (2022). Content Moderation as Systems Thinking. *Harvard Law Review, 136*(2), 528-607.

Douek, E. (2024). The Meta Oversight Board and the empty promise of legitimacy. *Harvard Journal of Law & Technology*, *37*(2) – 373-445. https://doi.org/10.2139/ssrn.4565180

Duffy, C. (2025, January 7). Meta is getting rid of fact checkers; Zuckerberg acknowledged more harmful content will appear on the platforms now. *CNN*. https://edition.cnn.com/2025/01/07/tech/meta-censorship-moderation

Dvoskin, B. (2023). Expertise and participation in the Facebook Oversight Board: From reason to will. *Telecommunications Policy*, *47*(5). https://www.sciencedirect.com/science/article/abs/pii/S0308596122001653

European Observatory of Online Hate. (2025). *How Meta's new content moderation policies affect gender-based violence*. https://eooh.eu/articles/meta/online/gender/based/violence/content/moderation

Fertmann, M., & Kettemann, M. (2022). Platform-proofing Democracy – Social Media Councils as Tools to increase the Public Accountability of Online Platforms. In M.C. Kettemann (Ed.), *How Platforms Respond to Human Rights Conflicts Online: Best Practices in Weighing Rights and Obligations in Hybrid Online Orders* (pp. 156-175). Verlag Hans-Bredow-Institut. https://doi.org/10.21241/ssoar.81873

Fertmann, M., Ganesh, B., Gorwa, R., & Neudert, L.-M. (2022, May 16). Hybrid institutions for disinformation governance: Between imaginative and imaginary. *Internet Policy Review*. https://policyreview.info/articles/news/hybrid-institutions-disinformation-governance-between-imaginative-and-imaginary/1669

Floridi, L. (2019). Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. *Philosophy & Technology, 32*, 185–193. https://doi.org/10.1007/s13347-019-00354-x

Fox-Kirk, W, Gardiner, R. A., Finn, H. & Chisholm, J. (2020). Genderwashing: the myth of equality. *Human Resource Development International*, *23*(5), 586–597. https://doi.org/10.1080/13678868.2020.1801065

Fry, H. (2018). *Hello World: Being Human in the Age of Algorithms*. W.W. Norton & Company

Howard, J. W., & Kira, B. (2024). Remove or reduce: Demotion, content moderation, and human rights. *SSRN*. https://doi.org/10.2139/ssrn.4891835

Kaplan, J. (2025, January 7). More Speech and Fewer Mistakes. Meta Newsroom. https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/

Kettemann, M. C., & Klausa, T. (2020). Regulating Online Speech: Ze German Way. Lawfare. https://www.lawfaremedia.org/article/regulating-online-speech-ze-german-way

Kettemann, M. C., & Schulz, W. (Eds.). (2023). *Platform://Democracy – Perspectives on Platform Power, Public Values and the Potential of Social Media Councils*. Verlag Hans-Bredow-Institut. https://graphite.page/platform-democracy-report/

Kettemann, M. C., & Tiedeke, A. S. (2020). Backup: Can users sue platforms to reinstate deleted content? *Internet Policy Review, 9*(2). https://doi.org/10.14763/2020.2.1484

Khosla, R., Mishra, V., & Singh, S. (2023). Sexual and reproductive health and rights and bodily autonomy in a digital world. *Sexual and Reproductive Health Matters, 31*(4), 1-5. https://doi.org/10.1080/26410397.2023.2269003

Kira, B. (2025) Regulatory intermediaries in content moderation. *Internet Policy Review, 14*(1), 1-26. https://policyreview.info/articles/analysis/regulatory-intermediaries-content-moderation

Meta. (n.d.). Corporate Human Rights Policy. https://humanrights.fb.com/policy/

Meta Oversight Board. (n.d.–a). Meet the Board – Get to know our board members https://www.oversightboard.com/meet-the-board/

Meta Oversight Board. (n.d.-b). Strategic priorities. https://www.oversightboard.com/strategic-priorities/

Meta Oversight Board. (n.d.-c). Recommendation tracker. https://www.oversightboard.com/explore-our-recommendation-tracker/

Meta Oversight Board. (2023a). Public Comment Appendix for 2023-002-IG-UA. https://www.oversightboard.com/wp-content/uploads/2024/02/Violence-against-women-public-comments-appendix-4.pdf

