



监督委员会

2022年

年度报告

目录

联合主席寄语	4
信托机构主席寄语	7
执行摘要	8
认识委员会	10
推出七项战略重点	12
委员会如何处理用户申诉	14
2022 年重要事件时间表	16

建议和影响

概述	19
从承诺到行动：为用户造福	20
影响时间表：让用户知道他们的错处	26
Meta 对委员会建议的落实情况	28

案件甄选

概述	31
提交给委员会的案件	32
案件甄选委员会考虑的案件	34

案件审理决定与政策咨询意见

概述	37
2022 年案件审理决定与政策咨询意见	39
决定和政策咨询意见摘要	40
将国际人权标准应用于内容审核：第 19 条	51
委员会决策流程中的国际人权标准	52

互动和公众评议

概述	55
2022 年互动活动时间线	58

展望未来

改进我们与 Meta 的合作	59
分享独立监督的好处	61
帮助企业适应新的监管环境	63

联合主席寄语



EVELYN ASWAD、CATALINA BOTERO-MARINO、MICHAEL MCCONNELL、HELLE THORNING-SCHMIDT - 监督委员会联合主席

2022 年，我们给 Meta 的许多建议落到了实处，帮助 Meta 更好地对待世界各地的人们和社群。我们的工作促使 Meta 复审内容审核政策，更清楚地阐明并更一致地适用政策规则。现在，Meta 在删除用户帖子的时候会说明帖子违反了政策的哪方面规定，并且 Meta 在进行内容审核时也比以往更为符合人权原则。

作为对我们建议的回应，Meta 推出了一项危机政策规程，以确保在面临危机情况时采取更为一致的应对措施；启动了对危险人物和组织政策的复审；针对错误信息制定了一条新的社群守则。作为对我们在“乳腺癌症状宣传与裸露内容”案件审理决定中建议的回应，Meta 还加强了在 Instagram 内容中识别乳腺癌背景的技术，成千上万个此前会被自动一删了之的帖子因此被提交给了人工审查。

2022 年，我们向 Meta 提出了 91 条政策建议并将其中过半建议（包括有关 Meta 在交叉检查计划中应当如何对待高影响力用户的建议）列入了第一批政策咨询意见。作为回应，Meta 承诺对那些面临过度执行风险的人士（包括记者和人权捍卫者）加强保护。

我们还保护了用户言论，尤其是政治、社会转型和危机期间的用户言论。例如，在 2023 年初，我们在一项委员会决定中敦促 Meta 加强对伊朗政治言论的保护，当时伊朗发生了历史性的广泛抗议活动并受到了暴力镇压。作为回应，Meta 允许在以伊朗当前抗议活动为背景的帖子中分享口号“Marg bar Khamenei”（字面翻译是“[伊朗最高领袖]哈梅内伊去死”）。Meta 现在还改动了违规惩罚机制，以便更公平地对待用户。

每项决定和政策咨询意见都为原本不够透明的内容审核流程加强了透明度，包括披露 Meta 在执行规则过程中援引新闻价值例外的次数。我们的政策建议引发了关于数字平台应当如何处理内容审核方面的一些复杂难题的公开讨论。

为了加强我们的影响力，我们希望与利益相关者合作改善人们的线上体验，并为此确定了七项优先领域：选举和公民空间、危机和冲突局势、性别、针对边缘化群体的仇

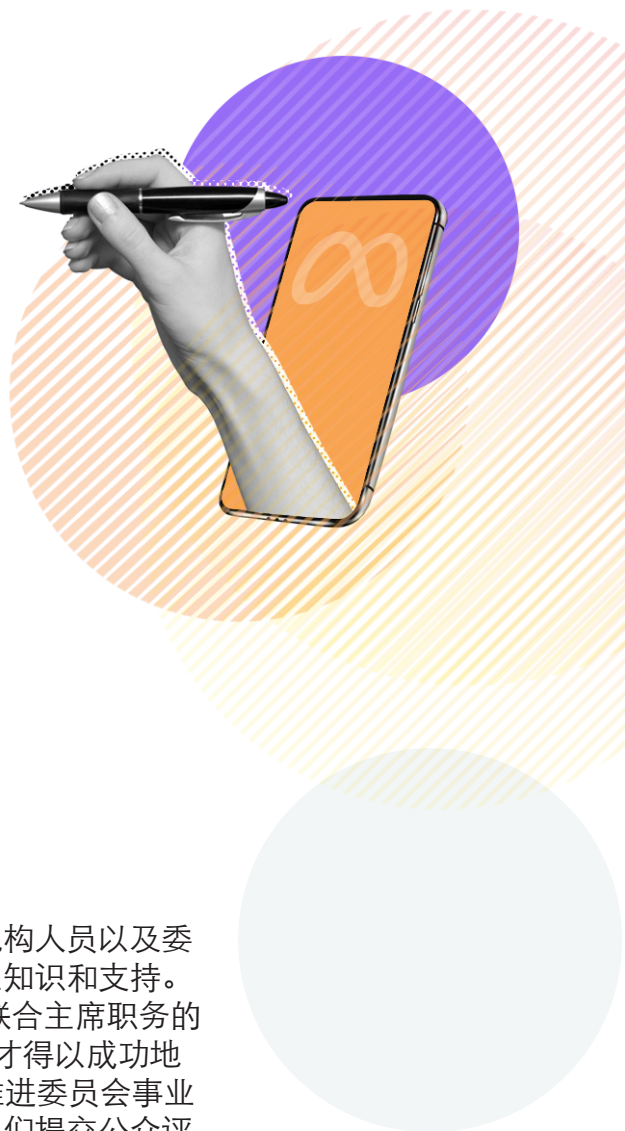


恨言论、政府对 Meta 平台的使用、公平对待用户以及自动执行政策和内容管理。我们还准备在 2023 年承担更多的案件量并更快地做出决定。

在 2022 年，我们还看到越来越多的人认识到，有关内容审核的重大决定不应由公司单方面做出。委员会从成立之初便致力于对内容审核进行独立监督（若能取得成功，也可以为其他公司所用）。独立监督要求公司敞开大门，欢迎外部人士对其工作提出质疑。在过去三年中，我们在独立监督方面积累了丰富的经验，可以帮助公司在有关尊重言论自由和其他人权方面做出更稳健的决定。我们相信，对于新法规带来的新要求，以及一些有关透明度和用户告知方面的特定领域，我们可以提供部分解决方案。

我们想借此机会感谢全体监督委员会受托人、行政管理机构人员以及委员会成员，感谢所有人迄今为止在我们工作中提供的专业知识和支持。我们要特别感谢在 2022 年 12 月辞去监督委员会成员兼联合主席职务的 Jamal Greene 的贡献。正是在 Jamal 的领导下，我们才得以成功地让 Meta 承担起种种责任，我们感谢他在建立委员会和推进委员会事业方面的付出。我们还要感谢许许多多利益相关者，感谢他们提交公众评议、参与工作并帮助我们取得了迄今为止的成就。

监督委员会正走在一条不曾有前人走过的道路上，未来将继续适应形势变化，找到各种新途径来完成自身使命。虽然目前一切进展顺利，但我们并不会因此对前方可能遇到的挑战掉以轻心。我们将与所有人携手合作，帮助人们克服社交媒体中的陷阱，自信地进行人与人之间的互动联系。





信托机构主席寄语



**STEPHEN
NEAL, 监督委员会
信托机构主席**

2022 年是我担任监督委员会信托机构主席的第一个整年，委员会的工作给我留下了深刻的印象。委员会成员继续就一些重大疑难案件展开审议，并针对一系列问题发布了重大决定。这些决定中包括委员会的第一批政策咨询意见，其中列出了委员会对分享私人住所信息、“俄语诗”（与俄乌冲突有关）案件审理决定的审议意见。作为受托人，我们帮助任命了来自埃及、墨西哥和美国的 3 名新委员会成员。

受托人的另一项重要职责是对监督委员会行政管理机构进行监督，行政管理机构由辅助委员会成员开展工作的全职工作人员构成。2022 年，行政管理机构完成了所有团队的招聘工作，目前约有 80 人，分布在伦敦、华盛顿特区和旧金山等地。我们招揽了一些优秀的新同事，其中许多人在言论自由和人权领域拥有独到的专业知识。行政管理机构与社交媒体本身一样富于全球化色彩，工作人员之间使用的语言达到了 40 种。

2022 年 7 月，我们宣布 Meta 在 2019 年信托成立之初承诺提供的 1.3 亿美元之外，承诺追加提供 1.5 亿美元。通过做出持续的财务承诺，Meta 对委员会的工作以及委员会致力于让 Facebook 和 Instagram 内容相关规则在具体实施时以保护言论自由并尊重相关人权标准为前提的努力投下了信任票。

在 2023 年，我们将继续监督委员会的运作并维护委员会的独立性，这两者对委员会能否起到应有的作用至关重要。委员会将继续做出案件审理决定并提供政策咨询意见，完善 Meta 的产品和政策，提升 Facebook 和 Instagram 的用户体验。委员会希望通过与民间社会团体、监管机构和其他平台开展合作，树立正面形象。Meta 员工是另一个在很大程度上左右着 Meta 未来的重要群体，委员会将致力于让他们拥护、宣传委员会的工作。

Meta 致力于将委员会打造成社交媒体新型治理模式的努力以及为此做出的持续承诺，值得所有人的赞扬。作为受托人，我们将支持委员会继续对 Meta 发挥作用，并将帮助委员会与业内其他公司和合作伙伴分享成功之道。

“通过做出这一持续的财务承诺，Meta 对委员会的工作投下了信任票”

监督委员会信托机构主席
Stephen Neal

执行摘要

2022 年，监督委员会向 Meta 提出了 91 条建议

迄今为止，针对我们的建议，Meta 做出了如下回应：



开始在删除用户内容时向用户说明相关内容违反了哪项具体政策。



改进了在 Instagram 内容中识别乳腺癌背景的方式，成千上万个此前会被自动一删了之的帖子因此被提交给了人工审查。



在社群守则中新增了有关错误信息的部分。



开始系统地衡量向用户发送消息告知执行决定时的透明度。



在全球范围内推出新的提示消息，告知用户导致内容被删除的是人工审查还是自动审查。



推出了一项新的危机政策规程。

在 2022 年，监督委员会：



收到了世界各地用户提交的近 130 万个案件 – 比 2021 年增加了近四分之一



发布了 12 项案件审理决定，涉及范围从俄乌冲突到执法部门对内容删除的影响等



发布了第一批政策咨询意见，内容涉及私人住所信息分享、Meta 交叉检查计划等



在 12 项案件审理决定中，以四分之三的比例推翻了 Meta 的决定 – 推翻内容审核决定 9 次，支持 3 次

获得了在符合条件的内容中增加警告画面的权限

促使 Meta 在案件甄选委员会考虑选择的 32 个案件中推翻了最初对帖子的错误决定

在 2023 年，我们将：



发布第一批针对 Meta 推翻了对相关内容的最初决定的那些案件的简易决定。



发布第一批加急决定，我们将在受理此类案件后几天内发布决定。



按最新公布的计划中委员会成员人数实现满员招聘，最大限度提升工作效率。



围绕七项战略重点，深入开展互动活动。



追求长期计划，扩大业务范围。



监督 Meta 对建议的落实情况，敦促 Meta 提供证据说明落实情况及其影响。

我们相信独立监督有其重大价值，未来将探索与业界公司结成新型合作伙伴关系的可能性，并力求通过我们的工作尽力为相关领域内处于萌芽状态的监管制度填补空白。

认识委员会

监督委员会成员



Afia Asantewaa Asare-Kyei
西非开放社会倡议组织 (Open Society Initiative for West Africa) 的项目经理



Tawakkol Karman
诺贝尔和平奖得主



Evelyn Aswad
俄克拉荷马大学法学院教授兼系主任



Sudhir Krishnaswamy
印度大学国立法学院副校长兼法学教授



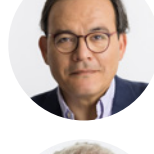
Endy Bayuni
《雅加达邮报》高级编辑兼董事会成员



Ronaldo Lemos
里约热内卢联邦大学法学院教授



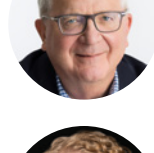
Catalina Botero-Marino
安第斯大学联合国教科文组织言论自由教席主持人



Khaled Mansour
作家



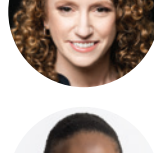
Paolo Carozza
圣母大学教授



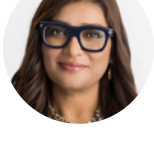
Michael McConnell
斯坦福法学院教授兼宪法中心主任



Katherine Chen
国立政治大学教授



Suzanne Nossel
美国笔会首席执行官



Nighat Dad
数字权利基金会 (Digital Rights Foundation) 创始人



Julie Owono
互联网无国界医生组织执行总监



Emi Palmor
以色列赫兹利亚跨学科研究中心的
讲师和权益倡导者



Pamela San Martín
墨西哥国家选举研究所 (INE) 前选
举委员



Alan Rusbridger
牛津大学玛格丽特夫人学堂校长



Nicolas Suzor
昆士兰科技大学法学院教授



András Sajó
中欧大学教授



Helle Thorning-Schmidt
丹麦前首相



John Samples
加图研究所副所长



吉野兼司
宪法学首席大法官厄尔·沃伦
(Earl Warren) 教授兼梅尔泽多样
性、包容性和归属感中心 (Meltzer
Center for Diversity, Inclusion,
and Belonging) 主任

监督委员会受托人



Kristina Arriaga
受托人



Kate O'Regan
受托人



Cherine Chalaby
受托人



Robert Post
受托人



Stephen Neal
信托机构主席



Marie Wieck
受托人

监督委员会行政管理机构



Thomas Hughes
总监

推出七项战略重点

2022 年 10 月，我们在对用户向委员会提起申诉的问题进行广泛、深入的分析之后，宣布了七项战略重点。我们目前按照案件是否涉及这些重点对案件进行甄选，因此建议用户提交申诉时在理由中列入涉及的重点。



1. 选举与公民空间

社交媒体公司面临着如何一视同仁地将其政策适用于世界多地（包括选举和大规模抗议期间）政治性言论的问题。我们在“俄罗斯境内声援纳瓦尔尼的抗议活动”案件审理决定中强调了保护政治性言论的重要性，在“新闻报道中提及塔利班”案件审理决定中则触及了媒体自由的问题。委员会希望探索 Meta 在选举、抗议和其他公民活动关键时刻的责任边界。



2. 危机和冲突局势

在发生诸如武装冲突、恐怖袭击、突发卫生事件等危机的时候，社交媒体既可能成为人们交流重要信息、讨论重要公共问题并争取安全度过危机的利器，也可能成为别有用心之人传播错误信息和仇恨言论的环境。我们在“Raya Kobo 指控暴行”和“提格雷通讯事务管理局”等案件的审理决定中审查了与埃塞俄比亚冲突有关的帖子；我们在前总统特朗普相关案件中的审理决定最终促成 Meta 通过了一项危机政策规程。委员会希望探讨 Meta 对在此类情况下保护言论自由可以发挥什么作用，以及它对旗下产品在武装冲突、内乱和其他紧急情况下可能造成的潜在危害可以做出哪些准备。



