



2021

监督委员会
2021 年年度
报告

目录

前言	2
联合主席序言	4
执行摘要	6
在全球范围内将国际人权标准应用到内容审核中：第 19 条	9
关键事件时间轴	12

案件甄选

概述	14
提交给委员会的案件	17
案件甄选委员会考虑的案件	21

案件审理决定

概述	22
表格：2021 年公布的决定	24
决定亮点：首批决定	26
决定亮点：关于封禁前总统特朗普帐户的审理决定	31
决定亮点：确保尊重言论自由	35
决定亮点：保护用户免受有害内容的伤害	40
委员会决策流程中的国际人权标准	43
委员会向 Meta 提出的帮助做出决定的问题	48

公众评议

51

建议

概述	54
Meta 对委员会建议的回应情况	57
从承诺到行动：为用户造福	60
责成 Meta 回应针对交叉检查提出的问题	62
交流和宣传	64

展望未来

2022 年及以后	66
-----------	----

附件

Meta 对我们的建议的回应和落实情况	68
---------------------	----



前言

监督委员会的使命是以保护言论自由和相关人权标准为指引，充分落实 Facebook 和 Instagram 的内容标准，促使 Meta 为世界各地的人们和社群带来更好的使用体验。为了履行使命，我们会对 Meta 的内容审核进行独立核查，针对最具棘手的内容争议做出约束性裁定。我们提出政策建议，促使 Meta 完善规定，使之在实际行动中坚持原则、保持透明度并对所有用户一视同仁。

截至 2018 年，Facebook 和 Instagram 的用户数量已经突破 20 亿。将人们联结在一起创造了巨大效益，它拉近许多人与其心爱之人的距离，巩固了现有社群的联系，甚至催生了全新的社群。但在获得这些效益的同时，也会社会带来诸多巨大的挑战。有些内容对人们的安全和自由构成威胁。有些人四处散布对民主和社会构成威胁的错误信息。仇恨言论和宣扬 Meta 所说的“危险人物和组织”的内容可能导致社群分裂，对社交媒体建立联结纽带的价值造成破坏。

为了讨论可以通过哪些措施来创建一个在全球范围内进行内容审核的治理系统，Meta 在 2018 年 11 月起草了一个构想，让人们可以通过一种新的方式向一个独立机构申诉内容决定，再由该独立机构做出具有约束力的裁定。随着人们越来越坚定地认为，Meta 不应该独自做出大量关于言论自由和安全问题的决定，建立一个在法律上和财务上区别于 Meta 的独立实体就成了自然而然的结果。

监督委员会就是基于这一初心建立起来的，旨在提高透明度，为 Meta 的内容裁定方式赋予更大的责任感、一致性和公平性，以此促进用户的权益。

2019 年 9 月发布的《章程》确立了委员会的机构独立性、成立原则和宗旨。《章程》授权委员会对特定内容是否应该出现在 Facebook 和 Instagram 上做出有约束力的决定，Meta 必须在决定公布后 7 日内予以落实。《章程》还授权委员会提出非约束性的建议，帮助 Meta 为世界各地的人们和社

群提供更好的使用体验。除了《章程》外，委员会还发布了一组《细则》，对自身的运营规程进行了细化。

监督委员会由三个相互关联的部分组成：委员会成员、受托人和行政管理机构。受托人负责保障委员会做出独立判断，确保其有效运作。包括 4 名联合主席在内的委员会成员负责甄选和裁定案件，向 Meta 提供建议，并领导我们落实工作。行政管理机构包含一个由全职员工组成的团队，负责协助委员会成员处理工作。

委员会成员是拥有多样化背景的领导者，他们在应对棘手问题方面经验丰富，包括与人权有关的问题。他们包括学者、民间团体领导者、前法官以及联合国和区域人权机构的负责人。我们的委员会成员生活在 27 个国家/地区，至少使用 29 种语言，男女人数相等。受托人同样男女人数相同。行政管理机构虽然人数较少，但他们之间也使用 28 种语言，强化了委员会的全球化属性。我们的员工约有 60% 为女性，40% 为男性。

为了践行我们在透明度和责任感方面的承诺，我们已经发布了几份季度透明度报告。现在，随着这份年度报告的发布，我们将公开披露以下信息：提交给委员会的案件综合摘要、委员会的决定和建议的摘要、公众评议概览。最后，本报告首次分析了委员会的决定是如何考虑或跟踪几个代表性案件中所涉及的国际人权标准的。

联合主席序言



Catalina Botero-Marino, Jamal Greene, Michael McConnell, Helle Thorning-Schmidt
监督委员会联合主席

我们成立监督委员会的初衷是因为我们认为社交媒体公司不应该独自做出关于内容审核的关键决定，然而这个想法说起来非常简单，实施起来却非常复杂。从成立的第一年，我们就开始落实这个想法。

我们制定了一套独立申诉系统，向全世界的数十亿用户开放。我们制定了公众评议程序，让人们可以对我们的决定发表意见。我们以不同的国籍、背景和观点，学会了如何审议棘手案件。在每个案件中，我们都会根据相关的 Facebook《社群守则》或 Instagram《社群准则》，并结合 Meta 自身的价值观，认真检查有争议的内容。在此过程中，我们考虑了这些原则是否得到了正确而一致的落实，用户是否得到了充分的通知和处理，这些守则和准则是否符合 Meta 承诺遵守的国际人权标准。

委员会的组织原则是：言论自由是民主社会的重要组成部分，必须得到充分尊重，以保护人权。我们对言论自由的承诺与《公民权利和政治权利国际公约》相一致，该公约规定“人人享有言论自由的权利。”我们所有

的决定都表明，国际人权标准是委员会权威的重要来源。

我们第一份年度报告的时间跨度从 2020 年 10 月我们开始接受申诉开始，一直到 2021 年 12 月为止，它阐述了我们在改进 Meta 为世界各地的用户和其他受影响的人提供更好的使用体验方面取得的进展。在全球疫情的背景下，建立一家有能力完成这些任务的机构自然要面临更多挑战。

Facebook 和 Instagram 用户显然存在大量被压抑的需求，他们希望通过某种渠道，向一个独立于 Meta 的组织申诉 Meta 做出的内容审核决定。在此期间，用户向委员会提交了 **100 多万次申诉**，其中绝大多数都希望恢复因为涉嫌违反 Meta 关于欺凌、仇恨言论或暴力与煽动暴力的规定，而从 Facebook 或 Instagram 上删除的内容。从 2021 年 4 月开始，用户还可以针对他们希望从 Facebook 或 Instagram 上删除的内容向监督委员会提出申诉。

我们在 2021 年公布了 20 起重大案件的完整解释，涉及从仇恨言论到新冠疫情错误信息

等诸多问题。在此过程中，我们以人权为基础分析内容审核决定，并收到近 1 万条公众评议，帮助我们做出了首批裁定。在更多的案件中，Meta 会因为委员会的工作而自愿修改错误的内容审核决定。

我们还向 Meta 提出了 86 项建议，推动该公司提高政策透明度。Meta 对我们的案件审理决定和政策建议做出的回应已经开始改善用户体验。Meta 现在为违反仇恨言论规定的 Facebook 英语用户提供更多详情，帮助其了解问题究竟出在哪里。该公司还在某些地方提供了新的提示信息，告知用户自动审查或人工审查是否会导致其内容被删除，并且承诺在透明度报告中增加关于政府要求和允许保留具新闻价值内容的信息。Meta 还承诺将《社群守则》翻译成印度使用的多种语言，一旦完成，就又有 4 亿多人可以用自己的母语阅读 Facebook 的规定。

Meta 承诺落实我们在 2021 年提出的大部分建议，那么我们的下一个任务是确保该公司将承诺转化为行动，切实改善人们使用 Facebook 和 Instagram 的体验。因此，本报告采用了新的数据驱动型方法来跟踪该公

司如何落实我们提出的每一条建议。我们还在向 Meta 索要新的数据，以便了解我们的建议对用户产生的具体影响。

在我们结束第一年的裁定工作时，委员会已经开始为一个能够成功推动 Meta 内部变革的项目打基础，以便帮助该公司更好地服务于受到其社交媒体平台非凡触角影响的人群和社群。2022 年，我们将以这个强劲的开端为基础，增加新的委员会成员，将工作延伸到新的领域，继续推动 Meta 向正确的方向发展。监督委员会在促使 Meta 肩负责任方面取得了初步成功，我们相信这证明了委员会的可行性，并为未来扩大和改善我们的业务提供了自我监管框架。通过这种方式，委员会将与企业、政府、学术界和民间团体携手塑造更光明、更安全的数字化未来，造福全世界。

执行摘要

用户和 Meta 提交
给委员会的案件
超过 110
万。

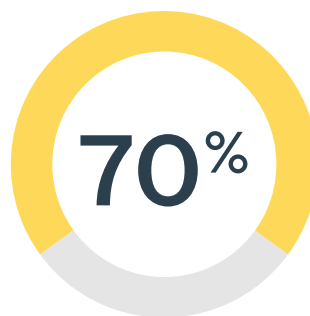
超过一半的决定与全球南方的国家/地区有关。



超过 8/10 的用户内容恢复申诉与 Meta 的欺凌、仇恨言论或暴力与煽动暴力规定有关。

公布了 20 项
决定，涉及从新冠疫情
错误信息到仇恨言
论等诸多方面。

使用国际人权
标准分析 Meta
的内容审核决定。



的案件审
理决定撤
销了 Meta
的决定 —
撤销其内容审核决
定 14 次，维持其
决定 6 次。

来自世界各地的个人和组织发表的

近 1 万条公众评议影响了我们的决定。



向 Meta 提出了 86 条建议。

Meta 承诺遵守的一些委员会建议：



在删除仇恨言论帖子时，为用户提供**更具体的信息**；



披露更多有关**政府要求**的报告；



将其规则翻译成**4 亿多人使用的多种语言**；



采用新的**危机政策方案**，规范危机状况应对方式；



在某些地方提供新的提示信息，告知用户**自动审查或人工审查**是否会导致其内容被删除。



本报告采用了一套**全新的数据驱动型方法**来跟踪落实情况，确保 Meta 落实对建议所做的承诺。



在我们提出的 86 项建议中，Meta 已证明落实或已汇报进度的建议达到**2/3**，与提高**透明度**有关的建议多数都属于这种情况。

展望 2022 年及未来，我们：

就**扩大我们的业务**

范围与 Meta 展开对话，包括审核用户对群组和帐户等方面的决定提出的申诉。

在亚洲、拉美、中东和非洲**扩大利益相关者的覆盖面**。



委员会如何促使 Meta 肩负责任

与 Meta 互动的第一年展现出的势头令委员会颇受鼓舞，但该公司必须尽快提高透明度。我们很高兴地宣布，Meta 已经完全履行了对案件审理决定做出的承诺，还同意落实委员会提出的一半以上的政策建议，而且越来越积极地回答委员会提出的问题。