Meta Oversight Board. (2023b, April 27). Oversight Board Announces New Cases About Gender-Based Violence. https://www.oversightboard.com/news/3353287251600384-oversight-board-announces-new-cases-about-gender-based-violence/

Meta Oversight Board. (2023c, July 12). Violence against women. https://www.oversightboard.com/decision/ig-h3138h6s/

Meta Oversight Board. (2023d, August 1). Image of gender-based violence. https://www.oversightboard.com/decision/fb-1rwwjuat/

Meta Oversight Board. (2023e, August 1). Public Comment Appendix for 2023-006-FB-UA. https://www.oversightboard.com/wp-content/uploads/2025/12/295372496346189.pdf

Meta Oversight Board. (2024a, March). Rulebook for case review and Policy Guidance. https://www.oversightboard.com/wp-content/uploads/2024/03/OB_Rulebook_March_2024-1.pdf

Meta Oversight Board. (2024b) Oversight Board Bylaws. Improving how Meta treats people & communities around the world. https://www.oversightboard.com/wp-content/uploads/2024/03/Oversight-Board-Bylaws.pdf

Meta Oversight Board. (2024c, June 27). 2023 Annual Report Shows Board's Impact on Meta. https://www.oversightboard.com/news/2023-annual-report-shows-boards-impact-on-meta/

Meta Oversight Board. (2025a, May 15). Breast Cancer Awareness. https://www.oversightboard.com/decision/bun-0w49p93l/

Meta Oversight Board. (2025b). Oversight Board Charter. https://www.oversightboard.com/wp-content/uploads/2025/07/Oversight-Board-Charter-June-2025.pdf

Meta Oversight Board. (2025c). Public Comments Portal. https://www.oversightboard.com/public-comments-portal/

Meta Transparency Centre. (n.d.). Hateful Conduct. https://transparency.meta.com/en-gb/policies/community-standards/hateful-conduct/

Meta Transparency Centre. (2023, September 11). First Bundled Case About Violence Against Women. https://transparency.meta.com/pt-br/oversight/oversight-board-cases/sweden-violence-against-women/

Morar, D. (2019). Facebook's Oversight Board: A toothless Supreme Court? https://www.internetgovernance.org/2019/10/02/facebooks-oversight-board-a-judiciary-with-no-constitution/

Muniz Da Conceição, L.H. (2024). A constitutional reflector? Assessing societal and digital constitutionalism in Meta's Oversight Board. *Global Constitutionalism, 13*(3), 557–590. https://doi.org/10.1017/S2045381723000394

Muniz Da Conceição, L.H. (2025). The Quantum State of the Individual in Platform Governance: Digital Constitutionalism and Global Democratisation. *Information, Communication & Society*, 1-28. https://doi.org/10.1080/1369118X.2025.2492572

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Pearson, J. (2010). Are We Doing the Right Thing? Leadership and Prioritisation for Public Benefit. *Journal of Corporate Citizenship*, *37*, 37–40.

Rubio-Marín, R. (2022). *Global Gender Constitutionalism and Women's Citizenship: A Struggle for Transformative Inclusion*. Cambridge University Press. https://doi.org/10.1017/9781316819241

Straza, T. (2024). *Changing the equation: Securing STEM futures for women*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391384

Suzor, N. (2020). A constitutional moment: How we might reimagine platform governance. *Computer Law & Security Review, 36*. https://doi.org/10.1016/j.clsr.2019.105381

United Nations General Assembly. (1999, March 9). *Declaration on the right and responsibility of individuals, groups and organs of society to promote and protect universally recognized human rights and fundamental freedoms* (A/RES/53/144). https://docs.un.org/en/A/RES/53/144

Valente, M. (2022). No place for women: Gaps and challenges in promoting equality on social media. In E. Celeste, A. Heldt, & C. I. Keller (Eds.), *Constitutionalising social media* (pp. 101–118). Bloomsbury.

Walker, A. R., (2025, January 10). Meta terminates its DEI programs days before Trump inauguration. *The Guardian*. https://www.theguardian.com/us-news/2025/jan/10/meta-ending-dei-program

Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist — it's time to make it fair. *Nature*, *559*, 324-326. https://doi.org/10.1038/d41586-018-05707-8