3. 性别

女性、非二元性别者和跨性别者在社交媒体上行使言论自由权时受到阻碍。例如，我们在“乳腺癌症状宣传与裸露内容”案件审理决定中提到，由于 Meta 自动化系统未能对乳腺癌宣传适用例外，导致重要健康信息被从 Instagram 中删除。我们在 2023 年初发布的“性别认同和裸露内容”案件审理决定中认为，Meta 有关成人裸体的政策加重了女性、跨性别者和非二元性别者在 Facebook 和 Instagram 上行使言论自由权时受到的阻碍。委员会希望探索女性和性少数群体 (LGBTQIA+) 在行使其言论自由权时面临的性别阻碍，包括基于性别的暴力和骚扰，以及内容政策中基于性别的区别对待所造成的影响。



4. 针对边缘化群体的仇恨言论

仇恨言论将营造歧视和敌视边缘化群体的环境。这类言论通常基于特定背景，使用密码式语言，其影响随时间累积并将造成伤害。我们在“扮演黑彼得”案件审理决定中认为，允许扮演黑人的图像在线上累积将助长营造歧视黑人的环境；我们在“贝壳串珠带”和“给某些阿拉伯词语重新赋予积极意义”案件审理决定中对“反语”进行了分析（“反语”也即反话正说，援引仇恨言论来表达抵制歧视的态度）。委员会希望探讨 Meta 在保护边缘化群体成员方面所应起到的作用，同时确保 Meta 在执行规

则时不致错误地处罚那些反对仇恨的人士。与此同时，委员会了解对仇恨言论的限制不应过度执行或被用于限制人们依法自由发表言论，包括发表不受欢迎或有争议的观点。



5. 政府使用 Meta 平台

政府会使用 Facebook 和 Instagram 来向公众传达政策，也会向 Meta 提出删除内容的请求。作为对我们在“奥贾兰单独监禁”案件审理决定中建议的回应，Meta 同意提供根据政府举报以违反社群守则为由删除的内容的相关信息。我们还在“英国 Drill 音乐”案件审理决定中建议 Meta 应当如何回应国家执法部门的要求。委员会希望探究国家行为者如何使用 Meta 平台，如何影响内容审核实务和政策（有时以不透明的方式），以及国家参与内容审核的影响。



6. 公平对待用户

人们发布的内容被从 Facebook 和 Instagram 中删除时，并不是每次都会被告知违反了哪条规则。还有一些时候，用户没有得到平等对待，没有得到适当的程序保障，没有机会对所犯的错误进行补救。委员会希望探索 Meta 如何通过提供更具体的用户通知来更好地对待用户，确保用户每次都有机会对 Meta 的决定提起申诉，同时在“违规”和交叉检查等领域加强透明度。



7. 自动执行政策和内容管理

虽然算法对大规模进行内容审核至关重要，但目前关于 Meta 自动化系统如何运作以及如何影响用户看到的内容，普遍存在着规则缺乏透明度、用户缺乏了解的问题。我们在“影射哥伦比亚警察的漫画”案件审理决定中说明了自动化系统如何放大错误的内容审核决定的影响。作为对我们在“乳腺癌症状宣传与裸露内容”案件审理决定中建议的回应，Meta 在全球范围内推出了新的提示消息，向用户说明是人工审查还是自动审查导致了他们的内容被删除。委员会希望探讨应当如何设计和审查自动执行系统，并希望探讨自动化系统的准确性和局限性，以及在此领域加强透明度的重要性。

与利益相关者合作，扩大我们的影响

委员会之所以能够取得如今的成就，在于委员会处理问题时会找到多年致力于解决这些问题的研究人员、民间社会团体和其他人士，听取多方的意见并与之合作。要想为我们的战略重点及其带来的重大复杂难题找到切实可行的解决方案，利益相关者的专业知识及其对当地情况的了解都必不可少。

我们将继续就所有战略重点，与足以体现 Meta 平台用户多样性的众多利益相关者开展合作。我们将通过这种方法帮助了解 Meta 有哪些政策和执行层面操作亟需改进，以及有哪些类型的案件可以让我们有机会解决这些问题。我们希望与全球各地组织合作实现这一目标，具体方式包括公众评议程序、圆桌会议和个人对话。如果您所在的组织希望参与此项事业，欢迎致信 engagement@osbadmin.com 与我们展开具体探讨。

委员会如何处理用户申诉

此图显示了适用于 2022 年用户申诉决定的申诉流程。



决定

审议小组就是否允许发布内容做出决定——支持或推翻 Meta 的决定。



发布

我们的决定将发布在监督委员会网站上。Meta 应在委员会发布决定后 7 天内落实决定，并在 60 天内对任何建议做出回应。



审批

决定草案将分发给所有委员会成员以供审查。一项决定要公布必须得到多数人的签署同意。

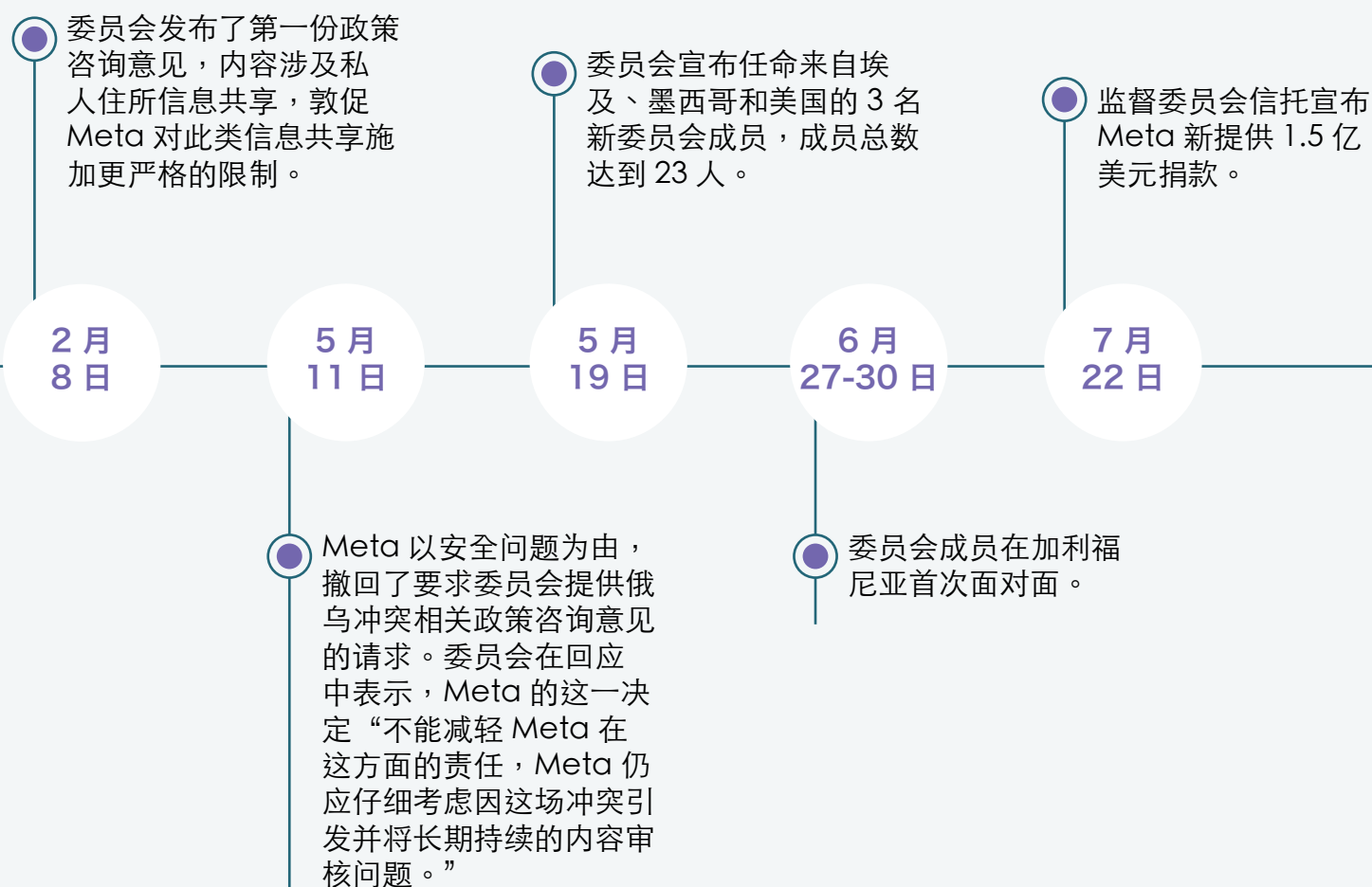


落实

委员会监督 Meta 对建议的落实情况，并在季度透明度报告中介绍最新进展。



2022 年重要事件时间表



7月
26日

委员会收到 Meta 要求就删除新冠疫情错误信息提供政策咨询意见的请求。

10月
20日

委员会宣布了七项战略重点，重点关注可能对人们的 Facebook 和 Instagram 体验产生最大影响的领域。

10月
20日

委员会开始能够要求 Meta 在恢复或保留符合条件的内容时附带“令人不安”或“敏感”警告画面。

11月
22日

委员会发布了“英国 Drill 音乐”案件审理决定。这是委员会首次审查应国家执法部门要求删除的帖子。

12月
6日

委员会发布了有关 Meta 交叉检查计划的政策咨询意见。委员会认为交叉检查计划在几个重大方面存在缺陷，并向 Meta 提出了 32 条建议。

建议和影响

我们在 2022 年向 Meta 提出了 91 条建议

迄今为止，针对我们的建议，Meta 做出了如下回应：



在全球范围内推出新的通知消息，向用户告知违反了仇恨言论、危险人物和组织以及欺凌和骚扰政策中的哪项具体政策。



推出新的通知消息，向用户告知应政府要求基于当地法律限制他们访问相关内容。



在社群守则中新增了有关错误信息的部分。



开始系统地衡量向用户发送消息告知执行决定时的透明度。



推出了一项新的危机政策规程。



改进了在 Instagram 内容中识别乳腺癌背景的方式，成千上万个此前会被自动一删了之的帖子因此被提交给了人工审查。



开始深入审核危险人物和组织政策，按危险高低对名单中的人物和组织排定优先等级。



在全球范围内推出提示消息，告知用户导致内容被删除的是人工审查还是自动审查。



概述

我们通过案件审理决定和政策咨询意见等途径，就如何完善适用于数十亿用户的政策向 Meta 方面提出了许多具体建议。虽然我们的建议并不具约束力，但 Meta 必须在 60 天内公开予以回应。Meta 公开承认，我们的建议在促使他们改变一些以往的做法。2022 年 8 月，Meta 表示委员会“继续推动我们更加审慎地对待全球内容审核造成的影响，并更加公平地实施政策和使用资源。最为重要的是，委员会还促使我们变得更加透明，因为外部声音有助于让我们对自己的承诺负责。”

委员会通过每季度发布透明度报告，要求 Meta 担负起相关责任。这些报告运用严格、独立、用数据说话的分析方法来评估 Meta 在落实建议方面的进展。委员会通过公开提出建议并公开监督 Meta 的回应和落实情况，为与 Meta 之间进行透明对话创造了一个前所未有的空间。这种开放性有助于在用户和民间团体中间树立正面形象，并赢得他们的信任。

“[委员会的建议]还促使我们变得更加透明，因为外部声音有助于让我们对自己的承诺负责”

Meta 发布的 2022 年第二季度监督委员会最新情况通报

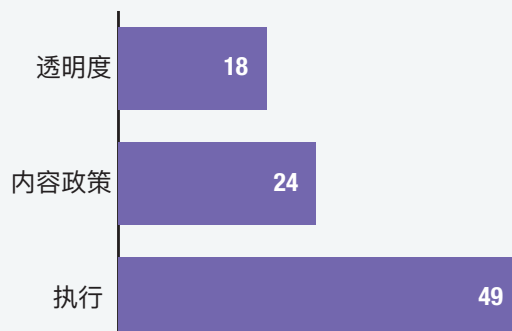
关于民间团体在我们形成建议过程中所起的作用，再怎么强调也不为过。在许多情况下，这些组织会通过公众评议流程提交对建议的具体设想。而在其他情况下，我们的提议是对这些组织多年呼吁的回应，以此迫使 Meta 考虑并公开回应多年来要求其采取行动呼吁。我们在案件审理决定中明确地对这些组织的影响表达了感谢，这里想再次感谢这些组织与我们分享他们的想法和专业知识。

2022 年建议和影响

委员会在 2022 年向 Meta 提出了 91 条建议，略高于我们在 2021 年提出的 86 条建议。截至 2023 年 4 月初，我们又提出了 14 条建议，也即总共向 Meta 提出了 191 条建议。

总共有 41 条建议被归入“已证明落实”或“已证明部分落实”类别，另外有 84 条建议被评估为“已汇报进度”。在 2022 年，我们尤为欣慰地看到，Meta 在落实建议时采取的几项变革措施对 Meta 的做法产生了系统性的影响。这其中包括我们自 2021 年 1 月以来始终要求 Meta 采取的一项措施，也即向用户提供更具体的通知消息。由于委员会反复建议，对于因用户内容违反相关政策而采取执行措施时发送给用户的通知消息，Meta 还更新了衡量其是否足够具体、透明的标准。Meta 并未止步

委员会在 2022 年提出了哪些类型的建议？



于这项系统性变革措施，还在更大范围内希望向用户更具体地说明情况。Meta 还深入复审了危险人物和组织政策中“赞扬、实质性支持和代表”（PSR）框架下对“赞扬”的定义。Meta 使用此框架来评估用户内容中对危险人物和组织的正面描述程度。

经验教训

2022 年，与 Meta 的合作可以得到改善的一个领域是数据访问。在 2021 年底和 2022 年初，我们花了八个月的时间试图访问 Meta 的 CrowdTangle 工具，以便获得更多信息帮助甄选案件并评估建议产生的影响。在几次受挫后，我们在 2022 年初将这个问题上报给 Meta 领导层，最终获得了访问权限。2022 年晚些时候，Meta 在数据共享方面采取了几项积极措施，包括共享进行中的用户申诉研究，以及聘请数据科学家来验证对建议的落实情况。我们期待继续与 Meta 的数据科学团队合作，获得有意义的数据来证明建议的落实情况及其影响。

从承诺到行动：为用户造福

由于我们的建议每每目标远大，而且经常需要技术变革才能实现，我们理解建议的落实需要时间。2022 年，我们看到我们在 2021 年提出的许多建议取得了进展，也看到 Meta 针对我们近期提议做出了新的承诺。我们欣慰地看到，Meta 首次对规则及其执行方式进行了系统性的变革，包括涉及用户通知消息以及危险组织相关规则。下文中的例子说明了我们的建议对 Meta 如何对待全球用户和社群所产生的影响。

|| META 规则及其执行方式发生系统性变革

- ◎ 委员会最常向 Meta 提出的一条建议是，在删除用户内容的时候向用户告知他/她做错了什么。我们最早在 2021 年 1 月提出这条建议，之后 Meta 逐步朝着这个目标前进。作为对这条建议的回应，Meta 在全球范围内推出新的通知消息，向用户告知其违反了仇恨言论、危险人物和组织以及欺凌和骚扰政策中的哪项具体政策。作为对我们建议的回应，Meta 现在还在系统地衡量所有与内容删除相关的用户通知消息的详细程度。
- ◎ 作为对我们在前总统特朗普相关案件审理决定中建议的回应，Meta 对回应危机和冲突的方式完成了系统性变革。2022 年 8 月，Meta 在咨询了 50 多位全球专家后，发布了危机政策规程。这项规程将有助于为 Meta 应对危机状况提供更加一致、透明的依据。2023 年 1 月 25 日，Meta 指出，在对前总统特朗普的账户做出决定之前，使用了危机政策规程来评估当前环境，包括研究了美国 2022 年中期选举情况。
- ◎ 我们有许多提议都在敦促 Meta 就规则和例外向用户提供更大的透明度。作为回应，Meta 于 2022 年 8 月首次披露了以符合公共利益为由向违规内容适用“允许保留具新闻价值内容”例外的次数。在 2021 年 6 月 1 日到 2022 年 6 月 1 日期间，Meta 档案中记录了 68 次适用例外的情况，