然而，委员会仍有一些重大关切，有些涉及 Meta 的透明度，有些涉及提供某些案件的相关信息 and 政策建议。

从 2020 年 10 月到 2021 年 12 月，Meta……

- ✔ 100% 落实了我们的案件审理决定，并承诺落实我们提出的多数建议。
- ✔ 到 2021 年底，更大比例地回答了我们在案件审理决定中提出的问题，2021 年第四季度达到 94%。
- ✔ 对我们的建议进行了更好的回应，2021 年，Meta 的回应中被认为是“全面”或“比较全面”的比例每个季度都在增加。
- ❑ 未向委员会提供关于其交叉检查系统的全面信息。
- ❑ 遗漏了一份重要的政策例外，在三年间始终没有实际应用，如我们在奥贾兰单独监禁决定中所述。
- ❑ 承认在委员会甄选的 130 起案件入围名单中，有 51 起案件的原始决定是错误的。



在全球范围内将国际人权标准应用到内容审核中：第 19 条

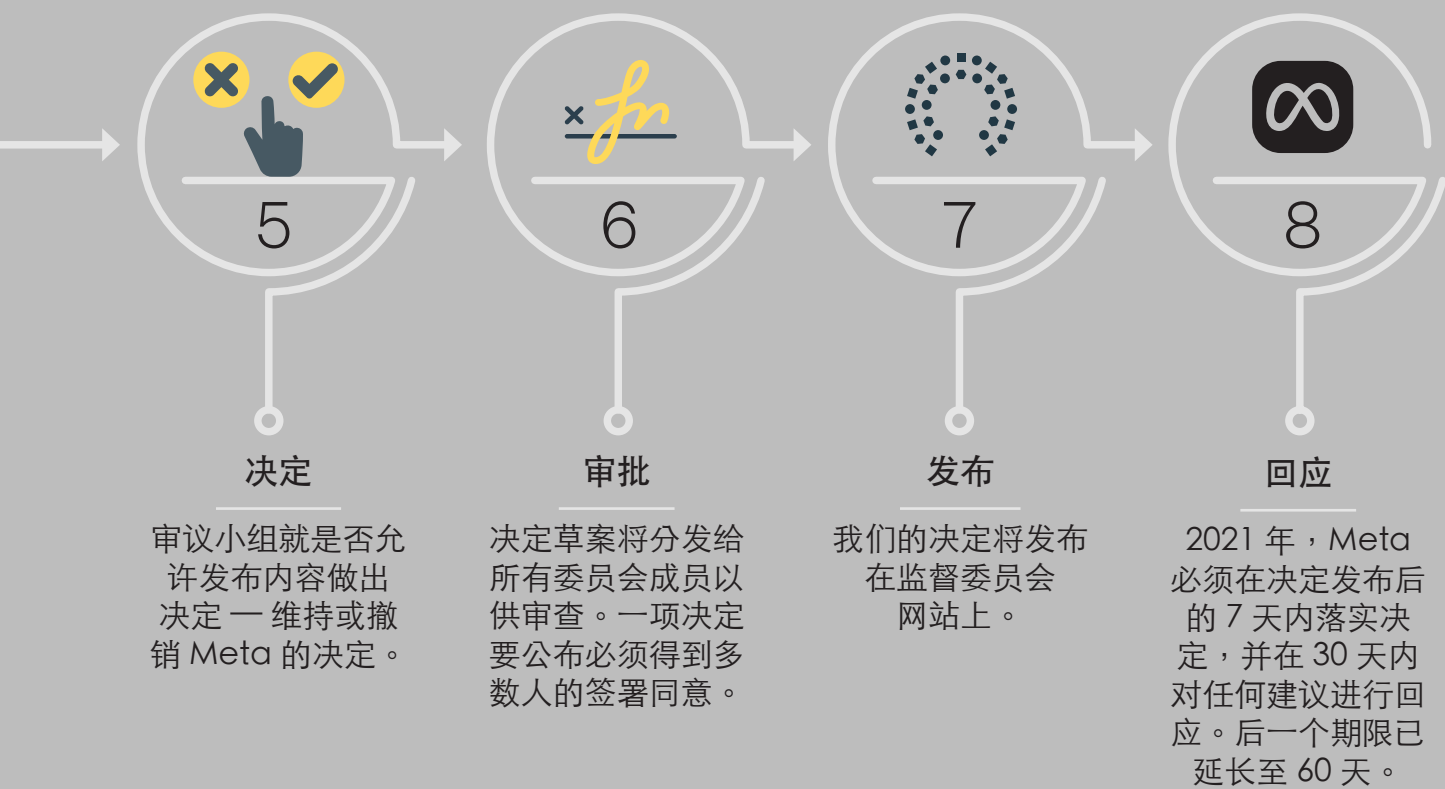
委员会 2021 年的工作最重要的主题是，确信 Meta 只要能遵守承诺，以国际人权标准作为基础，就能以更公平、更有原则的方式做出内容审核决定。为此，《监督委员会章程》规定，我们将“根据保护言论自由的人权准则，特别关注删除内容产生的影响。”《公民权利和政治权利国际公约》第 19 条就是这些准则中的一项，它规定，虽然“人人享有言论自由的权利……但行使该权利可能……受到一些限制，但仅限于……法律规定且必要的情况下。”该条款提供了评估言论限制的三步测试法：

- 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？委员会研究了《公民权利和政治权利国际公约》第 19 条中认可的限制言论的目的，借此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 该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？

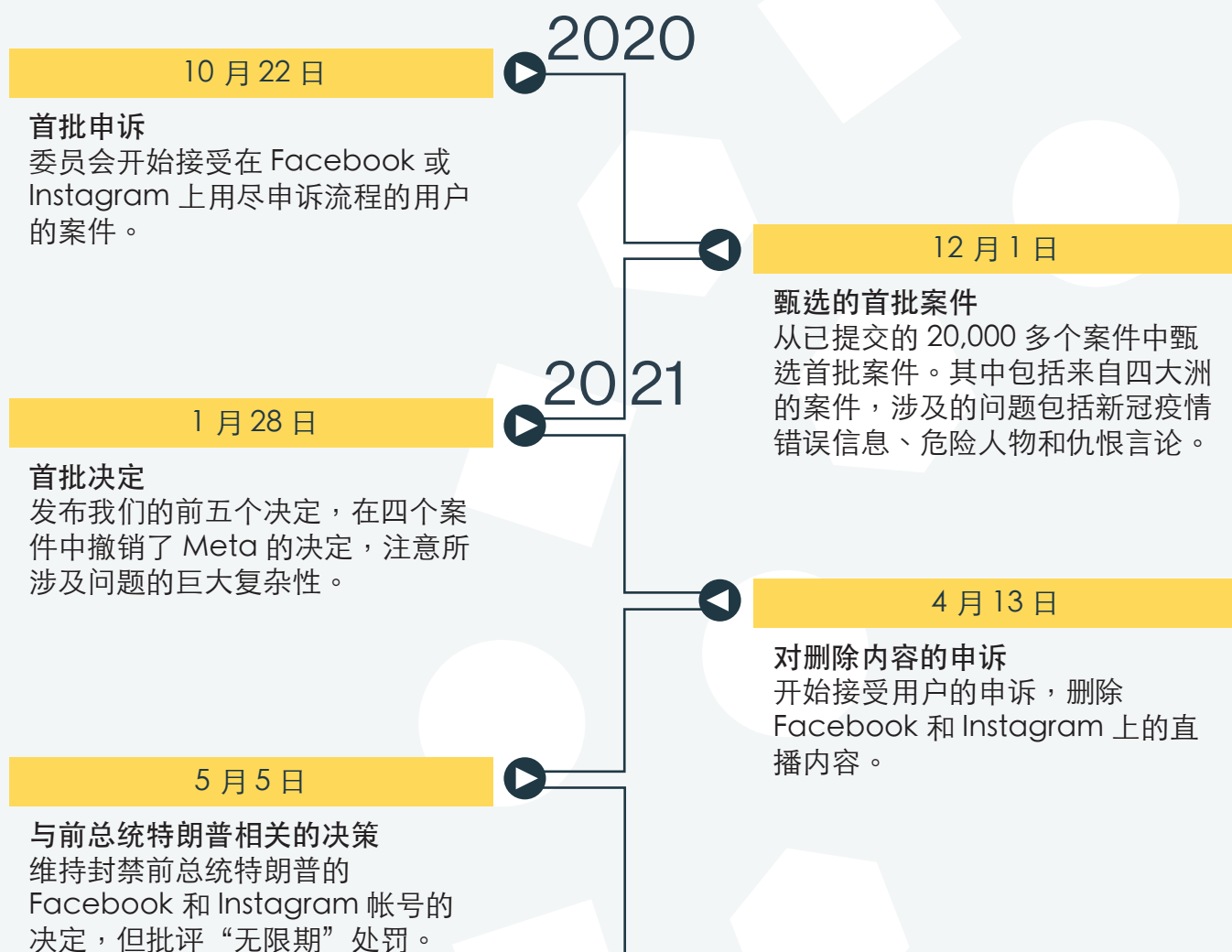
委员会如何 处理用户申诉



1 图表：监督委员会。资料来源：如监督委员会案件管理工具所示。



关键事件 时间轴





案件甄选



2020 年 10 月至 2021 年 12 月提交给委员会的案件达到 **1,152,181** 起，其中 **47** 起来自 Meta。

委员会平均每天收到 **2,649** 起案件。



超过 **8/10** 的用户内容恢复申诉与 Meta 的欺凌、仇恨言论或暴力与煽动暴力规定有关。



超过 **2/3** 的案件来自美国和加拿大以及欧洲。

有 **51** 起

共 **130** 起

在委员会的 130 起入围案件中，有 51 起被 Meta 确认为原始内容决定有误。

概述

Facebook 和 Instagram 用户只要向委员会申诉符合条件的内容，即可质询 Meta 的决定。从 2020 年 10 月委员会开始接受用户申诉，到 2021 年 12 月本报告的报告期结束，我们收到超过 110 万份用户要求独立复审 Meta 内容审核决定的申请。Meta 也向委员会提请复审 47 起案件。从 2020 年 10 月到 2021 年 12 月，委员会平均每天收到 2,649 起案件。

提交的案件数量说明了委员会的工作对用户的重要性。在 2020 和 2021 年，我们专门优先处理了三类案件，分别是：有可能对世界各地的大量用户产生影响的案件；对公共话语至关重要的案件；引发有关 Meta 政策的重要问题的案件。为了解决特定国家/地区的人特有的问题，我们还甄选了来自世界各地的案件。我们甄选的几起案件还对在全球范围内将国际人权标准应用于内容审核产生了重大影响。

委员会还倾听用户的意见，当某起案件集中体现了申诉中反复出现的问题时，便会获得优先处理。也就是说，尽管我们可能只甄选一起提出特定问题的案件进行复审，但委员会却可以解决更多的人面对的共性问题，包括那些案件没有被选中的人。例如，我们发现 Meta 删除了很多提到纳粹人物的帖子，但有些内容其实并没有支持或赞扬任何“危险人物或组织”

，于是我们甄选了纳粹引语案件来审查一个影响大量用户的重要问题。

为了确保我们在案件甄选过程中听取丰富的意见，委员会定期任命一个新的五人案件甄选委员会，负责处理用户广泛关切的各种问题。迄今为止，根据我们的总体甄选标准，委员会优先处理的案件包括：Meta 的自动化系统有可能对内容进行审核的案件，特别是在用户报告缺乏以原始语言进行人工审查的国家/地区；涉嫌赞扬或宣传“危险人物和组织”的案件和内容；以及提出重要的新闻自由问题的案件，特别是在冲突地区面临“紧迫伤害”危险的记者。