建议亮点：

进一步认识我们的影响力

我们在 2021 年 1 月发布了“乳腺癌症状宣传与裸露内容”案件审理决定，并在决定中建议 Meta “提高对带文字图像的自动检测水平，以免将旨在让更多人了解乳腺癌症状的帖子误标为待审核的内容。”作为回应，Meta 的落实团队增强了 Instagram 通过文本识别乳腺癌背景的技术，并于 2021 年 7 月部署了这些技术。这些增强功能已经到位，在 2023 年 2 月 26 日至 3 月 27 日的 30 天内，这些增强功能让 2,500 条此前会被一删了之的内容提交给了人工审查。

在没有简单判断方法的情况下，要找出 2,500 条包含乳腺癌症状宣传背景的内容无疑是一件难事，但我们可以从中看到，Meta 在严格按照我们设计的框架落实建议，而且此举成功减少了平台过度执法的情况。这是独立治理的胜利，在认识到我们对 Meta 系统影响力的前提下，我们与 Meta 之间的合作也将揭开全新的篇章。

其中 13 次例外给了政治人物发的帖子。这种开放性有助于在用户和民间团体中间树立正面形象，并赢得他们的信任。

- 为了回应我们关于危险人物和组织规则的建议，Meta 启动了对该政策领域的深入审查。这次审查的重点是按风险高低指定危险人物和组织名单，其中被评估为风险最高的实体将被优先执行。我们在几个案件审理决定中都认为，Meta 在此项政策中有关“赞扬”的定义过度限制了用户的言论自由。作为回应，Meta 正在审查它在判断有关内容是否构成赞扬、实质性支持或代表指定人物或组织时的依据。

公平对待用户

- 2022 年，Meta 更新了应政府要求基于当地法律删除内容时发给用户的通知消息。这些消息将向用户告知内容已受限，并将说明 Meta 如何处理此类政府要求。
- Meta 在全球范围内推出新的提示消息，告知用户是人工审查还是自动审查导致内容被删除。
- 作为对我们在“宣称疗法可治愈新冠病毒感染”案件中决定的回应，Meta 在社群守则中新增了有关错误信息的部分，将所有相关规则整合到一处并澄清了这些规则。



建议亮点：

有关 Meta 在 2021 年 5 月以巴冲突中影响的独立人权报告

2021 年“分享半岛电视台帖子”案件中提交的公众评议声称 Meta 在删除巴基斯坦用户帖子时未遵守“相称性”原则，我们在阅读这些评议以后敦促 Meta 聘请一个独立实体来调查有关 2021 年 5 月以巴冲突的内容审核活动是否存在偏见。

Meta 同意并委托非营利组织商务社会责任国际协会 (BSR) 进行审查，审查报告于 2022 年 9 月公布。报告得出的结论是，Meta 在 2021 年 5 月以巴冲突期间的内容审核活动似乎对人权造成了负面影响，具体是对巴基斯坦用户的言论自由权，以及他们根据经历分享信息和见解的能力。BSR 报告中指出的大部分偏见与 Meta 内容审核员欠缺语言能力和相关指导有关。委员会曾在 3 个不同的案件审理决定中敦促 Meta 将审核员内部指南翻译成所审核内容的语言。但 Meta 一直拒绝落实这一项建议，理由是审核员都精通英语，有英文版指南便已足够。BSR 报告表明，由于 Meta 方面欠缺语言能力、审核员对文化背景了解不足，才导致对巴基斯坦阿拉伯语内容执行规则过度，对反犹太主义内容执行规则不力。正如我们之前曾指出的那样，纯英文版指南可能导致审核员不能掌握不同语言的背景和细微差别。此外，由于人工审查数据被用来训练分类器，这种偏见在 Facebook 和 Instagram 上被放大了。

“Meta 在 2021 年 5 月的做法似乎对人权产生了不利影响……对巴基斯坦用户的言论自由权。”

有关 Meta 在以巴冲突中影响的人权尽职调查报告 - 商务社会责任国际协会 (BSR) 报告



|| META 需要进一步予以关注的建议

1. 确保 Facebook 和 Instagram 适用相同规则。
2. 将审核员内部指南翻译成所审核内容的语言。
3. 更具体说明 Meta 内部如何上报具新闻价值的帖子。

Meta 落实了我们之前提出的许多建议，切实改善了用户在 Facebook 和 Instagram 上的体验。但是，到目前为止，Meta 未落实三个方面的关键建议。

确保 Facebook 和 Instagram 适用相同规则。最早在 2021 年 1 月的“乳腺癌症状宣传与裸露内容”案件审理决定中，我们建议 Meta 确保 Facebook 和 Instagram 适用相同规则。Meta 虽然一开始接受了建议，但一再推迟最后落实期限。Meta 也还没有按照我们的建议，向用户告知 Meta 在 Instagram 上执行 Facebook 社群守则，如果有关内容在 Facebook 上会被认为违规，那么它在 Instagram 上也会被认为违规。

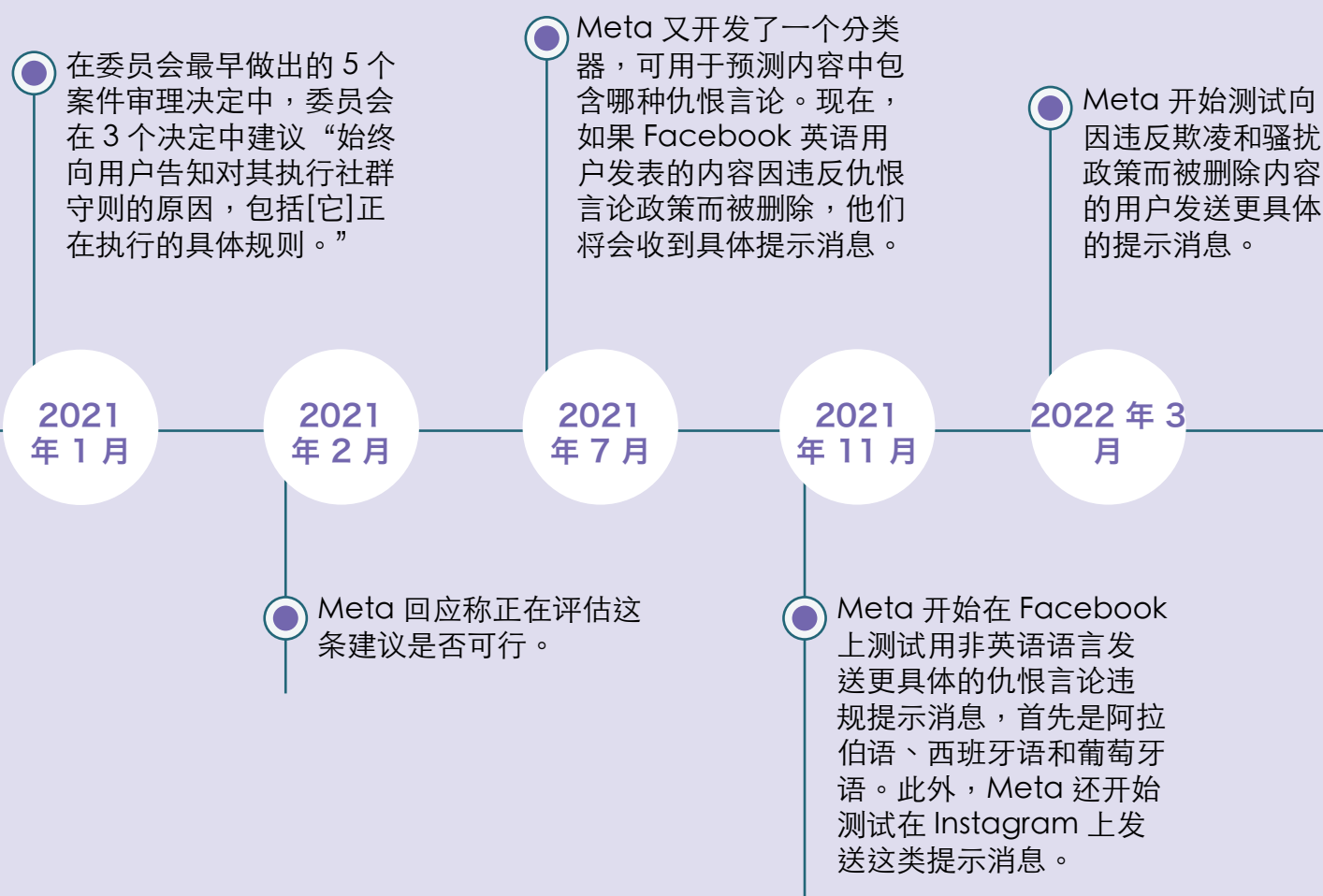
将审核员内部指南翻译成所审核内容的语言。在与旁遮普语、缅甸语和阿拉伯语内容相关的三个案件审理决定中，我们都建议 Meta 将审核员内部指南（内部执行标准）翻译成相关内容的语言。但 Meta 无视这个建议以及 BSR 在近期报告中提出的类似考量，一再表示不会对这项建议采取“进一步行动”，因为审核员全都精通英语。更具体说明 Meta 内部如何上报具新闻价值的帖子。今年，Meta 首次披露了适用“允许保留具新闻价值内容”例外的次数。不过，对于它判断内容是否具有新闻价值时适用的流程，人们还是知之甚少。我们在 2021 年“哥伦比亚抗议”案件审理决定中，呼吁 Meta 制定并公布内容审核员在上报违反 Meta 规则但可适用“允许保留具新闻价值内容”例外的内容时所应遵循的明确标准。Meta 对此建议的回应显示它似乎误解了此项建议的用意，它只是强调某些内容可以在经上报后适用“允许保留具新闻价值内容”例外，没有回应建议中的要求：说明在什么情况下、基于什么理由上报这些内容。因此，我们建议 Meta 接下来优先分享这方面的信息。

|| 关于 META 落实和回应我们建议的及时性的报告

- ⊙ 根据委员会细则，Meta 必须在委员会发布案件审理决定或政策咨询意见后 60 天内公开回应委员会的建议。
- ⊙ 在 2022 年发布的 12 项案件审理决定和 2 项政策咨询意见中，除了有关交叉检查的政策咨询意见，Meta 在所有其他情况下都在此时间范围内回应了建议。至于有关交叉检查的政策咨询意见，由于委员会提供了众多建议（总共 32 条），Meta 在 90 天内而不是通常的 60 天内回应了建议。

影响时间表：让用户知道他们的错处

作为对我们提议的回应，Meta 现在向全球更多用户说明被删除的内容违反了哪项具体政策。



2022
年 3 月

Meta 表示，提供更具体的通知消息导致“在多个市场上，人们对 Meta 拥有透明度和正面形象的看法在统计数据上出现了显著增加。”

2022
年 5 月

Meta 称其研究表明，内容审核流程需要类似促进相互理解的对话，才能确保让用户感到自己的意见得到倾听。

2022
年 8 月

Meta 推出新的提示消息，向用户确切说明依据哪项具体政策采取执行行动。提示消息在全球范围内以英语提供，正在翻译成其他语言。

2022
年 11 月

Meta 表示，到 2022 年底，这种新的提示消息将适用于 Facebook 上的绝大多数违规类型，并将在 2023 年将适用范围扩大到 Instagram。Meta 还将评估让通知消息更加具体的可行性。

Meta 对委员会建议的落实情况

为确保 Meta 兑现承诺，我们将监督 Meta 在落实建议方面的进展。我们将通过了解是否已达到特定建议给定的标准来进行监督。我们将把 Meta 在落实建议方面的进展分为七大类，每季度更新一次评估结果：



“通过公布的信息证明了落实情况” – Meta 向委员会提供了足够多的数据来证明建议已得到落实。



“通过公布的消息证明了部分落实情况” – Meta 已落实建议的核心部分，并向委员会提供了足够的证据来证明这一点。



“已汇报进度” – Meta 已经承诺落实这个建议，但尚未执行完所有必要的行动。



“Meta 报告称已落实建议或表明已完成相关工作，但没有公布信息来证明落实情况” – Meta 表示已落实此建议，但未向委员会提供足够多的证据来证明这一点。



“Meta 进行可行性评估后拒绝了建议” – Meta 了解了建议，但决定拒绝落实建议，并在拒绝前向委员会提供了相关信息。



“建议被忽略、拒绝或修改” – Meta 不会对委员会的提议采取任何进一步的行动。



“等待 Meta 的首次回应” – 委员会提出了建议，但 Meta 尚未公开回应（因为它在委员会公布建议后有 60 天的时间来回应）。

采用这种用数据说话的方法，意味着我们对 Meta 是否落实了建议的评估结果可能会与 Meta 自身的说法有所不同。不过我们认为，这种独立验证方式对于责成 Meta 落实建议以及确保让用户感受到建议的作用至关重要。

截至 2023 年 4 月上旬本报告定稿时间，我们评估了 Meta 落实全部 191 条建议的进展情况并加以分类，请参阅本页中的图表。随附这份年度报告，我们还发布了一份附件，在其中完整列出了全部 191 条建议，以及我们对 Meta 回应和落实情况的评估。附件反映了 Meta 发布的截至 2022 年第四季度的委员会季度报告信息。

落实进展类别	建议数量
 通过公布的信息证明了落实情况	27
 通过公布的消息证明了部分落实情况	14
 已汇报进度	84
 Meta 报告称已落实建议或表明已完成相关工作，但没有公布信息来证明落实情况	29
 Meta 进行可行性评估后拒绝了建议	10
 建议被忽略、拒绝或修改	23
 等待 Meta 的首次回应	4
建议总数	191

案件甄选



2022 年委员会收到了 1,290,942 起案件，比 2021 年增加了约四分之一。

在 2022 年，委员会平均每 24 秒收到一起案件。



在用户要求恢复内容的所有申诉中，超过三分之二的申诉只和 2 条社群守则有关：暴力与煽动暴力，以及仇恨言论。



在委员会于 2022 年选择的所有案件中，Meta 在近三分之二的案件中认识到它最初对内容的决定有误。

概述

每天都有数亿条帖子在 Facebook 和 Instagram 上被分享，数以万计的内容审核员需要瞬间决定哪些内容应予保留、哪些内容应予删除，决定选择哪些案件进行复审就成了摆在我们面前的一大难题。2022 年，我们收到了近 130 万份要求对 Meta 内容审核决定进行独立复审的用户申请。Meta 也向委员会提请复审 21 起案件。2022 年，我们平均每天收到 3,537 起案件。

我们重点关注那些提出了重大难题，同时牵涉到广泛地理区域和众多主题领域的案件。通过在众多案件中做出取舍，我们得以审视对用户影响最大的那些政策，对政策展开深入细致且有原则的审核，从而面向全体用户而不仅仅是案件得到审理的那部分用户来完善相关政策。例如，由于选择了“影射哥伦比亚警察的漫画”案件，我们才有机会针对 Meta 媒体匹配库（用于自动删除违反 Meta 规则的图像）提出建议。同理，由于“英国 Drill 音乐”案件，我们才有机会针对 Meta 应如何回复全球执法机构的要求提出建议。

为了解决与全球用户相关的问题，我们继续从全球不同地区选择案件。我们还选择了一些对公共言论至关重要的案件，例如“俄语诗”案件（研究了俄乌冲突背景下的表达问题）以及“给某些阿拉伯词语重新赋予积极意义”案件（涉及影响性少数群体的贬损词语的使用）。

“我们会尽量选择有实际影响，同时触及 Meta 常见问题的那些案件”

监督委员会联合主席
Michael McConnell



经验教训

2022 年底，我们在撰写“英国 Drill 音乐”案件审理决定和 Meta 交叉检查计划相关政策咨询意见时，发现 Meta 的申诉流程存在一个问题：对于某些经高级别审核的案件，用户没有机会向委员会提起申诉，而依照委员会细则，这些案件有资格接受审核。这会带来透明度和用户申诉权方面的严重问题。我们在“英国 Drill 音乐”案件审理决定和 Meta 交叉检查计划相关政策咨询意见中，敦促 Meta 纠正这一错误。

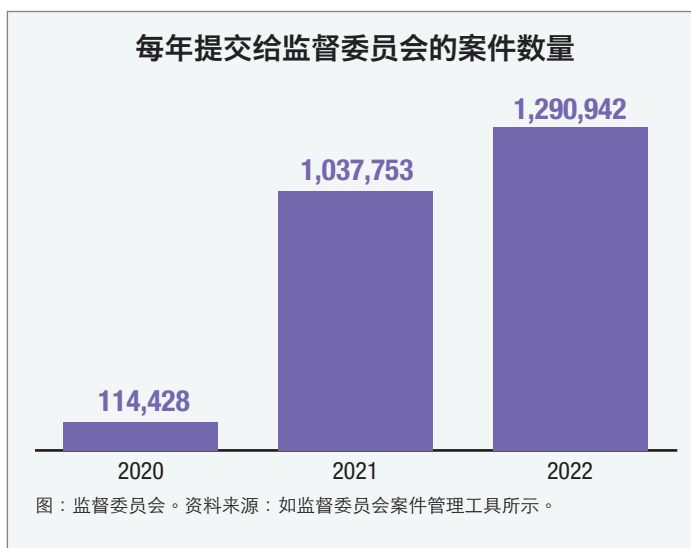
2023 年 1 月，Meta 回应称，欧盟、英国和印度用户不久后将可以对符合条件的经高级别审核的内容决定，向 Meta 和委员会提起申诉。“经高级别审核”的决定，由 Meta 内部专家团队而非外包人员执行。至于其他国家/地区的 Facebook 和 Instagram 用户，Meta 称计划为这些用户制定一套替代方案，对于符合条件的经高级别审核后删帖且不能在 Meta 内部提起申诉的决定，允许这些用户直接向委员会提起申诉。Meta 表示希望在 2023 年下半年之前落实该解决方案。

提交给委员会的案件

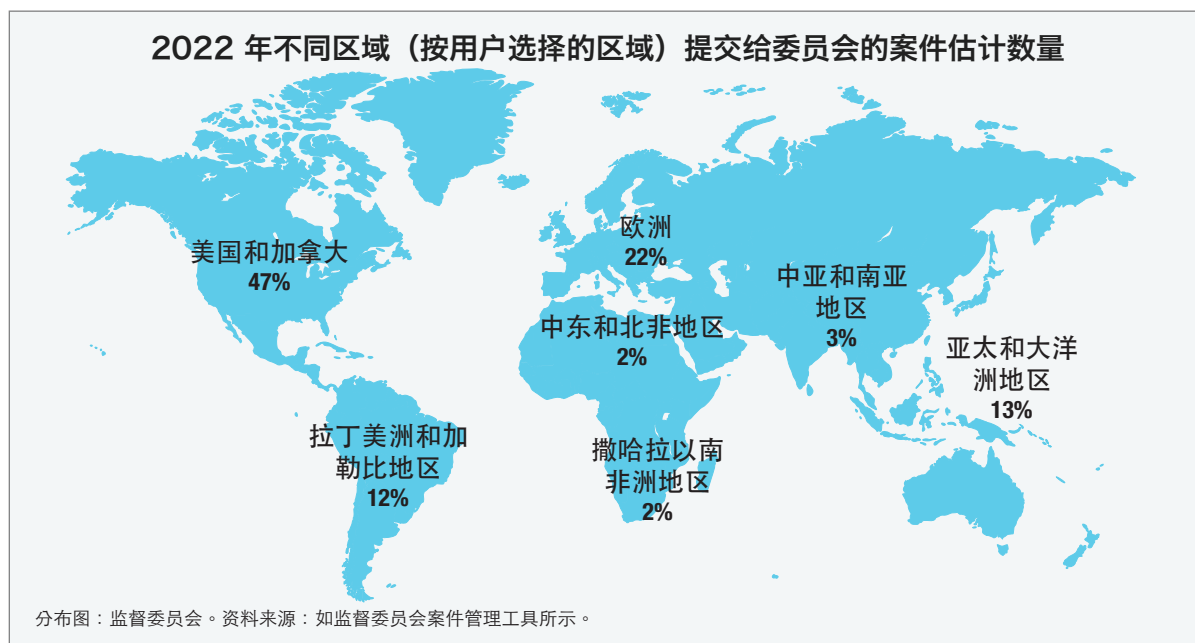
2022 年，我们收到的用户申诉数量增加了约四分之一——从 2021 年的约 100 万增加到 2022 年的约 130 万。2022 年第一季度和第二季度的申诉数量高于 2021 年的相应季度，第三季度和第四季度的申诉数量则比一年前有所减少。

从我们开始受理案件的 2020 年 10 月开始，到 2022 年 12 月，在此期间我们总共收到了超过 240 万份申诉——这表明不断有用户希望向独立机构申诉 Meta 的内容审核决定。虽然我们只能审核其中少数案件，但我们会继续选择涉及全球大量用户共同面临的基础问题的那些案件，并提出解决问题的建议。

2022 年，与 Instagram 内容有关的案件比例大幅增加。在 2022 年第一季度之前，这些案件只占每个季度申诉案件的 1%，但到了 2022 年第四季度，所占比例已经增加到 10% 以上。



2022 年，来自不同区域（指用户选择的区域）的案件比例与 2020 年 10 月至 2021 年 12 月期间大致相同，来自美国、加拿大和欧洲的申诉案件占全部案件比例超过三分之二。来自美国和加拿大的案件比例从 2020/21 年的 49% 略微下降到 2022 年的 47%。来自欧洲的案件比例从 2020/21 年的 20% 增加到 2022 年的 22%。来自亚太和大洋洲的案件比例增幅显著，从 2020/21 年的 9% 增加到 2022 年的 13%。2022

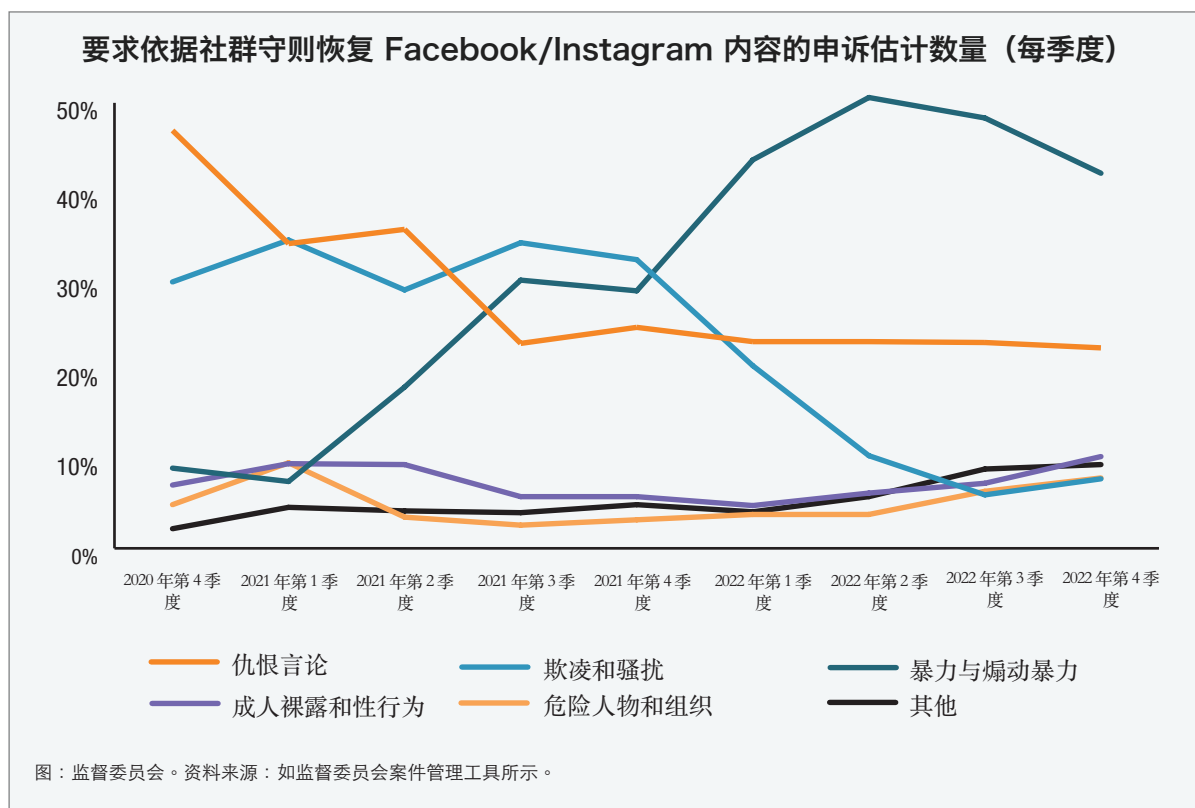


年，12% 的案件来自拉丁美洲和加勒比地区，其次有 3% 来自中亚和南亚，2% 来自中东和北非，2% 来自撒哈拉以南非洲。我们认识到，这些数据未能反映 Facebook 和 Instagram 用户在全球范围内的分布情况，也未能反映内容审核问题在全球范围内的实际分布情况。2023 年，我们将加强与美国 and 欧洲以外地区的联系和接触，以便宣传委员会的工作并鼓励这些地区的用户向委员会提出申诉。

2022 年，在用户要求恢复内容的所有申诉中，超过三分之二 (69%) 的申诉只和 2 条社群守则有关：暴力与煽动暴力（在要求恢复内容的申诉中占 46%）和仇恨言论（在要求恢复内容的申诉中占 23%）。以上数据只涉及用户要求恢复被 Meta 认为违规而加以删除的内容的申诉（在 2022 年全部申诉中占 92%），未涉及用户要求删除 Facebook 或 Instagram 上其他用户内容的申诉（在 2022 年全部申诉中占 8%），因为后者未被认为违反社群守则。

如图所示，在 2022 年，用户要求恢复被 Meta 认为违反暴力与煽动暴力社群守则而加以删除的内容的申诉比例增加，在 2022 年第二季度达到 51% 的高点。用户要求恢复被 Meta 认为违反欺凌和骚扰社群守则而加以删除的内容的申诉比例则有所下降，从 2021 年第四季度的 30% 左右下降到 2022 年的 10% 以下。

Meta 社群守则执行报告表明，2022 年 Facebook 上处理的暴力与煽动暴力内容也出现了类似的增长，从 2021 年第四季度的 1270 万条增至 2022 年第一季度的 2170 万条和 2022 年第二季度的 1930 万条。这可能部分解释了为什么委员会收到的申诉数量会有所增长。Meta 的报告还显示，在 2022 年第三季度，Meta 在 Facebook 上处理的欺凌和骚扰内容数量降至 660 万条，为近两年来的最低水平。这可能部分解释了为什么委员会收到的要求恢复依据此条社群守则被删除的内容的申诉比例有所下降。



案件甄选委员会考虑的案件

在案件甄选委员会确定了复审案件的入围名单后，Meta 有时也认为其针对某项内容制定的原始决定是错误的。

在 2022 年入围的 50 个案件中，Meta 认为它在其中 32 个案件中的最初决定有误 (64%)；而在 2020/21 年入围的 130 个案件中，Meta 认为它在其中 51 个案件中的最初决定有误 (39%)。Meta 认为最初决定有误的比例上升了 25 个百分点。

虽然这只是一个样本，而且委员会特意选择了重大疑难案件，但 Meta 认为最初决定有误的案件占入围案件近三分之二，这点让人颇感担忧。这让人对 Meta 内容审核的准确率，以及在委员会受理案件前 Meta 内部申诉流程的有效性产生了更大的疑问。2023 年，我们将开始发布“简易决定”，审理未被选中进行全面审理但导致 Meta 推翻最初执行行动的那些案件。

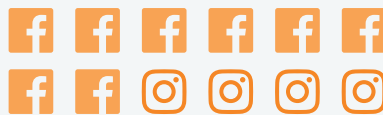


案件审理决定与政策 咨询意见

2022 年发布了
12 个案件审理决
定，其中 9 个推
翻了 Meta 之前
的决定。



7 起案件来自用户；
5 起案件来自 Meta。



8 起案件来自 Facebook；
4 起案件来自 Instagram。

发布了委员会的第一批政策
咨询意见，内容涉及私人住
所信息分享、Meta 交叉检
查计划等。



开始有权要求
Meta 在保留或
恢复内容时附带
警告画面

2022 年，我们在审理案件过程中向
Meta 提出了 356 个问题，Meta 完全
回答了其中 86% 的问题



概述

委员会针对用户申诉案件和 Meta 提请审核案件做出的决定，是委员会工作的重心。同时，委员会还在 2022 年发布了第一批政策咨询意见，内容涉及私人住所信息分享、Meta 交叉检查计划等。

案件审理决定

案件审理决定将审查 Meta 删除或保留内容的决定是否符合 Meta 规则、价值观和人权承诺。决定对 Meta 具有约束力，Meta 必须在决定发布后 7 天内落实，除非这样做会违反法律。2022 年，我们发布了 12 个案件审理决定，其中 75% 推翻了 Meta 对相关内容的决定——略高于 2021 年的 70%。

我们还对审查流程做出了一些改变，以便加强我们的影响力。2022 年 7 月，我们宣布将首次在一份决定中同时处理两起关于性别认同和裸露内容的案件。通过同时处理多起案件，我们将有机会探索更广泛的用户体验，比较不同擦边帖子之间的差异。2022 年 10 月，我们还扩大了我们的权限范围，有权做出具有约束力的决定，要求 Meta 在保留或恢复符合条件的内容时将内容标记为“令人不安”或“敏感”。11 月，我们发布了“英国 Drill 音乐”案件审理决定。我们在做出这个决定的过程中向伦敦警察厅提交了信息自由法请求，了解它向社交媒体公司要求删除内容的一些新细节。



政策咨询意见

除了审核单个案件并决定删除或恢复内容，我们还可以应 Meta 要求提供政策咨询意见。我们在这些意见中向 Meta 提供如何修改特定主题政策的详细建议。我们相信，这些建议对于那些经常和 Meta 面临类似内容审核问题的其他公司也会有所帮助。

我们在 2022 年发布了第一批政策咨询意见，内容涉及私人住所信息分享、Meta 交叉检查计划等。有关交叉检查计划的政策咨询意见以前所未有的详细程度披露了计划内容，向 Meta 提出了 32 条建议。