经验教训

在案件甄选过程，委员会遭遇了许多运营方面的挑战和障碍：

适应用户在案件甄选后删除自己内容的自由言论权

举例：就在我们 2020 年 12 月公布首批案件后不久，一位用户决定删除一条与马来西亚前总理马哈蒂尔·默罕默德 (Mahathir Mohamad) 发布的推文有关的内容。这种行为导致委员会无法对该案件进行复审。2021 年末也发生过类似的情况，一位用户删除了涉及美军撤出阿富汗期间和之后在当地工作的一名记者的内容。在这两起案件中，虽然删除内容造成委员会无法涉及普遍关注的领域，但我们尊重所有用户出于任何原因删除自己内容的权力。由于我们预料到这种情况会不时

发生，所以每当案件因为用户行为被撤回时，我们都会及时公布并澄清状况。

针对 Meta 不透明的“攻击系统”删除的内容引入更高的透明度

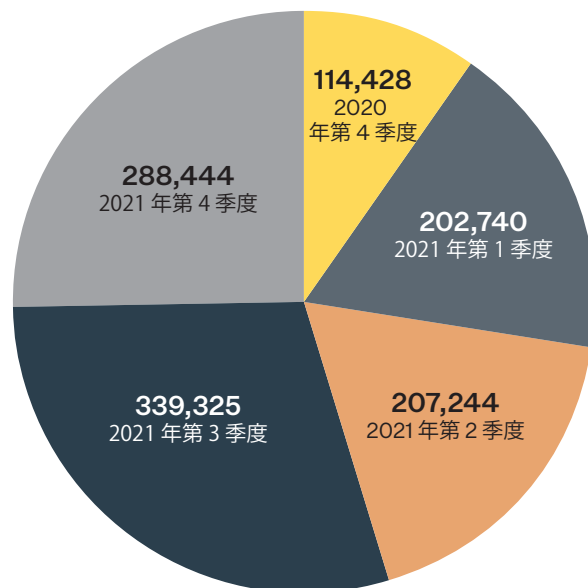
2021 年第二季度，有两起案件虽然分配给了审议小组，但由于 Meta 采取了执行措施，所以并未公布。一起关于新冠病毒起源的案件之所以没有公布，是因为 Meta 在收到太多针对该内容发出的“攻击”后删除了它所在的页面。在另一起关于性别错误的案件中，Meta 也因为收到太多“攻击”而删除了该帖子所在的页面。

提交给委员会的案件²

自从我们 2020 年 10 月开始接受案件以来，提交给委员会的案件数量逐季增长，并一直持续到 2021 年第三季度，2021 年第四季度略有降低。这表明全球各地的 Facebook 和 Instagram 用户都普遍面临内容审核问题。

在 2021 年第二季度和第三季度之间，申诉数量增长了 64%。这种增长可能源自两个原因，一方面是因为 Meta 改进了用户通过 Facebook 移动应用向委员会提交申诉的方式，另一方面则是因为委员会获得了较高的知名度。在始于 2021 年 8 月 30 日星期一的那个星期内，用户提交的案件数创下单周最高记录，总共向委员会提交了超过 3.2 万份不同的申诉。

每个季度提交给委员会的案件估算数据



图表：监督委员会。资料来源：如监督委员会案件管理工具所示。

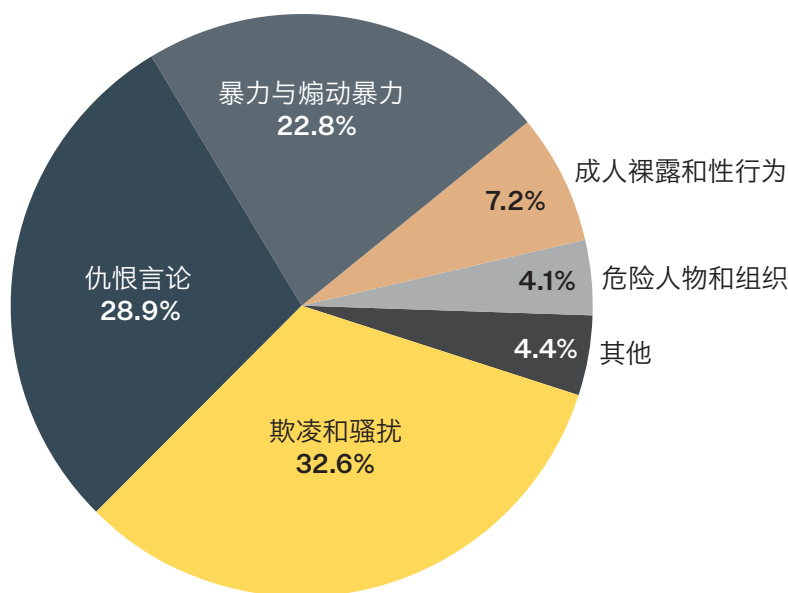


2 由于案件管理工具（通过该工具向委员会提交案件）以及我们目前可获得数据的局限性，提交的案件有时会由委员会的案件甄选团队手工统计，就像它们在当时出现在 CMT 中。这些数字的细分也可能是基于样本。因此，这些数字应被视为估计值。

3 该数据仅涵盖 2021 年第二季度至第四季度，因为只有这一时期的数据可用。

依据社群守则提交给委员会的案件估算数据 (2020 年 10 月-2021 年 12 月)

(仅包括用户要求恢复 Facebook 和 Instagram 内容的请求)



图表：监督委员会。资料来源：如监督委员会案件管理工具所示。

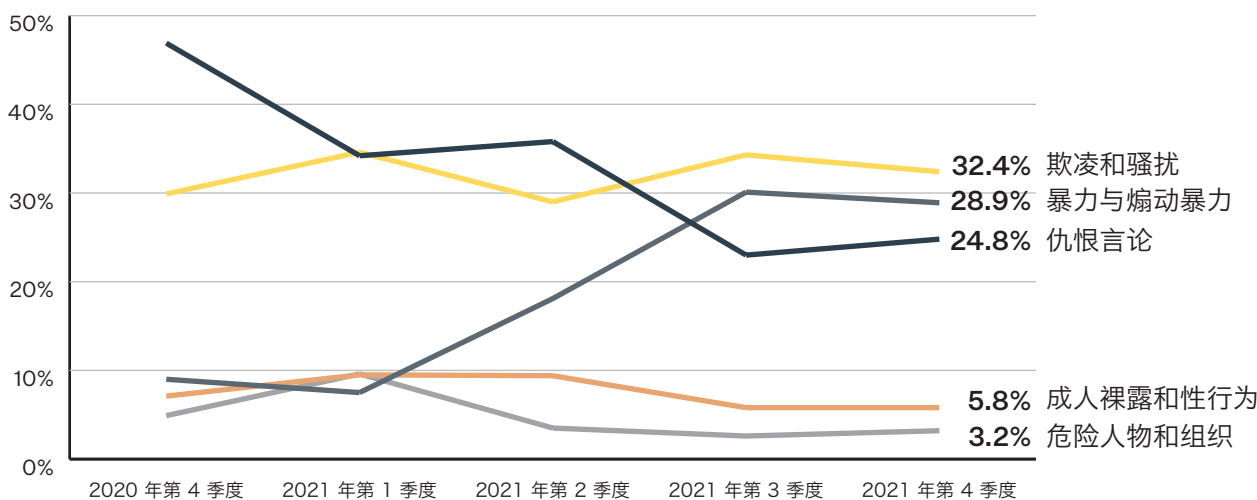
在提交的这些案件中，99% 都与 Facebook 上的内容有关，³ 只有 1% 与 Instagram 上的内容有关。从 2021 年第二季度开始，当用户可以通过向委员会提交申诉来删除其他人的内容后，94% 的用户申诉是为了恢复 Facebook 或 Instagram 上的内容，还有 6% 的申诉是为了删除这些平台上的内容。

绝大多数为恢复内容而提交的用户申诉 (84%) 都涉及 Facebook 的三项社群守则：欺凌和骚扰、仇恨言论、暴力与煽动暴力。其余的内

容恢复申诉绝大多数都是与成人裸露和性行为以及危险人物和组织有关的案件。

把这些数字放到更宏观的背景来看，从 2020 年第四季度到 2021 年第四季度，Meta “采取行动” 最多的内容（包括删除内容，但也包括发出警告等其他行动）涉及的社群守则是成人裸露和性行为，其次是暴力和图片内容以及仇恨言论。但 Meta 在此期间收到最多内容申诉的社群守则是仇恨言论，其次是欺凌和骚扰以及成人裸露和性行为。相比而言，委员会在此期间收到的内容恢复申诉涉及最多的两项

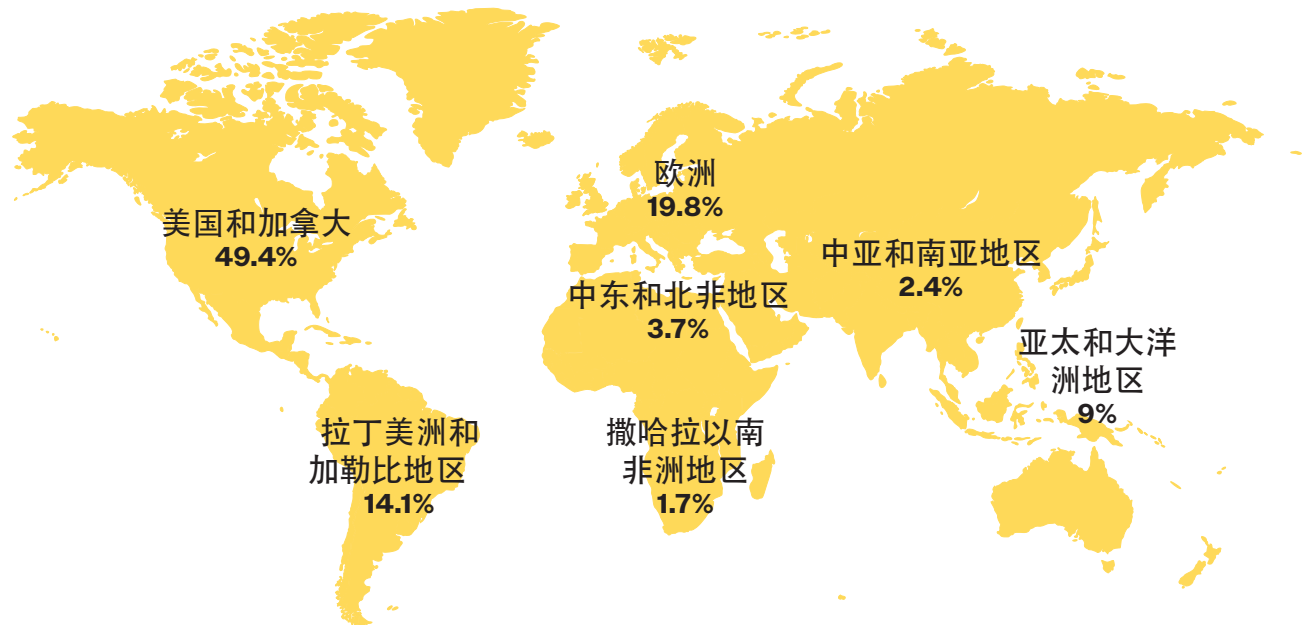
每个季度提交给委员会的违反社群守则的案件估算数据 (仅包括用户要求恢复 Facebook 和 Instagram 内容的请求)



图表：监督委员会。资料来源：如监督委员会案件管理工具所示。

4 资料来源：Meta 社群守则执行报告 (CSER)。此处提及的社群守则不包括委员会工作范围之外的政策领域，以及 CSER 中未提及的政策领域。

按用户选择的区域划分时，各地区提交给委员会的案件估算数据（2020年10月-2021年12月）



Map: Oversight Board. Source: As seen in the Oversight Board Case Management Tool.

社群守则中，第一是欺凌和骚扰，其次是仇恨言论。

按季度来看，我们收集的有关在这两大平台上恢复内容的用户申诉数据显示出一些有趣的趋势。虽然欺凌和骚扰社群守则始终占到申诉数的 1/3 左右，但其他社群守则在全年的变化较大。与 Facebook 的暴力与煽动暴力规定有关的案件比例同比增长 3 倍，从 2020 年第四季度的 9% 增加到 2021 年第四季度的 29%。在同一时期，与仇恨言论有关的案件比例减少近半，从 2020 年第四季度的 47% 减少到 2021 年第四季度的只有 25%。

超过 2/3 的用户申诉来自全球北方，美国和加拿大占到总数的 49%，欧洲占到 20%。14% 的申诉来自拉美和加勒比地区，9% 来自亚太和大洋洲，只有 4% 来自中东和北非，2% 来自中亚和南亚，2% 来自撒哈拉以南非洲地区。

我们意识到，这种分布与 Facebook 和 Instagram 用户在全球的分布状况并不一

致。例如，2019 年，Facebook 用户数最多的 20 个国家/地区中，只有 6 个位于欧洲和北美，印度的 Facebook 和 Instagram 用户数位居全球第一。来自欧洲以及美国和加拿大之外的用户申诉数量较少也说明，世界其他地方的许多 Facebook 和 Instagram 用户并没有意识到他们可以向委员会申诉 Meta 的内容审核决定。我们也不认为这张地图上的申诉数据分布情况反映了内容审核问题在全球的实际分布状况。事实上，我们有理由相信亚洲、非洲和中东用户在 Meta 平台上经历的问题较之于世界其他地区只多不少。我们目前的决定涵盖了来自印度和埃塞俄比亚的帖子，外界因此担心 Meta 是否投入了足够的资源审核非英语内容。

除了在欧洲以及美国和加拿大之外的其他地区扩大宣传，我们还在考虑和甄选全球北方之外的许多案件，力求解决全球用户关切的问题。案件甄选委员会考虑选择的 278 起案件中，超过一半来自美国和加拿大以及欧洲以外的地区，而我们在 2021 年公布的 20 项决定中，有一半以上涉及全球南方国家/地区。

Meta 提请复审的案件

除了 Facebook 和 Instagram 用户提交的申诉外，Meta 也可以向我们提请复审重要而棘手的案件。

截至 2021 年 12 月底，在 Meta 提请委员会复审的 47 起案件中，有 41 起涉及 Facebook 上分享的内容，还有 6 起涉及 Instagram 上的内容。在 Meta 提请复审的 47 起案件中，有 29 起涉及的内容仍保留在 Facebook 或 Instagram 上，未被视为违反该公司规定。

剩余 18 起案件涉及的内容已经因为涉嫌违反以下社群守则而被 Meta 从 Facebook 或 Instagram 上删除：暴力与煽动暴力（6 起案件）、仇恨言论（3 起案件）、危险人物和组织（3 起案件）、儿童性剥削、虐待和裸露（3 起案件）、欺凌和骚扰（2 起案件）、成人裸露和性行为（1 起案件）。

在 Meta 提请委员会复审的案件中，将近 2/3 与来自欧洲以及美国和加拿大的内容有关，即 47 起案件中有 30 起来自这些地区。Meta 提请复审的案件中有 17 起涉及世界其他地区：4 起来自拉美和加勒比地区，4 起来自中东和北非，3 起来自中亚和南亚，3 起来自亚太和大洋洲，还有 3 起来自撒哈拉以南非洲地区。

各地区 Meta 提请审核的案件数据



案件甄选委员会考虑的案件

为解决全球 Facebook 和 Instagram 用户面临的问题，委员会甄选了涉及多种不同语言的多样化案件。截至 2021 年 12 月底，案件甄选委员会考虑的 278 起案件覆盖了从斐济到乍得再到特立尼达和多巴哥在内的 70 多个国家/地区。这些案件包含的内容涉及 38 种不同语言。

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。我们将此称作“执行错误”。

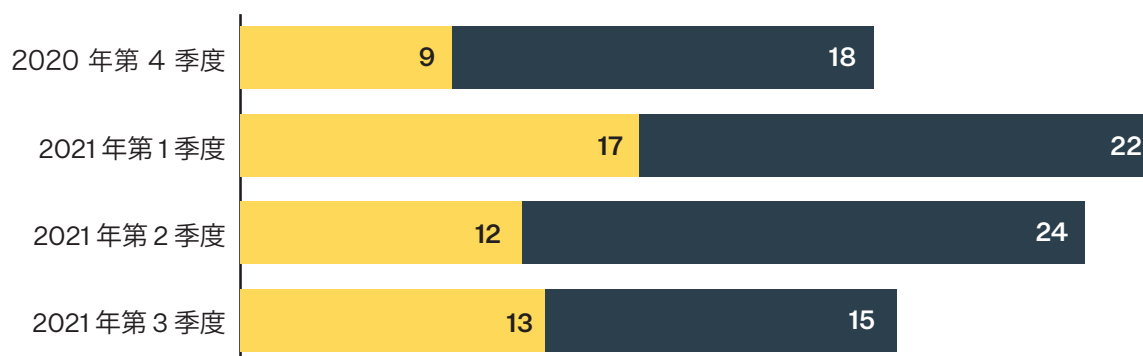
本页的图表按季度显示了被 Meta 认定为“执行错误”的案件。2020 年第四季度和

2021 年第二季度，Meta 发现有 1/3 的入围案件的原始决定有误。2021 年第一季度这一比率为 44%，2021 年第三季度为 46%。2021 年第四季度没有任何案件入围。

截至 2021 年 12 月底，共有 130 起案件入围委员会的复审名单，Meta 认定其中 51 起案件的原始内容决定有误。这是一份数量不大的样本，而且委员会有意识地选择了棘手案件，但接近 40% 的入围案件都被 Meta 认定为决定有误，这令人颇感担忧。如此高的错误率引发了更广泛的问题：Meta 的内容审核准确率究竟有多高？在案件被委员会受理前，Meta 的申诉流程落实情况如何？

入围案件中与“执行错误”相关的数据

■ 认定为“执行错误” ■ 未认定为“执行错误”



图表：监督委员会。资料来源：如监督委员会案件管理工具所示。

案件 审理决定



2021 年
公布了
20 项决定。



撤销 Meta 决定 **14** 次，维持 Meta 决定 **6** 次。



这些决定中有 **16** 起案件由用户申诉，有 **4** 起案件由 Meta 提请复审。



17 起案件来自 Facebook；
3 起案件来自 Instagram。

决定和落实平均耗时
83 天。



超过一半的决定
与全球南方国家/地区有
关。



在案件复审过程中，
向 Meta 提出
313 个问题。



使用国际人权标
准分析 Meta 的内容
审核决定。



概述

监督委员会存在的目的是在 Facebook 和 Instagram 最棘手的内容问题上独立做出约束性的决定。我们第一批 20 项完整决定以及相关的政策建议反复提到了关于该公司如何对待全球用户的关键问题。委员会不仅在必要时撤销了 Meta 的决定，还提出了能够改善用户体验和提高透明度的政策调整建议。总的来说，在公司政策不明确、不一致或不适用的情况下，这些措施会为用户提供支持。我们的决定涵盖的主要主题包括：

- **保护言论自由和其他人权。** Meta 充当了数十亿人的主要沟通网络，尽管它认为允许人们畅所欲言拥有“至高无上”的价值，但在必要时必须限制言论自由，以保护用户和其他人免受真正的伤害。协调这些关切是委员会最重要也是最困难的责任。
- **避免断章取义。** 我们通常会根据问题的背景做出决定。这包括特定词汇和帖子本身的意思，以及发布内容时所处的环境，如暴力冲突或政治抗议。
- **向用户提供明确的信息。** 我们对下面三起案件做出了审理决定：一起是 Meta 没有告知用户违反了哪一项社群守则，一起是 Meta 没有审核用户申诉，还有一起是 Meta 遗漏了一份政策准则三年之久。还有三个针对 Instagram 内容的审理决定也凸显出用户对 Facebook 与 Instagram 的规定之间的关系不够明确。
- **对不同地区的用户一视同仁。** 我们的决定凸显出一个问题：Meta 为 Facebook 和 Instagram 制定的规定没有翻译成用户使用的所有语言。我们还担心 Meta 是否投入了足够的资源审核非英语内容。
- **确保 Meta 的规定得到正确执行。** 我们的决定表明，Meta 需要改进自动和人工内容审核的准确性，并向委员会和公众更详细地解释错误发生的原因。

经验教训

修订时间表以反映现实运营状况

我们最初制定的计划是从 Meta 最近一次对某起案件做出决定后 90 天内完成复审，但当我们起草首批决定时，却很快意识到在此时间内无法完成任务。为了确保所有案件都获得相同的审议时间，我们在 3 月将 90 天内决定和落实一起案件的起点调整成从该案件被分配给审议小组开始计算。由于委员会制定了自己的裁决流程，并开始向 Meta 提出更多问题，我们在 11 月又将 90 天时限的起点调整成从公布案件开始计算。

优化流程和标准

我们在审议第一批案件时遇到了几个复杂问题。例如，当我们研究一条内容时，应当考虑案件复审过程中提供给我们所有信息，还是只需根据当时既定的审核环境，评估 Meta 依据当时所掌握的信息做出的决定是否属于合理回应？另外，委员会应当如何将主要针对国家/地区制定的国际人权标准，应用到社交媒体平台制定的决策中？本章的人权部分详细讨论了这个问题。

2021 年公布的决定

	平台	资料来源	社群守则	国家/地区	委员会决定
缅甸贬低穆斯林的帖子 案件编号 2020-002-FB-UA		用户		缅甸、法国、中国	撤销 Meta 的删除决定
亚美尼亚人侮辱阿塞拜疆 案件编号 2020-003-FB-UA		用户		亚美尼亚、阿塞拜疆	维持 Meta 的删除决定
乳腺癌症状和裸露 案件编号 2020-004-IG-UA		用户		巴西	撤销 Meta 的删除决定
纳粹引语 2020-005-FB-UA		用户		美国	撤销 Meta 的删除决定
宣称新冠肺炎疗法 案件编号 2020-006-FB-FBR		Meta		法国	撤销 Meta 的删除决定
印度抗议法国 案件编号 2020-007-FB-FBR		Meta		印度、法国	撤销 Meta 的删除决定
封禁前总统特朗普帐户 案件编号 2021-001-FB-FBR		Meta		美国	维持 Meta 的删除决定
扮成“黑彼得” 案件编号 2021-002-FB-UA		用户		荷兰	维持 Meta 的删除决定
旁遮普语用户提出对印度 RSS 的担忧 案件编号 2021-003-FB-UA		用户		印度	撤销 Meta 的删除决定
俄罗斯发生支持纳瓦利内抗议活动 案件编号 2021-004-FB-UA		用户		俄罗斯	撤销 Meta 的删除决定