META 对委员会问题的回答情况

为了更好地做出案件审理决定和政策咨询意见，我们会在进行审议前向 Meta 提出一些问题。针对我们在 2022 年发布的 12 项案件审理决定和 2 份政策咨询意见，我们向 Meta 提出 356 个问题。在这些问题中，Meta 完全回答了 306 个，部分回答了 37 个，未回答 13 个。

2022 年，Meta 完全回答的问题在全部问题中占比 86%，与 2021 年相同。但是，Meta 没有回答的问题占比从 2021 年的 6%（313 个问题中的 19 个）下降到 2022 年的 4%（356 个问题中的 13 个）。我们认为 Meta 在 2022 年的回答比 2021 年更全面。如果无法回答问题，Meta 也比 2021 年更经常地解释为什么会出现这种情况。

经验教训

2022 年，我们在细则中规定，委员会原则上应在网站上公布案件后 90 天内发布案件审理决定。细则中还规定了可以超出此期限的例外情况，比如发生了特殊情况或者技术、运营事故。2022 年，因例外情况延期做出决定的案件多于如期做出决定的案件。有些是因为花费了比预期更长的时间与 Meta 协商能够在最终决定中列入多少信息（Meta 一开始提供这些信息时附了保密义务），有些是因为翻译问题推迟了发布时间，还有一些是因为委员会日程限制和其他问题导致委员会不能在 90 天内发布决定。我们的目标仍然是在 90 天内发布案件审理决定（如 2023 年 2 月细则更新中所述），并在案件时间表方面保持透明。

虽然委员会在 2022 年发布了 2 项政策咨询意见，但也在同年 5 月宣布 Meta 已告知将撤回之前要求就当前俄乌冲突相关内容审核问题提供政策指导的请求。Meta 这么做的时候列举了一些具体、长期的安全方面的担忧。我们理解 Meta 的这些担忧，但认为之前的请求提出了一些十分重要的问题，因而对 Meta 撤回请求的决定感到失望。不过，我们随后选择“俄语诗”案件并发布了案件审理决定，在其中涉及了与俄乌冲突有关的几个内容审核问题。

2022 年案件审理决定与政策咨询意见

案件审理决定

	平台	资料来源	社群守则	国家/地区	委员会决定
咨询阿得拉 (Adderall®) 获取途径 案件编号：2021-015-FB-UA		用户	受限商品和服务	美国	已推翻 Meta 的删除决定
瑞典记者报道未成年人遭受性暴力 案件编号：2021-016-FB-FBR		Meta	成人裸露和性行为	瑞典	已推翻 Meta 的删除决定
克宁动画片 案件编号：2022-001-FB-UA		用户	仇恨言论	克罗地亚	已推翻 Meta 的保留决定
苏丹图像视频 案件编号：2022-002-FB-MR		Meta	暴力和血腥内容	苏丹	支持 Meta 的保留决定
给某些阿拉伯词语重新赋予积极意义 案件编号：2022-003-IG-UA		用户	仇恨言论	摩洛哥、埃及、黎巴嫩	已推翻 Meta 的删除决定
影射哥伦比亚警察的漫画 案件编号：2022-004-FB-UA		用户	危险人物和组织	哥伦比亚	已推翻 Meta 的删除决定
新闻报道中提及塔利班 案件编号：2022-005-FB-UA		用户	危险人物和组织	阿富汗	已推翻 Meta 的删除决定
提格雷通讯事务管理局 案件编号：2022-006-FB-MR		Meta	暴力与煽动暴力	埃塞俄比亚	支持 Meta 的删除决定
英国 Drill 音乐 案件编号：2022-007-IG-MR		Meta	暴力与煽动暴力	英国	已推翻 Meta 的删除决定
俄语诗 案件编号：2022-008-FB-UA		用户	仇恨言论	拉脱维亚、乌克兰、俄罗斯	已推翻 Meta 的删除决定
尼日利亚教堂遭袭后的视频 案件编号：2022-011-IG-UA		用户	暴力和血腥内容	尼日利亚	已推翻 Meta 的删除决定
印度性骚扰视频 案件编号：2022-012-IG-MR		Meta	成人性剥削	印度	支持 Meta 的保留决定

政策咨询意见

分享私人住所信息 案件编号：PAO-2021-01	Meta 的交叉检查计划 案件编号：PAO-2021-02
------------------------------	----------------------------------

决定和政策咨询意见摘要

2022 年 10 月，我们对用户申诉中最常提出的问题和担忧进行了深入分析，并以此为依据宣布了七项战略重点。本节将根据这些重点，对我们在 2022 年发布的决定和政策咨询意见进行分组。

选举与公民空间



新闻报道中提及塔利班

[已推翻]

2022 年 1 月，印度一家受欢迎的乌尔都语报纸在其 Facebook 公共主页上发布了阿富汗塔利班政权官方发言人的一则声明，宣称女子学校和女子大学将在 2 个月内开学复课。Meta 以违反 Facebook 的“危险人物和组织”社群守则为由删除了该帖。该守则禁止“赞扬”“参与严重的现实伤害事件”的实体，包括恐怖组织。Meta 还对公共主页管理员处以“违规记分”，并限制该帐户使用某些 Facebook 功能。

在该用户对 Meta 的决定提出申诉后，第二名人工审核员认可 Meta 的初步评估，依然判定该帖违规。但当我们选择对该案件进行复审后，Meta 承认最初的删帖决定构成“执行错误”，随后恢复了该内容，并取消了对用户的违规记分和对用户帐户的限制。尽管 Meta 主动纠错，但我们还是推翻了 Meta 关于删除塔利班政权这条积极声明的最初决定。我们引用了“危险人物和组织”社群守则中的一条例外规定，即允许发布有关恐怖组织活动的新闻报道。我们还认为，有关危险人物和组织的政策中允许“报道”的例外规定与更笼统的“允许保留具新闻价值内容”例外规定之间的关系还不够明确。

“委员会认为，尽管塔利班是一家危险组织，但由于此事涉及公共利益，所以与之有关的新闻报道不应受到限制”

监督委员会成员
Paolo Carozza



2022 年，
我们的决定涉及
最多的社群守则
是哪些？

仇恨言论
3 项决定

危险人物和组织

2 项决定

暴力与煽动暴力

2 项决定

暴力和血腥内容

2 项决定



瑞典记者报道未成年人遭受性暴力

[已推翻]

2019 年 8 月，瑞典的一名用户在其 Facebook 公共主页发布了一张照片。照片中，一名小女孩双手抱头坐着，脸被遮挡住。该照片附带的文字详细描述了两名未提及姓名的未成年人被强奸的事件。该帖列出了两名被定罪的犯罪者年龄，但未披露其姓名，并以生动的方式详细描述了一名受害者受侵犯的细节。此外帖子还引述了其中一名罪犯炫耀强奸行为的言论，以及罪犯提及未成年人时使用的色情露骨语言。2021 年 9 月，在该帖发布 2 年后，Meta 以违反“儿童性剥削、虐待和裸露”政策为由将其删除。

我们认为，Meta 的删帖决定是错误的。传达此强奸案后果的客观描述和罪犯对强奸行为的露骨描述并不等同于“性剥削儿童或在色情语境中描述未成年人”的语言。这篇帖子所在的大背景表明，该用户发帖的意图是在报道一个公众显著关注的问题，并谴责针对未成年人的性剥削行为。我们在针对这一案件提供的建议中，敦促 Meta 明确区分某条内容究竟是对儿童性剥削表达赞同和宣扬，还是旨在提高人们对这一问题的意识。

危机和冲突局势



提格雷通讯事务管理局

[支持]

2022 年初，Meta 提请复审一个关于提格雷州通讯事务管理局官方 Facebook 公共主页上发布的一个帖子的案件。该机构由埃塞俄比亚的一个州政府运营。该帖是在政府军队与提格雷军队在该地区持续爆发冲突时发布的，其中提到了政府军近期遭受的损失，还呼吁政府军“turn its gun”（将枪口指向）“Abiy Ahmed group”（Abiy Ahmed 一党）。Abiy Ahmed 是埃塞俄比亚的总理。该帖还对拒绝遵守这一警告的联邦军队发出死亡威胁。

在该帖遭到多名用户举报且被 Meta 的自动化系统判定为违规内容后，两名讲阿姆哈拉语的审核员认为该帖并未违反 Meta 的政策。随后该帖被提交给 Meta 为审核冲突中出现的内容而设立的诚信产品运营中心 (IPOC)，进行专家审查。两天后，Meta 推翻了最初的决定，并以违反 Facebook 的“暴力与煽动暴力”政策为由删除该帖。

虽然我们承认 Meta 采取了积极措施（包括建立 IPOC 系统）来监控高风险冲突局势中的内容滥用问题，但我们在决定中强调，这条违规内容在平台上展示了 2 天，浏览量超过 30 万次，然后才被删除。仅就

此事来看，这至少表明 IPOC “并非是用来应对多年冲突的长期可持续解决方案”。对此，我们建议，“Meta 可能需要投资建立一个更加可持续的机制”，以履行其在冲突地区和危机局势中的人权责任。



苏丹血腥暴力视频

[支持]

2021 年最后几个星期，Meta 提请复审一个案件：在 2021 年 10 月苏丹军事政变后某 Facebook 个人主页上发布的一段血腥视频。在该视频中，一个人头部严重受伤，一颗眼珠掉了出来，躺在一辆汽车旁边。从背景声中可以听到，有人用阿拉伯语说有人遭到殴打后被扔到街上。字幕也用阿拉伯语呼吁人们不要相信军队，称其长期存在虐待行为。

在 Meta 以违反 Facebook 的“暴力和血腥内容”社群守则为由删除该帖后，该公司发布了“允许保留具新闻价值内容”的政策。在延误近 5 个星期后，Meta 恢复了这条内容，并添加了警告画面，限制浏览量。

我们的决定支持 Meta 恢复内容并附带警告画面的决定。我们还认为，想要规模化审核 Facebook 上分享的血腥内容，应用“允许保留具新闻价值内容”的规定并非有效手段。为强调这一结论，我们引用了 Meta 自己的说法：在 2021 年前三个季度，记录了与暴力和血腥内容政策有关的 17 条因具有新闻价值而被保留下来的内容。相比之下，他们在同一时期据此社群守则删除的内容超过 9000 万条。因此，我们建议 Meta 修改其社群守则有关暴力和血腥内容的规定，允许用户“为提高人们对暴行的认知或记录暴行”而分享血腥视频。



俄语诗

[已推翻]

2022 年 4 月，一名位于拉脱维亚的 Facebook 用户发布了一张横尸街头的照片。Meta 证实，照片上的死者系在乌克兰遭到枪杀。图片随附的俄文说明称，在第二次世界大战期间，苏军士兵在德国犯下的种种罪行指控之所以被原谅，是因为这是他们对德国纳粹士兵在苏联犯下的累累罪行进行的复仇。在列举俄军士兵在乌克兰犯下的当代暴行时，该帖引用了苏联诗人康斯坦丁·西蒙诺夫 (Konstantin Simonov) 的诗句：“Kill the fascist...Kill him!”（杀死法西斯……杀死他！）

在 Meta 以该帖违反“仇恨言论”社群守则为由将其删除后，我们选择对该案件进行复审。Meta 随后将其恢复，并添加了警告画面。我们在决定中对两类内容进行了区分：一类内容以俄罗斯士兵的国籍为攻击目标，构成违规；另一类则与纳粹行为进行历史对照，并不违规。我们发现，引用苏联时代的诗歌是出于艺术和文化参照目的，旨在对一种精神状态进行描述，而非鼓励。我们承认，在国际冲突局势中评估暴力言论非常复杂，因为国际法允许战斗人员成为攻击目标。尽管如此，我们还是建议 Meta 修改政策，将“非法军事干预”的背景明确纳入考虑范围。

针对边缘化群体的仇恨言论



克宁动画片

[已推翻]

2021 年 12 月，一个 Facebook 公共主页发布了迪士尼动画短片 “The Pied Piper”（《花衣魔笛手》）的剪辑版本，并配上了克罗地亚语的说明文字，Meta 将这段文字翻译为 “The Player from Čavoglave”（Čavoglave 的魔笛手，其中 Čavoglave 是克罗地亚的一个村庄）“and the Rats from Knin”（与克宁的老鼠，其中克宁是克罗地亚的一个城市）。旁白说道老鼠誓要生活在一个 “pure rat country”（只有老鼠的国家），于是它们开始骚扰和迫害城里的人类，占领这座城市。后来，一位来自 Čavoglave 的魔笛手用他的 “魔笛” 吹奏了一段旋律，把老鼠引到城外的一辆拖拉机上，然后，拖拉机消失了。旁白总结道：“the rats disappeared forever from these lands...[and] everyone lived happily ever after.”（老鼠从这片土地上永远消失了，从此大家都过上了幸福的生活。）

即使收到近 400 次用户投诉，Meta 却并未删除这段内容。在委员会收到该案件的申诉后，Meta 进行了额外的人工审查，但依然认为该内容并不违规。当我们选择对该案件进行复审时，Meta 推翻了原来的决定，认为该帖虽然没有违反 “仇恨言论” 政策的字面规定，但的确违反了该政策的宗旨，并将其删除。而在起草提交给监督委员会的说明文件时，Meta 再次改变说法。这一次，他们认为该帖也违反了这项政策的字面规定。

在发现该内容的确违反了 “仇恨言论” 和 “暴力与煽动暴力” 社群守则后，我们引用了一些提及 1995 年克罗地亚军事行动 “风暴作战” (Operation Storm) 的评论。这一行动导致克罗地亚的塞尔维亚平民流离失所、惨遭处决、失踪失联。有鉴于此，我们感到担忧的是，接近 40 名克罗地亚语审核员认为此内容没有违规，因为他们一直理解为必须在塞尔维亚人和老鼠之间发现明确的贬义才算该内容违规。我们建议 Meta 进一步解释说明 “仇恨言论” 社群守则，明确表达该政策禁止以含蓄的和明确的方式恶意提及受保护群体。



给某些阿拉伯词语重新赋予积极意义

[已推翻]

2021 年 11 月，一个自称要为 “阿拉伯文化中的 ‘酷儿’ 词语” 提供讨论平台的 Instagram 公开帐户发布了一组图片，并用阿拉伯语配文解释称，每张图片中都包含一个阿拉伯语文化中贬损 “举止柔弱” 的男性的词。该用户称其发帖的目的在于 “to reclaim [the] power of such hurtful terms”（给这些伤害性词语重新赋予积极意义）。

Meta 先是以违反 “仇恨言论” 为由删除该内容，在收到用户申诉后，又将其恢复。但在收到另一名用户投诉后，Meta 又再次将其删除。我们选择对该案件进行复审。Meta 决定对该内容进行额外审核，最终认定其没有违反政策。Meta 随后将其恢复，解释称最初的删除决定仅仅依据对包含贬损用语的图片的审核结果，但没有考虑到内容本身或作者对其意图的解释。

在发现该删帖决定与“仇恨言论”政策不符时，我们指出，该内容属于“出于自述或自我激励而使用禁用词语”的例外情况。换言之，Meta 允许用户以谴责或提高公众意识为目的引用仇恨言论。我们认为，Meta 在本案件决策过程中的反反复复，表明他们没有始终如一地“将其仇恨言论政策中的豁免情形应用于边缘化群体发布的言论”。为解决这一问题，我们建议 Meta 将其提供给审核员的“内部实施标准”和“已知问题”翻译成现代标准阿拉伯语。

政府使用 Meta 平台



英国 Drill 音乐

[已推翻]

2022 年 1 月，一个自称致力于宣传英国音乐的 Instagram 帐户发布了说唱歌手 Chinx (OS) 最近发行的“英国 Drill 音乐”歌曲《Secrets Not Safe》的音乐视频片段。英国 Drill 音乐是一种极富本地特色、草根流派、说唱音乐的分支，在英国城市黑人青年中流行。不久后，负责在大伦敦地区执法的伦敦警察厅 (Metropolitan Police) 发邮件给 Meta，请求 Meta 审核所有包含《Secrets Not Safe》的内容。警方还额外提供了与伦敦地区帮派暴力盛行有关的资料，他们担心这段音乐可能会增加报复性帮派暴力活动发生的风险。

“Meta 不缺价值观和守则，缺乏的是执行”

监督委员会成员
Endy Bayuni



Meta 的一个专家小组根据警方提供的信息做出回应，认定这个帖子构成“隐藏威胁”。公司得出此结论是因为歌曲提到了 2017 年的一起枪击事件，在他们看来这增加了这首歌曲煽动暴力的可能性。Meta 不仅以违反“暴力与煽动暴力”政策为由从该帐户中删除这一内容，还删除了其他帐户中的另外 50 多条包含该歌曲的内容，其中也包括 Chinx (OS) 本人的帐户。此后，Meta 的自动化系统还对包含该歌曲的内容进行了另外 112 次删除操作。

在 Meta 将该案件提请委员会复审后，我们要求复审 Chinx (OS) 的原始帖子。Meta 回应称，他们为从整个平台上删除该视频和歌曲所采取的行动导致 Chinx (OS) 的帐户被删除。在判断 Meta 的行动是否符合其守则、价值观和人权责任时，该帖子的背景至关重要：

“[说唱] 歌手们经常以第一人称的叙事方式细致讲述正在发生的街头暴力冲突，并用图像和歌词描绘或描述暴力行为。潜在的暴力描述和虚张声势的表演被认为是该流派的一部分，这种艺术表达形式可能会模糊事实与虚构之间的界限。

这类描述是歌手们的竞争法宝，他们借此来引起共鸣和积攒人气。关于 Drill 音乐是否会引起现实世界中的暴力，尤其是这方面争论中所提到的所谓证据的可靠性，是有争议的。”

在评估 Meta 的决定是否符合其规则、价值观和人权责任时，我们认为缺乏足够的证据来支持该内容构成“实质性威胁”的观点。我们进一步发现，“在缺乏此类证据的情况下，Meta 应该更加重视该内容的艺术性。”我们的决定指出，“虽然执法部门有时可以提供背景信息和专业意见，但并不是执法部门希望移除的每一条内容都应该被移除。”

我们敦促 Meta 在未来收到这种请求时，采取切实措施“独立评估这些请求，特别是当它们涉及到少数群体或边缘群体成员的艺术表达时，因为他们的内容极有可能遭受文化偏见。”



我们的决定还凸显出一个问题：Meta 在回应政府和执法部门的请求时所遵循的流程缺乏透明度。我们引用了向伦敦警察厅提出信息自由请求后收到的回应，结果显示，在截至 2022 年 5 月的一年中，该警察厅共向社交媒体公司和流媒体服务提出了 286 项审核和删除音乐内容的请求，均与英国 Drill 音乐有关。但在 2023 年 1 月 4 日，伦敦警察厅联系我们称，他们在上述回应中发现错误，并予以纠正。在截至 2022 年 5 月的一年中，伦敦警察厅实际发出的请求并非 286 项，而是 992 项，均与英国 Drill 音乐有关。

公平对待用户



关于分享私人住所信息的政策咨询意见

2022 年 2 月，我们发布了首份政策咨询意见 (PAO)。委员会利用政策咨询意见审核 Meta 的政策，并就如何改进这些政策提出建议。在本案件中，Meta 请求委员会详细评估它可能需要对 Facebook 社群守则中有关侵犯隐私的规定做出哪些更改。除特定情况外，该守则禁止分享“与自己或他人有关的个人身份识别信息”。根据该政策条款，用户不得分享包括住址在内的“个人身份识别信息”，但“为了推广慈善事业、寻找失踪人员/动物/物品或联系商业服务提供商而分享或索取此类信息”的情况除外。

为了保护用户隐私，该政策禁止分享“展示私人住所外观的图像”，尤其是该住所为独栋住房，从图像/说明文字中可识别住所的门牌号，当事住户“反对暴露其私人住宅”，或者“有人在组织针对当事住户的抗议活动”。如果该住所是政府大使馆，可能成为抗议活动的焦点时，则属例外情况。

Meta 请我们帮忙解决的一个潜在问题是，获取这类信息对记者和社会活动家非常重要。相反，在未经住户同意的情况下暴露这些信息，可能会“给当事住户造成安全隐患以及侵犯个人隐私”。Meta 列举了用户的安全和隐私面临的多项潜在的现实伤害，其中包括“人肉搜索”。“人肉搜索”是指在未经授权的情况下发布文件，故意向可能

滥用个人信息的人透露个人信息。Meta 指出，人肉搜索会在现实世界中造成骚扰、跟踪、暴力和死亡等不良后果。

如果私人住宅信息可能已经公开发布，包括该信息已经通过“新闻报道、法庭文件、通讯稿或其他信息源公开”，则视为该政策的例外情况。Meta 在向内容审核员提供的内部指南指出，信息如果“至少经五家新闻媒体发布过”，则不再属于政策中规定的隐私信息。

我们同意 Meta 的观点，认为侵犯住所隐私权可能在现实世界中引发严重伤害，妇女、儿童和 LGBTQIA+ 人群等造成的伤害尤其严重。我们也认同 Meta 的另一个观点：在现实世界访问公开记录和其他“公开可用”的信息来源需要投入资源，付出努力，但在数字平台上获取这些信息要容易得多，而且更容易在更大范围内分享。鉴于此类伤害的严重性，我们建议 Meta 对允许分享“可公开访问”的信息的情况进行限制。



关于 Meta 的交叉检查计划的政策咨询意见

2021 年 10 月，继《华尔街日报》(Wall Street Journal) 披露 Meta 的交叉检查计划后，我们接受了该公司关于审议交叉检查计划并就其可以如何改进提供建议的请求。在此背景下，Meta 指出公司每天对内容做出的政策执行活动大约为 1 亿次。面对如此巨大的工作量，即使 Meta 的内容决定能够达到 99% 的准确率，每天也仍然会有 100 万次误判。

由于在 Facebook 和 Instagram 上发布的内容不仅数量多，而且性质复杂，这对构建能够履行 Meta 人权承诺的系统提出了挑战。要构建一套能够履行 Meta 人权承诺的系统确实需要面临艰巨的挑战。虽然认识到这一点，但我们还是指出，以现有结构来看，交叉检查并不符合该公司的人权承诺。多年以来，某些政治家、商业伙伴、名人和其他人发布在 Facebook 和 Instagram 上的内容原本应当很快就被删除，但却因为交叉检查计划而得以保留数日。Meta 向我们透露，平均而言，他们需要 5 天以上才能对交叉检查名单上的用户发布的内容做出决定。

Meta 坚称，该计划的初衷是“推动人权承诺的履行”。但我们经过细致调查后发现，这项计划似乎“更多地是直接按照满足商业需求的方式来设计的”。我们在审核中发现的最突出的缺陷在于，该公司在删除内容时，为选定的用户赋予比其他用户更大力度的保护，进一步放大了原本就普遍存在的担忧，即 Meta 为名人以及政治和商业领袖赋予凌驾于其他数百万人之上的特权。因此，我们敦促该公司解决这一问题。

如果某篇帖子被认定违反了公司的规定，但是它是由 Meta 交叉检查名单中的用户发布的，则它仍将保留在平台上，等待进一步审核。之后，Meta 会对照其所有政策来审核这篇帖子，包括所有特例规定和

针对特定背景的规定，而这很可能会增大内容保留在平台上的概率。相比之下，对于普通用户发布的内容，到达能够对照 Meta 所有规定进行内容审核的审核员手中的可能性要小得多。

在决定哪些实体可以入选交叉检查名单，使其内容获得额外保护时，Meta 究竟遵循了哪些标准？由于在这方面缺乏透明度，导致这种不公平的待遇进一步加剧。虽然 Meta 在将商业伙伴和政府领导人加入交叉检查名单时有明确的标准，但在人权保护方面扮演重要角色的群体（如记者或民间团体领导者）却缺乏明确的入选渠道。

这种缺乏透明度的情况还延伸到其他方面：Meta 并未披露它通过哪些核心指标来衡量该计划的有效性。至于 Meta 是否对比过交叉检查机制下和正常质量控制机制下的决定准确度，该公司没有提供证据来证明这一点。此外，Meta 也一直未能告知用户他们是否入选交叉检查名单，未能披露这些名单的创建和审核程序，同样未能披露“持续发布违规内容的实体因为他们的身份继续留在名单中”的案件。

我们为了解决这些问题提出了 32 条建议。这些建议敦促 Meta 在其交叉检查名单中应该优先考虑具有重大人权意义的言论，包括“对公众有特殊重要性”的言论。我们还呼吁 Meta 从根本上提高该计划及其运作方式的透明度，并提议减少因接受“进一步审核”而保留的违规内容造成的伤害。Meta 于 2023 年 3 月回复了我们的建议，承诺提高透明度，并为那些特别容易遭受过度执行的人提供更多保护，包括记者和人权维护者。



咨询阿得拉 (Adderall®) 获取途径

[已推翻]

2021 年 6 月，一位被确诊患有注意力缺陷多动障碍 (ADHD) 的美国 Facebook 用户，在非公开小组中发帖询问如何与医生讨论药物治疗问题。该用户称医生为其开具的是处方药阿普唑仑，但其过去服用的阿得拉 (Adderall) 很有效。该用户担心，如果直接向医生索要不同的处方药，可能会表现为觅药行为。

2021 年 8 月，Meta 删除了该帖子，理由是它违反了 Facebook 的“受限商品和服务”社群守则。在我们选择复审该案件后，Meta 将当时的删帖行为标记为“执行错误”，并将其恢复。

我们在决定中认为 Meta 最初的删帖决定是错误的。Facebook 的“受限商品和服务”社群守则并不禁止用户在“医疗”背景下，就特定药物寻求建议。我们还发现，该标准对酒精或药物的公开定义对用户来说较为模糊，因为即使根据其内部定义，阿得拉和阿普唑仑也要视具体情况才能确定属于非医用药物还是医用药物。我们在建议中敦促该公司“当内容层面的执行措施会引发帐户层面的惩罚时，Meta 应及时审核用户的申诉”。

自动执行 政策和内 容管理



影射哥伦比亚警察的漫画

[已推翻]

2020年9月，哥伦比亚的一位 Facebook 用户发布了一幅酷似哥伦比亚国家警察官方徽章的漫画。画中有三个身穿警服的人将警棍举过头顶，显然是在拳打脚踢一个头破血流、倒地不起的人。徽章上用西班牙语写着一串字，Meta 将其翻译为：“National Police – Republic of Colombia – Baton and Kick”（国家警察 – 哥伦比亚共和国 – 暴力之师”。

在这条内容发布 16 个月后，Meta 移除了该内容，因为该图片与其某个媒体匹配服务库中的一张类似图片相匹配。这些库可以自动确定和删除经人工审核员认定为违规的图片。在我们选择复审该案件后，Meta 认为该帖并未违反规则，并将其恢复。公司还恢复了其他包含该漫画的内容，并承认这些内容是被媒体匹配服务库误删的。

“内容决定和内容的算法处理不可分割”

监督委员会联合主席
Catalina Botero-Marino



通过这个案件可以明显看出，当对库中的内容做出的错误决定后，其影响会被这些库进一步放大。尽管有 215 名用户对删除决定提出申诉，且有 98% 申诉成功，但 Meta 并未从库中删除该漫画，也没有恢复类似的内容，直到我们选择复审该案件。我们敦促 Meta “紧急改进其程序，迅速从这些库中移除未违规内容。”



尼日利亚教堂遭袭后的视频

[已推翻]

2022年6月5日，恐怖分子袭击了尼日利亚西南部的一座天主教堂。几小时后，一位尼日利亚的 Instagram 用户发布了一段视频，显示血迹斑斑的尸体一动不动地躺在教堂地板上，其中有些人的面部是可见的。在 Meta 的一个媒体匹配服务库发现该视频与一位人工审核员认定的违规视频相似之后，自动内容“归类技术”决定将其保留，并插入了“该内容可能令人不适”的警告画面。

几天后，当该内容被第二个媒体匹配库审核时，该用户增加了一段英文说明，并将该事件形容为“sad”（令人痛心）。该用户还增加了话题标签，其中一些提到了真人电子游戏《airsoft》，还有一些通常被用于向收藏家推销枪支的标签。Meta 认为，虽然视频本身没有违规，但话题标签却存在违规行为，因为它们可能被解读为“美化暴力，淡化受害者遭受的痛苦”。委员会的大多数成员都不同意 Meta 的决定，认为最初恢复内容并配上警告画面的做法是正确的。多数派得出此结论是因为他们认为不应该“仅仅因为这些话题标签与枪支使用者有关，”就判定这些话题标签属于残酷言论。少数派不同意这一看法，他们指出，使用与枪击有关的话题标签可以被理解为“很残酷，可能会给幸存者或受害者家人带来精神创伤”。注意到这种模糊性后，我们建议 Meta 对“暴力和血腥内容”政策中的用语进行复审。



2022年3月，一个自称是“达利特人资讯平台”的Instagram帐户发布了一条在印度拍摄的视频，其中可以看到一名部落女性正在被多名男子侵犯。“Dalit”（达利特人）以前被称为“untouchables”（贱民），曾饱受印度种姓制度的压迫。尽管视频中没有出现裸露画面，而且无法通过视频确认这位女士的身份，但Meta还是以违反“成人性剥削”政策为由将其删除。后来，Meta的内部团队改变了决定，在执行“允许保留具新闻价值内容”规定并添加警告画面后，恢复了这段内容。理由是，该视频的内容意在提高人们的认识，帮助人们了解印度长期以来对“部落妇女”的歧视。

我们认为，Meta执行“允许保留具新闻价值内容”规定并添加警告画面后恢复该内容是正确的决定。尽管如此，我们还是重申了自己的担忧，认为对违规内容执行“允许保留具新闻价值内容”规定并不是“大规模处理此类案件”的正确方式。

“‘允许保留具新闻价值内容’的规定不够清楚明确，给执行者留下了极大的斟酌空间，且无法保证政策得到大规模的统一执行。此外，这条规定也没有明确的标准来评估违反‘成人性剥削’政策的内容所造成的潜在危害。监督委员会认为，Meta的人权责任要求其必须为此类案件提供更明确的标准以及更有效的执行流程。”

关于 META 在执行和回应我们决定的及时性方面的报告

- ⊙ 根据委员会的《细则》，Meta 必须在我们的决定公布后 7 天内予以执行。
- ⊙ 在我们 2022 年公布的 12 项决定中，Meta 都在 7 天内恢复或删除了相应的内容，除非内容已被恢复。
- ⊙ 我们将通过落实委员会（目前由 5 名委员会成员组成）继续敦促 Meta 提高透明度，帮助外界了解其如何识别和处理与我们的决定具有相似内容和背景的情况。这将确保我们的决定不会局限于具体案件，而是可以推广到具有类似背景的相关内容。