危险人物和组织



暴力与煽动暴力



仇恨言论



成人裸露和性行为



欺凌和骚扰

	平台	资料来源	社群守则	国家/地区	委员会决定
“双按钮”表情包 案件编号 2021-005-FB-UA		用户		亚美尼亚、土耳其、美国	撤销 Meta 的删除决定
奥贾兰单独监禁 案件编号 2021-006-IG-UA		用户		土耳其、美国	撤销 Meta 的删除决定
缅甸机器人 案件编号 2021-007-FB-UA		用户		缅甸	撤销 Meta 的删除决定
巴西新冠疫情封城 案件编号 2021-008-FB-FBR		Meta	无	巴西	维持 Meta 的保留决定
分享半岛电视台的帖子 案件编号 2021-009-FB-UA		用户		以色列、埃及	撤销 Meta 的删除决定
哥伦比亚抗议 案件编号 2021-010-FB-UA		用户		哥伦比亚	撤销 Meta 的删除决定
南非诋毁性措辞 案件编号 2021-011-FB-UA		用户		南非	维持 Meta 的删除决定
串珠带 案件编号 2021-012-FB-UA		用户		加拿大	撤销 Meta 的删除决定
死藤水酿造 案件编号 2021-013-IG-UA		用户		巴西	撤销 Meta 的删除决定
在 Raya Kobo 涉嫌犯罪 案件编号 2021-014-FB-UA		用户		埃塞俄比亚	维持 Meta 的删除决定



危险人物和组织



仇恨言论



管制品

首批决定

“

“直到现在，内容决定一直都由马克·扎克伯格负责……这是历史性的一天。这是第一次由 Facebook 之外的机构负责内容审核，而且 Facebook 必须遵守我们的决定。”

监督委员会联合主席 Helle Thorning-Schmidt



2021 年 1 月 28 日，基于我们从 2 万项申诉中甄选出首批案件以来进行的评审，监督委员会公布了 5 项决定。委员会成员进行评审时遵循了我们在《细则》中规定的流程：每起案件都会分配给由 5 名委员会成员组成的审议小组，其中至少有一名成员来自案件内容涉及的地区，且男女成员兼有。一旦审议小组做出决定，就需要将草案发给所有委员会成员评议，之后再由委员会全体成员投票表决。案件涵盖四大洲的内容：亚洲、欧洲、北美洲和南美洲。

在首批 5 起案件中，有 4 起撤销了 Meta 的决定，1 起维持了 Meta 的删帖决定。这些案件涉及的问题很广泛。在其中一起案件中，委员会成员审查了在武装冲突背景下，Meta 是否应该因为其中包含充满仇恨的诋毁性言论而删除一个原本允许发表的帖子。在另一起案件中，我们还审查了一个被控散布新冠肺炎错误信息的帖子是否会引发紧迫伤害。另一项决定通过指出 Meta 自动执行系统无法区分乳腺癌宣传活动中允许暴露的女性乳头图像，树立了一个重要的先例。



乳腺癌症状和裸露

公布日期	国家/地区	社群守则
2021 年 1 月 28 日	巴西	成人裸露和性行为

2020 年 10 月，巴西的一位用户在 Instagram 上发了一张配有葡萄牙语标题的图片，希望提高人们对乳腺癌的意识。那篇帖子被自动化系统删除了，理由是它违反了 Facebook《社群守则》中关于成人裸露和性行为的规定 — Facebook 解释称，这条规定也适用于 Instagram。我们撤销了 Meta 的决定，因为《社群守则》明确排除了一种情况，即当用户试图“出于教育或医疗原因，提高人们的意识”时，允许发布裸露图片，而且明确允许发布未遮挡女性乳头的图片以“提高乳腺癌意识”。

当该内容被删除后，那位用户向 Meta 提出申诉。Meta 之前在公开声明中表示，由于新冠疫情导致其审核能力暂时降低，所以无法始终为用户提供申诉选项。此外，Meta 还宣称，无法对所有申诉都进行人工审查。

我们驳回了 Meta 关于该案件“有争议”的抗辩，而在恢复被删除的帖子后，该用户和 Meta 对保留那条内容再无异议。我们明确表示，即使 Meta 后来纠正错误并恢复内容，委员会仍然有权复审用户提交的案件。最后，委员会建议，当用户的内容被发现违反了 Facebook 关于成人裸露和性行为的社群守则时，可以对自动化系统做出的决定提出申诉，要求人工审查。





纳粹引语

公布日期	国家/地区	社群守则
2021 年 1 月 28 日	美国	危险人物和组织

2020 年 10 月，一位美国用户用英语发帖，引用了被误认为是纳粹德国宣传部长约瑟夫·戈培尔 (Joseph Goebbels) 的话。这句引语称，在宣传中依靠理智并无作用，因为理智不能让人回心转意，他们往往会受街头强人的煽动而失去理智。这句引语进一步指出应诉诸于情感和直觉的煽动。Meta 以违反 Facebook《社群守则》关于危险人物和组织的规定为由删除了那篇帖子。该用户则向委员会坚称，他们的意图是将引语所传达的观点与唐纳德·特朗普总统的任期表现进行比较。

委员会审核该案件时发现，由于该引文并没有支持纳粹的意识形态或其政权的仇恨及暴力行径，所以并不违反 Facebook 的社群守则。我们指出，在本案件中，Meta 的模式要求内容审核员在不充分考虑背景的情况下对内容进行审核，导致用户的表达受到不必要的限制。委员会进而建议 Meta 应当解释并举例说明如何应用危险人物和组织政策中的关键术语，包括它对“赞扬”、“支持”和“代表”的定义，这应当与 Facebook 的内部执行标准中的定义相一致。



宣称新冠肺炎疗法

公布日期	国家/地区	社群守则
2021 年 1 月 28 日	法国	暴力与煽动暴力

另外一项决定考虑了与健康有关的错误信息何时会达到引发迫切身体伤害风险的阈值。2020 年 10 月，一位用户用法语发表了一段声明，内容是关于法国国家药品安全管理局 (Agence Nationale de Sécurité du Médicament) 决定不批准使用羟氯喹联合阿奇霉素治疗新冠肺炎的决定。该用户称这种组合为“无害药物”，并宣称这是一种“疗法”。Meta 删除这篇帖子的依据是暴力与煽动暴力政策，其中包含有关错误信息和紧迫伤害的规定。委员会驳回了 Meta 的决定，因为 Meta 并未证明这条内容会构成紧迫伤害——他们自己制定的政策标准。事实上，这篇帖子是针对公共政策，而不是向个人推荐治疗方法，而且其中涉及的药物在法国需凭处方购买。这些都是我们考虑的因素。我们建议，在处理批评卫生部门的立场以及构成潜在但非紧迫伤害的帖子时，Meta 应采取干扰性更小的措施，包括提供公共卫生机构的网站链接。



缅甸贬低穆斯林的帖子

公布日期	国家/地区	社群守则
2021 年 1 月 28 日	缅甸、法国、中国	仇恨言论

这项决定与缅甸用户发布的帖子有关，该用户分享了两张被广泛传播的照片：一名叙利亚库尔德族儿童在 2015 年 9 月偷渡到欧洲时溺水身亡。这篇帖子还提出疑问，与法国有人因展示侮辱先知穆罕默德的漫画而惨遭杀害的情形相比，为何穆斯林世界对中国维吾尔族穆斯林所遭受的对待普遍没有反应。图片配文使用缅甸语，Meta 将其理解为“穆斯林的心理确实有问题。”而委员会的译员则建议将其解读为“这些穆斯林男性的思维方式有问题。”委员会在撤销 Meta 的删帖决定时指出，结合背景来看，这段文字最好理解为“就穆斯林对待法国和中国所发生事件的明显不一致反应而发表的评论”。

我们承认，缅甸网络世界的仇恨言论与针对罗兴亚穆斯林少数族裔的反人类及种族灭绝罪行有关，但委员会发现，这篇帖子“并没有提倡仇恨或有意煽动任何形式的紧迫伤害”。简而言之，Meta 以违反 Facebook 仇恨言论社群守则为由删除该帖的决定是错误的。





亚美尼亚人侮辱阿塞拜疆

公布日期	国家/地区	社群守则
2021 年 1 月 28 日	亚美尼亚、阿塞拜疆	仇恨言论

在我们的首批 5 项决定中，唯一维持 Meta 决定的一起案件与 2020 年 11 月发布的一组历史照片有关。据称，这些照片拍摄的是阿塞拜疆首都巴库的教堂。这篇帖子以俄语配文称，巴库是亚美尼亚人建立起来的，现在其遗产（包括这些教堂）都已被摧毁。该用户在帖子中加入了话题标签，呼吁终止阿塞拜疆人的侵略和破坏行为。该帖子使用“taziks”一词来形容阿塞拜疆人，并且称阿塞拜疆人“与亚美尼亚人相比，他们没有自己的历史”。

“taziks”在俄语中的意思是“洗漱盆”，看起来像是对俄语单词“aziks”一词玩的文字游戏，而后者是对阿塞拜疆人的一种贬低性称呼，属于 Meta 禁止的诋毁词之一。我们在这项决定中指出，“从帖子中这个词语的上下文语境也能清楚看出，其使用者有意将对方非人化。”因此，我们认为，在这一背景下的删帖决定符合 Meta 的人权责任。我们进一步发现，如果保留这些内容，“这类内容越积越多可能会营造不良的环境，推高歧视和暴力行为出现的概率。”我们建议 Meta 立即将违规原因通知用户（本案件的违规原因就是一个词）。这样一来，如果用户愿意，便可重新发布这条信息的剩余部分。

委员会在公布第一批决定时注意到：

我们相信监督委员会的第一批案件审理决定证明我们将履行诺言，维护全世界的用户和社群的利益，逐步改变 Facebook 的内容审核模式，让 Meta 切实肩负起自己的责任。这只是个开始，后续还会耗费许多时间，我们也期待着通过委员会今后的大量案件审理决定分享我们的进程。



决定 亮点

关于封禁前总统特朗普 普帐户的审理决定

公布日期
2021年5月5日

国家/地区
美国

社群守则
危险人物和组织

美国东部标准时间 2021 年 1 月 6 日下午 4:21，就在一群暴徒强行闯入位于华盛顿特区的美国国会大厦，并在走廊上横冲直撞几个小时后，美国总统唐纳德·J·特朗普 (Donald J. Trump) 在 Facebook 上发布了一分钟的视频（同时分享到他的 Instagram 帐户中），他在视频中说：

我知道你们很痛苦、很受伤。我们在这场选举中的胜利果实被人窃取了。每个人都清楚，特别是我们的对手，我们以压倒性的优势赢得了选举，但现在，请别再纠缠，离开这里。我们必须拥护和平，拥护法律，必须尊重遵纪守法的伟

大人民。我们不想看到任何人受伤。这是一段异常艰难的时期。

他接着说道：

历史上从未有过这样的事：他们就这样将胜利从我们的手中夺走——从你我和这个国家手中夺走。这场选举的舞弊现象非常严重，我们不能再给舞弊者做嫁衣。我们必须拥护和平，所以请回家吧。我们爱你们，你们如此与众不同。你们看到了所发生的事情，看到他们对待其他人的方式是如此的恶劣与邪恶。我知道你们的感受，但是请回家吧，和平地离开。



拍摄者：Tyler Merbler

1 小时 20 分钟后，Meta 以违反其关于危险人物和组织的社群守则为由撤下了这段视频。下午 6:07，当警方正在保卫国会时，特朗普在 Facebook 上发表书面声明，表示：

这一幕幕场景，都是在伟大的爱国者们以压倒性优势获得神圣的选举胜利，又被人毫不客气、不坏好意地夺走时发生的...他们受到不公正且恶劣对待的时间已经太久了。请心怀爱与和平，然后回家吧。让我们永远记住这一天！

8 分钟后，Meta 以违反其《社群守则》的危险人物和组织规定为由，从 Facebook 上删除了这篇帖子，禁止特朗普在 Facebook 或 Instagram 平台上发布内容，封禁期为 24 小时。

第二天，也就是 1 月 7 日，Facebook 在进一步审核特朗普的帖子、他近期在 Facebook 外的言论，以及有关国会大厦暴力事件严重性的更多信息后，宣布将封禁期“无限期延长，至少延长至未来两周完成和平的政权交接之后”。

Meta 向委员会提请复审

两周后的 1 月 21 日，也就是乔·拜登 (Joe Biden) 总统就职典礼第二天，Meta 将无限期禁止特朗普访问其 Facebook 和 Instagram 帐户的决定提请委员会复审。他们还请求监督委员会就封禁政治领导人帐户的事宜提供意见或建议。

Meta 无限期封禁特朗普的 Instagram 帐户和 Facebook 页面的做法毫无意外地吸引了公众、政界和媒体的强烈关注，并引发了关于社交媒体平台内容审核政策的公平性、复杂性和局限性的激烈讨论。委员会指出，这也让各界重新关注“对 Facebook 等公司最重要的决定进行独立监督的价值”。我们在

努力解决这一冲突时发挥的关键作用促使《华盛顿邮报》发表了如下评论：

放眼世界，所有渴望为社交媒体行业设计新框架的官员和政策制定者都在密切关注监督委员会。倘若它的成功获得认可，就有可能降低监管呼声。倘若它以失败告终，可能会迫使许多国家/地区为内容审核加速建立更严格的法律屏障。

《纽约时报》简明扼要地总结了其中的利害关系。

这项决定的意义重大，不仅会影响美国政治，还会影响社交媒体监管方式，而在几乎不存在合法权力的当下，它还会对一种可能出现的新型跨国公司权力产生影响。

委员会在做出决定时考虑了 Meta 在 Facebook 和 Instagram 上对特朗普禁言是否符合 Meta 自己的政策、价值观和国际人权标准。

委员会参考了多个信息来源，包括：Meta 提供的决策理由，以阐明其决定背后的原因；Meta 对委员会提出的具体问题做出的回应；通过美国法律和司法中心和一位页面管理员代表特朗普提交给委员会的《内容创作者声明》；以及独立研究。

Meta 的解释

当 Meta 必须在紧急情况下做出影响广泛的决定时，它往往会通过 Newsroom 中的帖子分享这项决定背后的原因。具体到本案，Meta 表示之所以在 2021 年 1 月 6 日删除那两篇帖子，是它们因为违反了《社群守则》关于危险人物和组织的规定。具体来看，该内容被删除的原因是其违反了禁止“赞扬、支持和代表认定的暴力事件”这项规

5 “Facebook 新推出的“最高法院”推翻了公司的首批裁决”，Elizabeth Dwoskin, Craig Timberg, 《华盛顿邮报》，2021 年 1 月 28 日

6 “特朗普希望重返 Facebook。这个众星云集的陪审团可能会让他得逞”，Ben Smith, 《纽约时报》，2021 年 1 月 24 日

定。Meta 坚持认为，该内容还违反了禁止“对参与有组织暴力行为的个人予以赞扬”的规定。

Meta 指出，虽然特朗普的确在他的视频中让人们“和平地回家”，但他还重申了选举舞弊指控，并使用“我与你们感同身受”这样的说法暗示了共同目标。考虑到他发表评论时的动荡局势以及他言辞的整体基调，Meta 认定：“我们爱你们，你们如此与众不同”意在赞扬违法冲击国会的人。该公司还认为，第二篇帖子也包含了对此事的赞扬，因为特朗普将冲击国会的人称作“伟大的爱国者们”，并呼吁人们“让我们永远记住这一天”。

至于随后无限期延长对特朗普 24 小时禁言时限的决定，Meta 表示，他们在进一步评估了不断演变的局势和国会大厦暴力事件的最新细节后，认为 24 小时禁令不足以解除“特朗普利用其 Facebook 公共主页和 Instagram 帐户引发进一步暴力行动的风险”。该公司援引了美国国土安全部发布的《国家恐怖主义咨询系统公告》，该公告称“美国各地的威胁环境加剧，国土安全部认为这种情况将在总统成功就职后的几周内持续存在”。

在拜登于 1 月 20 日就职后，Meta 继续实施“无限期”封禁，既没有说明确切的时间段，也没有披露明确的恢复标准，而是完全根据未来的自由裁量来决定封禁时间。该公司称，对特朗普无限期封禁的决定是根据国际人权标准制定的，具体参考了《公民权利和政治权利国际公约》第 19 条和联合国人权事务委员会的第 34 号一般性意见，其中允许在威胁国家存亡的突发公共事件中对言论自由进行必要和相称的限制。

Meta 有一条“允许保留具新闻价值内容”的规定，即当该公司认为某项内容“具有新闻价值且符合公众利益”时，可以将违反其政策的内容保留在该平台上。但 Meta 解释称，他们并没有在本案件中将此豁免权应用于有争议的帖子。

委员会维持了封禁帐户的初始决定，但撤销了无限期延长禁令的决定。

“

对特朗普的裁决不可能让所有人满意。但这篇 38 页的文档对思考如何在信息混乱的时代处理言论自由贡献不菲。”

委员会成员 Alan Rusbridger



我们在 5 月 5 日公布了决定，并提出了一组建议。委员会发现，考虑到违规行为的严重性以及持续存在的暴力风险，Meta 有理由在 2021 年 1 月 6 日封禁特朗普的帐户，并将禁令延期到 1 月 7 日。委员会认为，存在问题的两篇帖子的确违反了 Facebook 的《社群守则》和 Instagram 的《社群守则》。Facebook 的规定禁止发布对参与恐怖主义、有组织暴力或犯罪活动的团体、领导者和个人表达支持或赞扬的内容。

委员会在决定中指出：

在这些帖子发布时，国会大厦的暴力活动正在进行中。这两篇帖子都对参与暴力活动的人表达了赞扬或支持。第一篇帖子中的“我们爱你们，你们如此与众不同”，以及第二篇帖子中的“伟大的爱国者们”和“让我们永远记住这一天”，都是在对那天参与国会大厦暴力事件的个人表示赞扬或支持。

但该决定也发现，Meta 实施的“含糊不清且无标准可循的无限期封禁处罚”有失偏颇，《社群守则》“不支持”无限期处罚。此外，决定进一步指出，这项处罚的模糊性“违反了言论自由原则……没有明确阐述根据什么标准触发这种处罚或者……根据什么标准保留或删除该内容。”

监督委员会拒绝了 Facebook 要求其“支持无限期封禁”的请求，因为这类限制措施的实行和撤销都没有明确的标准。我们坚持认为，“在本决定公布后 6 个月内，[Meta] 必须重新审议这项未经慎重考虑，于 1 月 7 日所做出的处罚决定，确定合适的处罚措施。此处罚必须考虑到违规行为的严重程度和将来可能造成的危害。它还必须符合 [该平台] 有关严重违规行为的规定，而且这些规定必须是明确、必要和相称的。”



我们要 [对 Meta] 说的是，他们不能突发奇想地发明惩罚措施，而是必须遵守既定规则。”

委员会联合主席
Helle Thorning-Schmidt



Meta 回应委员会的决定

Meta 于 6 月 4 日对委员会的决定作出回应，包括对其无限期处罚进行复审，并根据适用于其他 Facebook 用户的规定做出相称的处罚。该公司宣布了一套“适用于此类特殊情况”的最新执行规程。根据这些规程，Meta 宣布了对特朗普帐户封禁两年的最新决定，自 2021 年 1 月 7 日起生效。Meta 坚持认为：“鉴于导致特朗普被封禁一事的严重性，我们认为他的行为严重违反了我们的规定，应当根据新的执行规程给予最高处罚。”当处罚到期日临近时，Meta 承诺将

聘请专家评估公共安全风险是否已经降低。我们将评估外部因素，包括暴力事件、和平集会限制和其他内乱标志。如果我们确定公共安全仍然面临严重风险，就会延长限制时间，并继续重新评估，直至风险消除。当禁令最终解除时，将有一套能迅速升级的严格制裁措施。如果特朗普今后继续违规，就将触发这些制裁措施，直至并包括永久删除其主页和帐户。

确保尊重言论自由

在委员会 2021 年做出的 20 项决定中，有许多需要 Meta 恢复我们认为被误删的内容。我们发现，Meta 在这些案件中的行动与该公司的规定、价值观和国际人权标准承诺不符。在下面 5 起代表性案件中，我们发现 Meta 限制用户言论的决定在当时的情况下是不合理的。



旁遮普语用户提出对印度 RSS 的担忧

公布日期	国家/地区	社群守则
2021 年 4 月 29 日	印度	危险人物和组织

2020 年 11 月，一名用户分享了旁遮普语网媒公司 Global Punjab TV 发布的一篇视频帖。这是一段 17 分钟的采访视频，采访对象是被称为“社会活动家及旁遮普文化支持者”的教授。在帖子的配文中，该用户声称印度教民族主义组织国民志愿服务团（Rashtriya Swayamsevak Sangh，简称 RSS）和印度执政党印度人民党（Bharatiya Janata Party，简称 BJP）威胁要杀死锡克教徒（印度的一个少数宗教团体）。该帖子还指控 RSS 前领导人，即印度总理穆迪，在 RSS 主席的建议下，威胁“锡克教种族灭绝”。

在收到一名用户举报后，人工审核员以违反 Facebook 的“危险人物和组织”社群守则为由删除了这个帖子。此举也使得系统自动对发帖用户的帐户施加了限制。这名用户随即向 Meta 提出申诉。Meta 以新冠疫情导致审核能力暂时下降为由拒绝审核该用户的申诉。由于该案件被委员会选中，Meta 后来恢复了这条内容，并承认其最初的决定是错误的。

委员会在决定中表示，视频中没有内容违反 Meta 的政策。该帖子只是提出了印度少数群体和反对派的担忧，据称，他们正受到政府的歧视。在回应 Meta 关于新冠疫情导致其缺乏用户申诉审核能力的说法时，我们重申了乳腺癌症状和裸露决定中的观点，即所有案件在提交给委员会前都应先向 Meta 提出申诉。

关于 Meta 在执行和回应我们决定的及时性方面的报告

- 根据我们的《细则》，Meta 必须在我们的决定公布后 **7 天** 内予以执行。
- 在我们 2021 年公布的 20 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。
- 对于我们要求其采取行动的几乎所有决定，Meta 都在我们发布决定当天恢复或删除了相应的帖子。
- 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。