将国际人权标准应用于内容审核： 第 19 条

委员会工作最重要的主题是，确信 Meta 只要能遵守承诺，以国际人权标准作为基础，就能以更公平、更有原则的方式做出内容审核决定。为此，我们的《章程》规定，我们将“根据保护言论自由的人权准则，特别关注删除内容产生的影响。”《公民权利和政治权利国际公约》(ICCPR) 第 19 条就是这些准则中的一项，它规定，虽然“人人享有言论自由的权利……但行使该权利可能……受到一些限制，但仅限于……法律规定且必要的情况下。”第 19 条提供了评估言论限制的三步测试法：



1. 该限制是否符合合法性 原则？

我们研究了 Meta 做出决定时所依赖的规定是否便于用户获取，并且通过足够明确的表达让用户理解和遵守。确保规定的明确性非常重要，因为只有这样才能让负责执行的人做出公平、一致的决定。

2. 提出的限制是否有合法目的？

我们要评估一项决定所依据的规定是否在追求目标时兼顾权利。

3. 提出的限制是否必要且相称？

删除该内容是否是实现目标最温和的手段？
该限制是否与被保护的利益相称？



“企业必须尊重言论自由权，包括不分国界地传递和接受各种思想。”

监督委员会联合主席
Evelyn Aswad



委员会决策流程中的国际人权标准

2021 年，Meta 发布了《公司人权政策》，阐述了它基于《联合国工商企业与人权指导原则》(UNGPs) 做出的人权承诺。这份 2011 年的框架为私有公司提供了新的指导，帮助他们肩负起尊重人权的责任。

Meta 在其 2022 年发布的首份《人权报告》中，将建立委员会作为其为人权影响提供补救途径所采取的一项措施。《联合国工商企业与人权指导原则》呼吁企业提供“补救途径”，包括建立“有效的运营层面申诉机制，供可能受到（其业务）不利影响的个人和社群使用。”

作为一项运营层面的申诉机制，委员会竭力体现《联合国工商企业与人权指导原则》中规定的有效性标准，包括使我们的工作合法、易懂、公平、透明、兼顾权利、成为持续的学习源泉。我们通过选择案件、发布决定和提供建议，促使 Meta 尊重所有人的权利。在每项决定中，我们都对案件中涉及的人权影响和引发的担忧进行了详细分析。

委员会向联合国提交的文件

我们对 Meta 在案件和政策咨询意见中的人权责任进行了分析，并以此为基础在 2022 年向两个联合国机构提交了广泛的材料，阐述了我们将人权原则应用于数字平台的内容政策和内容决定时采用的方法是如何演变的。

在我们 2 月提交给联合国人权事务高级专员办事处关于“《联合国工商企业与人权指导原则》在科技公司活动中的实际应用”的文件，以及 7 月提交给联合国意见和言论自由特别报告员的文件中，我们重点讨论了《联合国工商企业与人权指导原则》中提出的“责任”和“补救”这两个主题。

我们提交的这两份文件都强调，社交媒体在冲突期间的内容决定会对人权产生严重影响。这一重点反映了我们对《联合国工商企业与人权指导原则》的重视，认为企业应当“加大努力，在后果最严重的领域尽可能降低人权风险”。根据《联合国工商企业与人权指导原则》，对于分配给审议小组的每个案件，我们都会进行细致的人权影响评估 (HRIA)，其中可能包含“冲突敏感性分析”，以此确保委员会能履行其尽职调查责任。此外，我们的案件甄选团队制定了专门的程序，可以提醒 Meta 注意可能需要“加急关注”的“遗留”内容，以减轻严重伤害的风险。

“我们不断提醒 Meta，人权原则应当成为其社群守则的核心”

监督委员会成员
Nighat Dad



来自阿富汗、埃塞俄比亚和苏丹的案件研究

我们在提交给联合国特别报告员的文件中，提到了 2022 年的 3 个说明性案件：“苏丹血腥暴力视频”、“新闻报道中提及塔利班”和“提格雷通讯事务管理局”。这三个案件都反映出 Meta 在“为审核冲突地区的内容建立一个原则清晰、公开透明的内容审核系统”方面缺乏进展。我们认为，制定有效的危机应对规程，将极大地提高该公司在危机和冲突局势下坚守其人权责任的能力。

我们在“苏丹血腥暴力视频”决定中，强调了这些内容发布时所处的危机和冲突背景：“警卫队……以记者和活动家为目标，搜查他们的住宅和办公室。记者遭到袭击、逮捕和拘留……由于军方接管国家媒体以及打压苏丹报纸和广播公司，社交媒体成为记录军方暴行的重要信息来源和空间。军方……逮捕平民领导人并切断了互联网。”

我们使用第 19 条的三步测试法，对 Facebook 的“暴力和血腥内容”政策提出质疑，因为它没有澄清如何允许用户“为提高人们对暴行的认知或记录暴行而分享血腥视频”。虽然 Meta 公司正确地将执行“允许保留具新闻价值内容”规定作为恢复有关内容的主要理由，但我们指出，该规定未能明确界定相关术语。至于在内容中增加警告画面这样的补救措施，我们认为此举构成了“对言论自由的必要且相称的限制”，从而“充分保护了所描述的个人及其家庭的尊严”。

在第二个案件，即“新闻报道中提及塔利班”中，Meta 删除了一个转述塔利班官方发言人宣布将在阿富汗开设女子学校的帖子。我们认为，Meta 最初决定删除该帖子的做法侵犯了用户的言论自由。这是因为它剥夺了阿富汗的 Facebook 用户“对关乎公共利益的事件的知情权……特别是当一个指定的危险团体强行推翻公认的政府



时。”我们补充道，塔利班政权对媒体自由的限制使得“国际报道的作用愈加重要”，因为“这些信息……对关注女童和妇女平等受教育权利的人来说至关重要。”

在第三个案件“提格雷通讯事务管理局”中，Meta 删除的帖子中包含提格雷州政府内的一个官方消息来源呼吁政府军反击中央政府军的内容。我们认为，Meta 的删帖决定符合其人权责任。为了支持那项决定，我们采用了拉巴特行动计划中的六因素测试法，这种方法就如何在保护言论自由的同时避免人们受到歧视、敌意或暴力的煽动提供了具体指导。

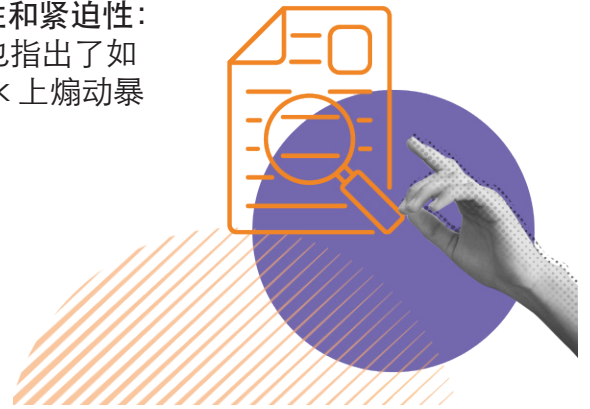
这六项拉巴特因素评估如下：(1) 背景：该内容是在持续且不断升级的内战背景下发布的。(2) 发言人：发言人隶属于冲突一方的地区政府。(3) 意图：明确呼吁杀死不投降的士兵；(4) 内容：该帖子可被解读为主张以战斗人员和政治领导人为目标，无论他们是否参与了敌对行动。(5) 传播范围：该内容被发布在与冲突一方有关的机构的公共页面上，其关注者约为 26 万。帖子在该平台上停留了两天后被删除。(6) 可能性和紧迫性：发布内容时，提格雷部队正在向提格雷以外的地区推进，总理宣布全国进入紧急状态，并呼吁平民拿起武器战斗。

“若要压制言论自由，必须阐明原因。”

监督委员会成员
Julie Owono



我们还使用拉巴特测试来支持我们在 2022 年“克宁动画片”案件中的决定，本案涉及的动画片将克罗地亚某城市的塞尔维亚人描绘成老鼠。(1) 背景：某地区最近经历了种族冲突和对少数民族的歧视。(2) 发言人：一个以反塞尔维亚人观点著称的克罗地亚新闻门户。(3) 意图：煽动民族仇恨。(4) 内容：由于动画视频很吸引人，所以这种形式尤其有害。(5) 传播范围：该内容被浏览超过 38 万次，被分享超 540 次，收到 2400 多次互动和 1200 多条评论。(6) 可能性和紧迫性：委员会认为该帖子不太可能导致紧迫的伤害。但我们也指出了如下决定，即 Meta “可通过隐晦的方式移除 Facebook 上煽动暴力的帖子”。



互动和公众评议

作为委员会，我们深知，只有听取其他组织的意见并与之合作，才能找到应对内容审核挑战的持久解决方案。为了实现这个目标，委员会成员和员工积极参加公共活动和辩论，阐述委员会的工作，并讨论有原则的全球性内容审核方法对用户的重要意义。个人和组织也可以提交公众评议，参与到我们的决定制定过程中来。公众评议让人们可以对我们的决定各抒己见，并且分享语言、文化、政治和人权方面的专业知识。在许多情况下，公众评议也会影响我们的决定以及我们向 Meta 提出的建议。

宣传和互动

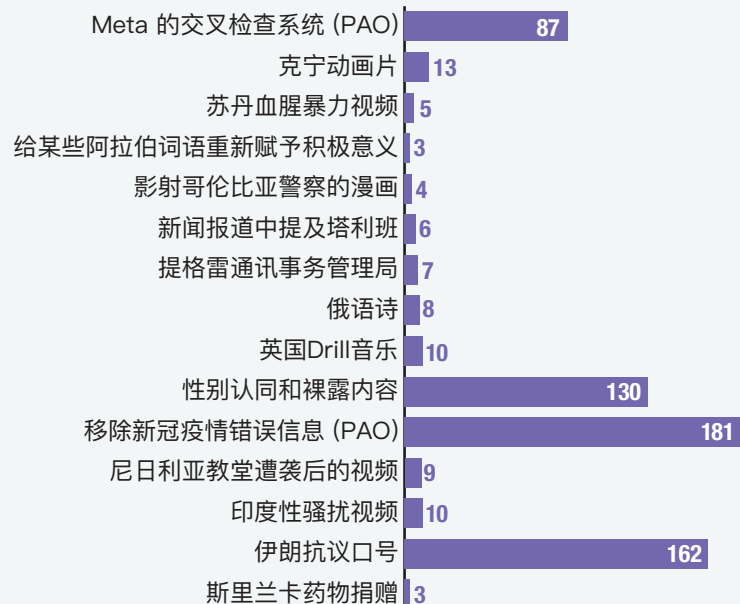
2022 年，超过 1000 名利益相关者参加了我们针对具体案件中提出的问题举办的圆桌会议。主题包括我们关于交叉检查的政策咨询意见和 Meta 针对新冠疫情错误信息制定的政策。我们还发行了一份名为《Across the Board》的简报，帮助公众更好地了解委员会的活动。2023 年，我们将继续推进这项重要工作，每月在全球各地至少召开一次圆桌会议，重点讨论我们的 7 个战略重点。通过将互动的重点集中在 7 个领域，我们也希望能从自己的关系网络以及专攻这些领域的更广泛的学术和宣传社区中受益。

公众评议

为了丰富我们的决定和政策咨询意见，我们认真考虑了个人和组织提交的公众评议。2022 年，我们共收到 600 多份公众评议，下半年收到的公众评议数量有所增加。

我们在“提格雷通讯事务管理局”决定中，参考了专家针对冲突地区的内容审核提出的宝贵见解。其中一个结论是“Meta 目前对冲突地区采用的内容审核方法可能导致出现不一致的情况”，这借鉴了几位专家提交的意见，包括 BI 挪威商学院副教授 Samson Esayas 博士。他指出，Meta 在俄乌冲突中审核内容的“措施非常迅速”，而且明显“与其他地区的冲突相比，特别是与埃塞俄比亚冲突和缅甸冲突相比，俄乌冲突获得了 Meta 的区别对待”，这两种表现之间存在明显不一致。

2022 年收到的公众评议数量，按时间顺序显示



我们的“英国 Drill 音乐”决定是根据几个组织提供的专业知识做出的。数字权利基金会 (Digital Rights Foundation) 认为，众所周知，逐行分析被删歌曲的歌词，无法准确地寻找过去的违法行为证据，也无法精确地判断未来的伤害风险。不仅如此，要核实 Drill 音乐歌词中所谓的事实陈述也并非易事。电子前沿基金会 (Electronic Frontier Foundation) 也批评了执法部门对合法 Drill 音乐的管制。我们还收到了伦敦警察厅的公众评议，这也是我们首次收到政府机构的公众评议。尽管伦敦警察厅目前不同意公布其评议，但他们表示，日后可能会在的某个时间点允许公开这一信息。

我们关于 Meta 的交叉检查计划的政策咨询意见也吸引了许多评议。民主与技术中心 (Center for Democracy and Technology)、里约热内卢技术与社会研究所 (Institute for Technology and Society of Rio de Janeiro)、美国笔会 (PEN America) 和 Mnemonic 等组织都与委员会分享了他们的专业知识。美国民主党和共和党的政治家们同样通过提交公众评议来表达他们对交叉检查的关注。

2022 年收到评议数量最多的案件已于 2023 年初公布。我们的“伊朗抗议口号”案件吸引了 162 条公众评议。公众评议证实，“marg bar Khamenei”（字面翻译为“（伊朗最高领袖哈梅内伊去死”）在最近的伊朗抗议活动中被广泛使用。他们还对 Meta 在这些抗议活动中误删波斯语内容的行为表示关切。

在“性别认同和裸露内容”案件中，我们共收到的 100 多条评议，其中有许多来自变性者、非二元性别者或顺性别女性。许多评议者指出，与这些案件中的用户一样，他们也有过被误删帖子的经历。一些评议对 Meta 为何删除包含筹款网站链接的帖子感到疑惑。在另一条评议中，从事 HIV 教育的澳大利亚非政府组织 ACON 写道，Meta 以包含性诱惑信息为由，删除了用于宣传 HIV 预防信息的帖子。研究员 Joanna Williams 对此表示赞同，她发现，在她采访的 12 家性健康组织中，有 9 家报告说 Meta 在这方面的审核对其产生了负面影响。

互联网实验室 (InternetLab) 研究中心在评议中表示，人们经常不由自主地“性化”女性、跨性别者和非二元性别者的身体，这种情况令人担忧，因为人们对顺性别男性没有类似的偏见。Zahra Stardust 博士的评议也指出，在被删除的内容中，包含女性身体的内容占比极高。最后，同性恋反诽谤联盟 (GLAAD) 和人权运动基金会 (Human Rights Campaign Foundation) 在评议中指出了如下担忧：来自边缘群体的用户发布的内容更有可能遭到重复或恶意举报，有些用户会通过举报未违规内容来骚扰发布者。

我们关于移除新冠疫情错误信息的政策咨询意见收到了 181 条公众评议。这是我们 2022 年收到评议数量最多的一条决定或意见。美国公民自由联盟 (American Civil Liberties Union) 在评议中提出了一项担忧，他们认为，由于难以大规模区分事实与虚构，以及观点、经验与对事实的断言，导致 Meta 扼杀本应允许的言论。

“我们从公众评议中寻找宝贵信息，帮助我们做出更好的决定”

监督委员会成员
Ronaldo Lemos



有几条评议指出，鉴于 Meta 的影响力及其系统在扩大错误信息传播范围方面的作用，Meta 有责任解决公共安全面临的风险。还有人担心，Meta 在处理伤害风险时采取的标记和降级措施可能还不够充分。比如，互联网与国际研究中心 (Center of Internet and International Studies) 高级副总裁提交的评议强调，添加标签不足以遏制政治家和有影响力的名人传播的错误信息。