“双按钮”表情包

公布日期	国家/地区	社群守则
2021 年 5 月 20 日	美国、亚美尼亚、土耳其	仇恨言论；残忍和无情

2020 年 12 月 24 日，一位美国 Facebook 用户发布了一条评论，将源自喜剧《Daily Struggle》的“双按钮”表情包修改成了具有土耳其政治背景的内容。这个表情包（2014 年发布）里面有一个卡通角色试图在两个内容可能相互冲突的按钮之间做出选择。该帖子将卡通人物的脸换成了一面土耳其国旗，上方的两个红色按钮分别用英语写着：“The Armenian Genocide is a lie”（亚美尼亚种族灭绝是个谎言）和“The Armenians were terrorists that deserved it”（亚美尼亚人是恐怖分子，他们罪有应得）。一名人工审核员认为该表情包违反了 Facebook《社群守则》有关仇恨言论的规定，而另一名审核员则认为它违反了 Facebook《社群守则》有关残忍和冷漠内容的规定。于是，Facebook 将其删除并告知该用户。用户向 Meta 提出申诉，后来又申诉到委员会。

这一次，背景同样是关键。Meta 的删帖决定是因为其中这样一句话：“亚美尼亚人是恐怖分子，罪有应得。”但我们发现，如果放到在帖子发表的背景中去理解就会发现，那个帖子实际上是在批评土耳其政府在亚美尼亚种族灭绝问题上一味否认的态度。因此，该帖子符合“仇恨言论”社群守则规定的例外情况，即允许用户“出于谴责或提高认知之目的”发布仇恨言论。我们还认为，该内容符合 Facebook 规定的讽刺内容例外——我们呼吁该公司将其纳入面向公众的《社群守则》。





奥贾兰单独监禁

公布日期	国家/地区	社群守则
2021 年 7 月 8 日	土耳其、美国	危险人物和组织

委员会 2021 年 7 月的这项决定凸显了对 Meta 的行为进行独立审查的重要性，因为它揭示了一份被遗漏三年之久的政策准则。

2021 年 1 月 25 日，一位美国 Instagram 用户发布了库尔德工人党创始人之一阿卜杜拉·奥贾兰 (Abdullah Öcalan) 的照片。

其中的英文说明鼓励读者讨论结束奥贾兰在土耳其伊姆拉里岛监狱的单独监禁，以及单独监禁的不人道性质。根据 Facebook 的“危险人物和组织”政策，库尔德工人党和奥贾兰被认定为“危险实体”。此外，该政策还禁止“赞扬”或“支持”认定的实体。

在审核员以违反“危险人物和组织”政策删除该贴后，用户对这一决定提出了申诉。在申诉遭 Meta 驳回后，该用户又向监督委员会提出了申诉。

该决定的关键是在委员会选中该案件后才浮出水面的，这也凸显出我们对 Meta 的行为进行独立审查的重要性。该公司披露，他们找到了 2017 年制定的一份内部准则，其中明确允许用户讨论被认定为危险人物的监禁条件。之所以起草这一政策，部分原因就是为回应人们对奥贾兰监禁条件的关注，并允许人们在该平台上讨论这个问题。在撤销 Meta 的删帖决定时，我们表达了对 Meta 缺乏透明度的担忧：“在移除内容时显示的默认政策为‘支持’经认定的危险人物，同时关键的例外情况并未对外公开，从而使得这个错误持续相当长一段时间，而没有被任何人发现。”

我们的决定涉及最多的社群守则是哪些？



仇恨言论

9 项决定



危险人物和组织

5 项决定



暴力与煽动暴力

2 项决定



分享半岛电视台的帖子

公布日期	国家/地区	社群守则
2021 年 9 月 14 日	以色列、埃及	危险人物和组织

2021 年 5 月 10 日，一位埃及 Facebook 用户分享了一篇半岛电视台阿拉伯语认证公共主页发布的帖子，其中包含一段阿拉伯文字和一张照片。照片上有两名身穿迷彩服的蒙面男子，他们还戴着印有卡萨姆旅徽章的头带。

帖子中的文字写道：“联合行动委员会的抵抗运动领导人决定给占领军一个喘息的机会，在 18:00 前将其士兵从阿克萨清真寺和谢赫贾拉区撤离，否则后果自负。阿布·乌拜达 (Abu Ubaida) 一卡萨姆旅军事发言人。”该用户分享了半岛电视台的帖子，并用阿拉伯语单字配文“Ooh”（哦）。根据 Facebook 的“危险人物和组织”社群守则，卡萨姆旅及其发言人阿布·乌拜达均被认定为危险实体。Meta 根据这项政策删除了该贴，但因为委员会选中该案件，所以又恢复了这个帖子。

Meta 无法解释为什么两名人工审核员最初都认定该内容违规。他们表示，审核员无需逐一记录他们做出每项内容决定的原因。委员会的决定强调，虽然 Facebook 的“危险人物和组织”政策禁止“代表认定的实体传播信息或资源，包括官方消息”，但也将作为“新闻报道”发布的内容列为例外情形。本案件中的内容是转发半岛电视台一条广泛传播的新闻报道，内容是关于 2021 年 5 月以色列军队与巴勒斯坦激进组织在以色列和加沙（ Hamas 统治的巴勒斯坦领土）发生武装冲突。除了增加“Ooh”（哦）这句非实质性的评论外，并没有其他改动。我们因此得出结论，该帖子属于 Facebook 的新闻报道例外情况。

本案件还涉及到公平对待用户的问题，因为这个帖子转发的是一家合法机构发布的新闻，但却与该新闻机构本身发布的内容获得了不同的待遇。我们在决定中指出，只要媒体机构有权发布新闻，个人就有权转发新闻。我们还注意到，半岛电视台的页面可以享受 Meta 为某些著名帐户提供的交叉检查系统，但分享该内容的用户却没有享受这一待遇。



哥伦比亚抗议

公布日期	国家/地区	社群守则
2021 年 9 月 27 日	哥伦比亚	仇恨言论

2021 年 6 月，哥伦比亚一家地方性新闻机构的 Facebook 公共主页分享了一段视频，显示抗议者在一个写着“SOS COLOMBIA”（救救哥伦比亚）的横幅后面游行。从这段视频中可以听到西班牙语的呼喊声和歌唱声，抗议者用憎恶同性恋的诋毁性言论谴责哥伦比亚总统伊万·杜克 (Ivan Duque)。Meta 以违反“仇恨言论”社群守则为由删除了这段视频（该视频最早发布于 Tik-Tok）。该守则禁止“以诋毁性措辞来描绘或消极攻击他人……用作侮辱性标签”。该政策尤其禁止对包括性取向在内的受保护特征进行负面描述的“侮辱性标签”。

该用户对 Meta 删除视频的决定提出申诉时自称是报道省内新闻的当地记者。该用户还坚称，他们分享这段视频不是为了造成伤害，只是为了展示一群年轻人使用有代表性的用语和平抗议并呼吁权利。

委员会决定 Meta 应当恢复视频。我们承认，这段视频表面看来的确违反了 Facebook 的“仇恨言论”社群守则，因为分享诋毁性言论会导致周围的环境威胁和排斥 LGBT 人群，但 Meta 的删帖行为依然是错误的，因为它没有在审核该帖时考虑“允许保留具新闻价值内容”这条原则。这一规定可能允许保留违反守则的内容，具体取决于公共利益价值和伤害风险。



保护用户免受有害内容的伤害

委员会在 2021 年做出的某些关键决定中维持了 Meta 的删帖决定，发现在某些情况下限制用户言论是合理措施。这些决定也明确表示，Meta 的删帖行为可以与该公司的人权责任保持一致，特别是当删除一个用户的言论是尊重他人权利的的必要和相称举措时。



扮成“黑彼得”

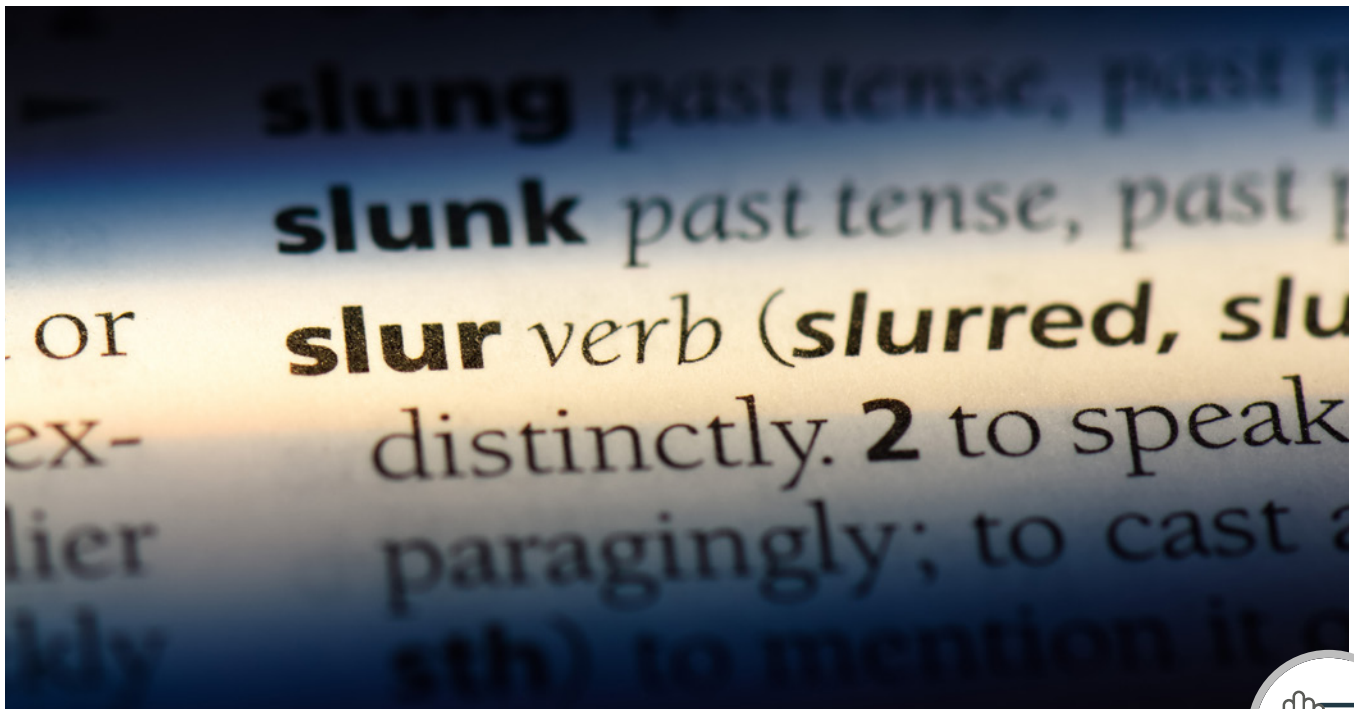
公布日期	国家/地区	社群守则
2021 年 4 月 13 日	荷兰	仇恨言论

2020 年 12 月 5 日，一位荷兰的 Facebook 用户在其时间线上分享了一篇帖子，其中配有荷兰语文字并包括一段 17 秒长、描述一名儿童与三名成年人见面的视频。其中一名成年人打扮成圣尼古拉斯 (Sinterklaas)，即圣诞老人，另外两名成年人的面部是黑色的，帽子下戴有爆炸头假发，身穿色彩鲜艳的文艺复兴时期风格的服装，扮演的是荷兰漫画中的人物“Zwarte Piet”。这个名字可以翻译成“黑彼得”。