2022 年互动活动时间线

2022 年，委员会成员、受托人和行政管理人员在全世界参加了近 100 次活动。如下互动活动时间线列出了其中的几个例子。

3 月

一位委员会代表在巴塞罗那举行的世界移动通信大会上谈到了独立数字内容监督的价值。

5 月

一位委员会成员在奥斯陆自由论坛上就社交媒体在加强民主、反抗暴政方面可以发挥的作用发表演讲。

6 月

多位委员会成员和一位受托人参加了世界领先的数字时代人权峰会 RightsCon。小组讨论的主题包括伊朗的在线言论自由和多利益相关方治理。

委员会成员齐聚加利福尼亚，首次与 Meta 高层、委员会受托人和众多利益相关者会面。

9 月

一位委员会成员在与联合国大会同时举行的美国纽约康科迪亚年度峰会上发言。

10 月

委员会成员与业界专家在纽约哥伦比亚全球言论自由会议上讨论内容审核的未来

11 月

一位委员会成员在葡萄牙里斯本年度网络峰会上发言。

一位委员会成员在黎巴嫩贝鲁特举行的面包与网络会议上，出席了中东利益相关者列席的圆桌会议。

展望未来： 2023 年及以后

改进我们与 Meta 的合作

2022 年，我们完善了与 Meta 的合作关系，发布了我们的首份政策咨询意见，并继续通过我们的决定和建议施加影响。2023 年，我们将在此坚实基础上再接再厉。根据利益相关者的反馈，我们正在继续改进自己的方法，并为我们与 Meta 的合作设定了 6 个目标。

监督委员会 2023 年目标：

1. 发布首批简易决定
2. 发布首批加急决定
3. 实现最新的委员会成员目标，追求效率最大化
4. 围绕七项战略重点，深入开展互动活动
5. 追求长期计划，扩大业务范围
6. 监督 Meta 对建议的落实情况，敦促 Meta 提供证据说明落实情况及其影响。

1. 发布首批简易决定。

在我们 2022 年最终选定的 50 个案件中，Meta 立即推翻原先决定的有 32 个，他们还在必要时对内容进行了恢复或删除。自 2020 年以来，Meta 因为委员会指出其错误而删除或恢复的帖子总数达到 80 多个，他们还为用户提供了申诉渠道，并深入揭示了该公司的内容审核方式。2023 年，我们案件甄选委员会的委员会成员将选择一些案件，以简易方式对其进行复审。届时，我们将阐述我们为何认为该案件意义重大，并指出我们在过去的决定和政策咨询意见中针对类似情况提出的建议。简易决定将由案件甄选委员会负责起草。这些决定将由案件甄选委员会（而非全体委员会）批准和投票，而且不会考虑公众评议。

2. 发布首批加急决定。

2023 年，我们预计将发布首批加急决定。这些案件将由 Meta 以加急方式提请委员会复审。我们将在几天内起草和发布决定，从而对公众关注的问题和对人权产生严重影响的情况做出贡献。审议小组将会审议、起草和批准书面决定，然后发布在我们的网站上。有关是否删除或保留某条内容的加急决定将对 Meta 具有约束力。受时间限制，这些案件将不考虑公众评议，并将基于审议时可用的信息做出决定。

3. 实现最新的委员会成员目标，追求效率最大化。

我们最初预计委员会的成员人数将达到 40 人，但三年来的实际运作表明，想要及时、定期、有效地进行审议和决策，最佳成员数量应该是 26 人。在我们 2023 年 4 月更新了现有的大部分委员会成员后，一旦正在进行的 2 位新委员会成员的遴选程序完成，Meta 就会退出遴选程序。在此之后，委员会成员的遴选将由委员会成员和受托人负责，Meta 不会参与其中。任何委员会成员的撤换始终由委员会成员和受托人独立决定，今后仍将如此。我们预计，到 2023 年末，委员会的成员构成将会完善。

4. 围绕七项战略重点，深入开展互动活动。

2022 年，我们基于对提请委员会复审的案件进行的分析，以及全球用户面临的问题，确定了七大战略重点。2023 年，在 Meta 向我们提请复审以及我们最终选定案件的过程中，这些战略重点都将提供指导。我们将在所有重点领域继续与利益相关者合作，了解 Meta 最迫切需要改进的政策和执行方式，以及哪些类型的案件有助于解决这些问题。我们鼓励在这些领域具备专业能力的利益相关者，通过公众评议程序、圆桌会议或个人对话与我们取得联系。

5. 追求长期计划，扩大业务范围。

2022 年，我们获得了制定约束性决定的能力，可以要求 Meta 在保留或恢复合格内容时增加警告画面。2023 年，我们将继续与 Meta 对话，以期将小组和帐户也纳入我们的业务范围。虽然我们预计这项工作将在 2023 年开始，但 Meta 向我们表示，由于技术和运营的原因，小组和帐户在 2024 年前不太可能纳入我们的业务范围。作为委员会，我们也有意向其他领域扩大业务，包括内容升级和降级。

6. 监督 META 对建议的落实情况，敦促 META 提供证据说明落实情况及其影响。

我们将继续密切监督 Meta 对建议的落实情况，并在季度透明度报告中介绍最新进展。我们还将敦促 Meta 提供证据，说明它已在各项政策和产品中落实我们的建议，并分享各种指标来说明这些建议如何影响使用其平台及受其平台影响的人的体验。

分享独立监督的好处

2022 年，越来越多的公司寻求外部专家协助其做出内容审核决定。Spotify 组建了安全咨询委员会，微博网站 Koo 也设立了咨询委员会。Twitter 前信任和安全负责人也呼吁为谷歌和苹果的应用商店建立内容审核委员会。

从根本上讲，独立监督标志着公司开放内部程序，邀请外部人士对决定进行复审。这样的挑战和审查可以催生更优质、更扎实的决策，有助于赢得用户的长期信任：这对社群和公司而言是一种“双赢”。

从一开始，监督委员会就不仅仅是为了改善 Facebook 和 Instagram 的内容生态，我们还希望借此机会尝试新的独立内容审核方法。若能取得成功，也可以为其他公司所用。作为委员会，我们在过去三年积累了丰富的经验，可以帮助其他公司制定更优质的决定，为用户提供更好的服务。

监督委员会的独家建设经验

之前从未有人尝试创建和运营一个内容管理监督委员会，事实上，这件事情远比表面看起来要复杂得多。从搭建技术基础设施到树立职场文化，我们从第一天起就在努力克服各种艰巨的挑战，以便在快速审核和审慎判断这两个方面达成平衡。几个月来，我们与 Meta 合作建立了一套独立申诉系统，并面向全世界的数十亿用户开放。我们制定了公众评议程序，便于人们对我们的决策流程发表意见。我们以不同的国籍、背景和观点，学会了如何审议棘手案件。

通过三年的运营，我们学习并应用了许多关于内容管理监督委员会的运作经验。我们专门明确了 5 个特征，可以为其他希望建立这种监督机构的科技公司提供帮助。

- ◎ **独立** - 任何监督机构都必须通过结构设计来实现独立判断。为了成功地审查内容决定，它绝不能因为商业、声誉或政治因素而受到平台的限制。作为委员会，我们会毫不犹豫地推翻 Meta 的决定。这种独立性是在世界各地的用户和公民社会团体中树立正当性的关键。
- ◎ **透明** - 要让人们信任一家公司的决定，就必须让他们了解这些决定是如何做出的。我们在准备决定时，会尽可能多地包含之前与政策及其执行情况相关的非公开信息。如果企业想要证明自己心怀责任和真诚维护言论自由的决心，就必须将删帖或封号的原因告知用户。此外还必须采取更多措施来解释这些决定背后的原因，并在政府或国家要求其删除内容时增加透明度。
- ◎ **多样** - 社交媒体平台的多数用户都来自美国和欧洲之外的其他地区。在美国和欧洲之外的地区，人们对社交媒体的负面和正面影响最为担忧。想要让他们信任一家公司的决策，不仅要与他们所在的世界“互动”，还要让他们感觉决策桌上有人倾听他们的声音，代表他们的利益。我们的监督委员会成员生活在 27 个国家/地区，使用 29 种语言。这样的多样性丰富了我们的最终决定。

“如何保障数十亿人的言论自由？此事非同小可，绝不能让一家公司独揽大权”

监督委员会成员
Alan Rusbridger



- ◎ **人权** – 当今的科技公司在审核内容时面临艰巨挑战：面对不同国籍、不同语言、不同文化的数十亿人，应该采用哪些规定？我们认为，国际人权标准给出了关键答案。这些标准对所有人都平等适用，同时还提供了统一的框架，将用户的自由表达权与其他人权（如生命权或隐私权）一同纳入考虑。
- ◎ **合作** – 最后，合作必不可少。只有在企业允许独立监督机构访问其数据和流程时，这些机构才能产生持久的影响。虽然在某些情况下，分享信息可能会损害用户的隐私，或为不良分子规避政策创造机会，但我们的工作所依赖的数据和信息只有 Meta 才能提供。这种开放性也有助于在用户和公民社会中建立正当性，并赢得他们的信任。企业也需要展示出落实建议的意愿，就像 Meta 对我们的许多提议表现出的态度。公司和监督委员会之间的合作关系要求双方都具备学习意愿。

作为委员会，我们希望上述方法能为其他希望通过独立监督获益的科技公司提供一个可信的框架。我们已经通过与 Meta 的合作克服了诸多运营障碍，并吸取了许多与建立这样一个机构有关的经验教训。虽然任何方法都需要根据公司的具体情况进行调整，但我们的经验可以帮助公司以更快的速度和更低的成本从独立监督中获益。

当我们寻找让科技行业负起责任的最佳途径时，我们鼓励所有公司都应考虑在其平台和服务中建立独立监督机制。

“我们发现，那些与我们面临相同问题的社交媒体平台正在关注我们的决策过程”

监督委员会受托人
Kristina Arriaga



帮助企业适应新的监管环境

如果一家公司与我们有着相同的信念，认为在独立机构的监督下打造透明而负责的内容管理机制，对于创造尊重言论自由和其他人权的网络环境至关重要，我们便有意与之合作。对于企业而言，如果愿意像 Meta 一样，对这种标准和结构做出有意义的承诺，使其用户和整个社会都能获益，便可赢得用户的信任，并向监管者展示其严肃的态度。

在未来的一年里，新的法规将为科技公司带来新的要求，迫使许多公司改变做法。以往，围绕社交媒体监管进行的辩论，往往将政府监管和行业自律视作“非此即彼”的两个选项，但我们现在看到，“共同监管”已经渐渐确立趋势。换言之，独立监管机制的构成要素将得到法律的支持。虽然许多参与者都将在这种新的监管环境中发挥作用，但我们相信，我们的独立监督模式，以及我们对透明度和用户公平对待的关注，将会成为这一解决方案中的重要组成部分。

“我们并不奢望成为整个行业的委员会。但我们希望分享自己的经验教训，而对于有意建立不同机构来制定标准和监督内容管理的公司，我们也有意与之合作”

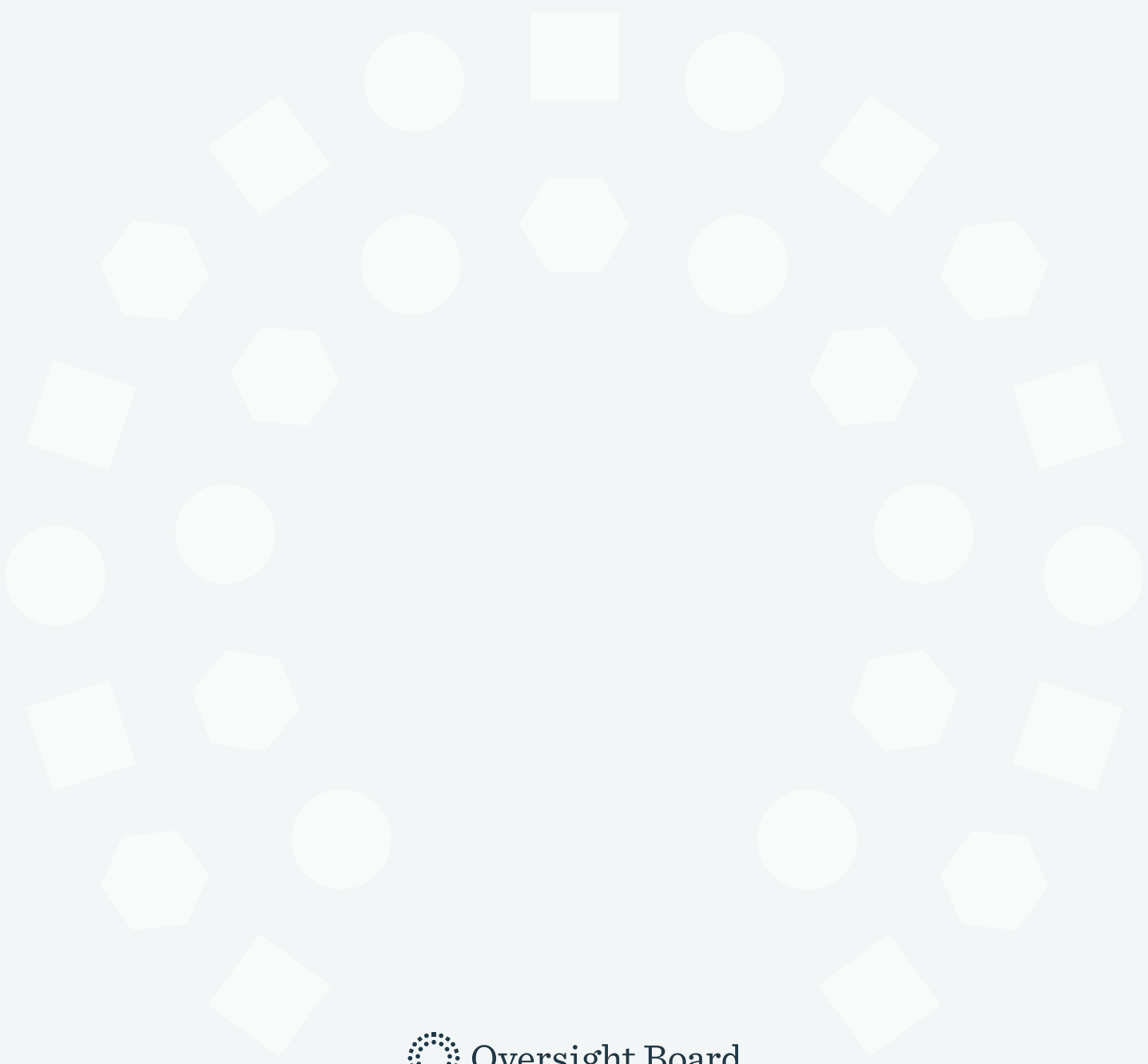
监督委员会主任
Thomas Hughes



总结

在最理想的情况下，社交媒体可以成为全球联系和对话的绝佳催化剂。但要仅利用社交媒体的优势而杜绝其带来的伤害，却并非易事。作为一个开创性的实体，监督委员会将继续运用我们吸取的经验教训来改进自己。我们相信，只要社交媒体公司能将国际人权标准作为决策基础，就能以更公平、更有原则的方式做出内容审核决定。对于与我们秉持相同目标，希望提高透明度，并改善人们上网体验的公司和组织，我们随时准备与之分享我们迄今为止吸取的经验教训。





Oversight Board

www.oversightboard.com

© 2023 Oversight Board LLC