我们维持了 Meta 的删帖决定，理由是这个帖子违反了“仇恨言论”社群守则。多数委员会成员都认为，该内容包含的漫画“与负面和种族主义成见存在密切联系，并被荷兰社会的部分人士视为一种在荷兰维持系统性种族主义的表现”。在做出这一决定时，委员会的大多数成员参考的一些研究表明，“圣尼古拉斯节期间，黑人儿童在家里感到害怕和不安全，并且不敢去上学。”

虽然少数委员会成员认为，没有足够的证据将这条内容与据称可通过移除这条内容减少的伤害直接联系起来，但大多数委员会成员发现，允许这类内容在平台上积累将助长营造贬低和骚扰黑人的歧视性环境。大多数委员会成员指出：

从大范围来看，[Meta 的] 政策清楚了，可以确保黑人在平台上的尊严、安全和表达自由。为了实现这一目标，当用户分享了涉及“扮黑脸”的内容且意图不是谴责种族主义时，限制其表达自由的做法是可以接受的。



南非诋毁性措辞

公布日期	国家/地区	社群守则
2021 年 9 月 28 日	南非	仇恨言论

2021 年 5 月，一名 Facebook 用户在一个自称致力于“解放思想”的公共小组中用英语发布了一篇帖子。该帖子讨论了南非的“多种族主义”，并指出，自 1994 年以来，南非黑人的贫困以及无家可归和无土地的情况愈加恶劣。这位用户表示，白人拥有并控制着大部分财富，而富有的黑人可能拥有一些公司的所有权，但却没有控制权。这篇帖子对持有相反观点的人提出挑战，让他们“检查一下脑袋”，并使用种族主义诋毁性措辞描述黑人群体。Meta 认为这些措辞违反了“仇恨言论”政策。

我们维持了 Meta 删除这条内容的决定，发现其中使用的一个词汇“在南非的语境中极具仇恨性和伤害性”。我们强调，公开讨论社会经济的不平等问题是推动平等的关键。然而，我们认为完全可以在不使用诋毁性措辞的情况下参与这种讨论并引起受众共鸣。此外，其中一个诋毁性措辞与该国的种族隔离历史有明确关系，至今仍会与贬低和排斥黑人联系起来。该决定重申，当用户的内容因为违反“仇恨言论”政策遭到删除时，Meta 就违反的规则进行明确解释是非常重要的。



在 Raya Kobo 涉嫌犯罪

公布日期	国家/地区	社群守则
2021 年 12 月 14 日	埃塞俄比亚	仇恨言论

2021 年 12 月的一项决定发现一篇包含针对某个族群的无法证实的传言的帖子违反了 Meta 的社群守则，而且删除该帖的做法符合人权原则。

2021 年 7 月下旬，一位来自埃塞俄比亚的 Facebook 用户用阿姆哈拉语发帖指控，埃塞俄比亚左翼政党提格雷人民解放阵线（Tigray People’s Liberation Front，简称 TPLF）不仅奸淫杀戮，残害妇女儿童，还在 Raya Kobo 和阿姆哈拉地区其他多个城镇实施抢劫，掠夺平民的财产。这位用户在帖子中声称提格雷族平民助纣为虐，协助 TPLF 犯下了这些暴行。该用户在帖子最后说道：“我们会奋起反抗，努力捍卫我们的自由。”此外，该内容发出后 24 小时已被数千名使用阿姆哈拉语的用户浏览，有些人还在评论中呼吁实施报复。

Meta 阿姆哈拉语内容审核团队的一位内容审核员认为，该帖子违反了 Facebook 的“仇恨言论”社群守则，并将其删除。这项政策禁止根据种族、民族或原国籍，使用“暴力言论”攻击某人。这位用户就该决定向 Meta 提出申诉，于是由 Meta 阿姆哈拉语内容审核团队的另一名审核员进行二次审核，证实该帖确实违反了 Facebook 的政策。

由于委员会选择复审该案件，Meta 认为移除这篇帖子属于“执行失误”，并将其恢复。Meta 之所以决定恢复该帖，是因为他们的评估发现，其中的言论没有达到仇恨言论的程度。

我们却持不同意见，我们发现 Meta 的解释“缺乏细节且存在错误”。我们对该帖适用了“暴力与煽动暴力”社群守则，发现它违反了 Facebook 禁止发布“有可能在近期诱发暴力活动或造成人身伤害的错误信息和无法证实的传言”的规定。委员会认为“埃塞俄比亚现在正处于暴力冲突时期，这时指控某个民族助纣为虐，参与大规模的暴行非常危险，会大大增加近期诱发暴力事件的风险”。


委员会决策 流程中的国际 人权标准

我们的责任之一是确保 Meta 的内容审核模式能以尊重人权为基础。这也反映在我们的《章程》和 Meta 自己的公司人权承诺中。因此，我们的决定不仅会审查 Meta 的内容决定是否遵守该公司的内容规定和价值观，还会判断这些决定是否符合国际人权标准。

国际人权标准提供了一套保护言论自由和其他人权的通用框架，还提供了一套全球一致的框架，使 Meta 为它在全世界的影响力负责。Meta 的《公司人权政策》阐明了它尊重人权的承诺，包括内容政策的制定方式。该政策还指出，《联合国商业与人权指导原则》(UNGPs) 为委员会的成立提供了依据。

我们相信，将人权作为我们工作的核心并推动 Meta 尊重人权，将帮助 Facebook 和 Instagram 为用户提供更好的使用体验。假

“
是时候讨论如何创造尊重人权的
技术了。”
委员会成员 Julie Owono



以时日，我们希望借助自己的决定使 Meta 的内容审核模式更符合国际人权标准。

新挑战

在 Meta 这种公司的内容审核中应用国际人权标准是一个相对较新的挑战。委员会是目前唯一一家在从事这种工作的实体，可谓孤军奋战。

使用主要由国家/地区制定且适用于国家/地区的法律框架来解释一家社交媒体公司的人权责任会引发许多问题。与在其他地方进行言论监管的政府相比，Meta 的内容审核应该在多大程度上遵守同样严格的规定？这种模式应当如何看待大规模内容审核面临的挑战？我们应该如何探讨 Meta 的设计选择可能带来的伤害，包括算法对内容的放大效应？我们的决定探讨了如何利用国际人权标准来界定 Meta 的人权责任，这可能与适用于国家/地区的责任有所不同。例如，我们在扮成“黑彼得”和亚美尼亚人侮辱阿塞拜疆两个案件的决定中维持了 Meta 的删帖决定，同时认识到国际人权标准可能不允许一个国家/地区通过刑事或行政措施对同样的言论进行惩罚。

建立植根于人权的组织

作为一个组织，我们的决定会对世界各地的人产生影响，所以在工作的方方面面都必须尊重人权。委员会的《章程》要求委员会在做出决定时特别关注人权问题。我们竭力遵循《联合国商业与人权指导原则》，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，并以合作和对话为基础。2020年，委员会委托非营利组织商务社会责任国际协会 (BSR) 编写的一份跟进报告指出，委员会根据其之前的报告提出的建议取得了“良好进展”。

人权如何支撑我们的言论自由模式

委员会在做决定时以各类国际人权标准作为权威指导来源，其中《公民权利和政治权利国际公约》(ICCPR) 第 19 条的引用次数最多。ICCPR 是 Meta 在其《公司人权政策》中引用的一项全球人权公约，自愿承诺尊重其中描述的人权。事实证明，联合国意见

和言论自由问题特别报告员的工作，特别是 2018 年关于内容审核的报告（该报告探讨了私有公司如何才能更好地尊重人权），对委员会来说特别有价值。

ICCPR 第 19 条第 2 款写道：“人人有自由发表意见的权利；此项权利包括寻求、接受和传递各种消息和思想的自由，而不论国界，也不论口头的、书写的、印刷的、采取艺术形式的、或通过他所选择的任何其他媒介。”然而，根据国际人权标准，委员会也不认为自由言论权是绝对的。虽然这项权利很重要，但在某些情况下依然可以加以限制。ICCPR 第 19 条第 3 款为评估言论自由的限制措施提供了三步测试法，考虑了合法性、合法目的以及必要性和相称性。这些原则与 Meta 运营的许多国家/地区的国内法和宪法（包括《美国宪法第一修正案》）规定的言论自由标准相似，但不完全相同。我们也在必要时考虑国际人权法对言论的限制，包括禁止鼓吹构成煽动敌意、歧视或暴力的仇恨主张 (ICCPR 第 20 条第 2 款)。在决定中的合法性分析中，我们调查了 Meta 的规定对 Facebook 和 Instagram 用户的明确度。规则应该方便用户查找和理

各大组织对委员会工作的反应

学术界、民间团体以及国家和国际机构已经开始分析委员会的工作和决定。重点包括：

- 2020 年 12 月，联合国少数群体问题特别报告员 Fernand de Varennes 敦促我们考虑少数群体的权利，称委员会是“一个创新且有雄心的网络言论监管项目，尤其是在仇恨言论方面，对在全球范围内有效保护处于弱势的少数群体至关重要。”
- 2021 年 3 月，国会法律图书馆的一名法律研究员分析了委员会的首批 5 项决定如何将国际人权法应用于社交媒体公司。
- 2021 年 9 月，包括 Access Now、Article 19、7amleh 在内的组织都对委员会“分享半岛电视台的帖子”案件的决定表示欢迎，认为这是“向着正确方向迈出了一步”，称该案件“体现了 Facebook 的系统性武断和对‘危险人物和组织’社群守则不透明的过度执行”。
- 2021 年 12 月，美国国际宗教自由委员会提到我们将人权标准应用于内容审核工作。
- “Lawfare”还设立了一个全面的监督委员会博客，汇总了迄今为止针对我们的决定发表的评论和分析。
- 我们的决定还被国际法律数据库 Westlaw 和 Lexis/Lexis+ 收录。

解。我们在许多决定中都敦促 Meta 公布其规定中不为公众所知的部分，包括被该公司认定为“危险的”人物和组织名单。

在决定中的合法目的分析中，我们研究了 Meta 制定 Facebook 和 Instagram 规定时所追求的目标，是否出于国际人权法规定的限制言论的合法目的。我们的多数决定都认为，尊重“他人的权利”是 Meta 限制言论的内容政策背后的合法目的。

在我们的必要性和相称性分析中，我们调查了 Meta 在特定案件中的行为是否是实现合法目的的必要举措，以及删除内容等措施是否是实现这一目的的最低限制性手段。

在首批 20 项决定中，我们因为 Meta 的某些内容规定过于模糊、过于宽泛或不够明确而提出过疑虑，并建议他们澄清规定，或者将内部对这些规定的秘密指导性解释对外公开。委员会特别重视政治内容和提高人权违规意识的内容（包括政治讽刺），以及艺术表达和对健康问题的讨论。对于可能引发伤害，但不会直接导致紧迫伤害的内容，我们还质疑删帖行为是否总是相称的应对措施。例如，在两个与新冠疫情有关的决定中（宣称新冠肺炎疗法和巴西新冠疫情封城），我们询问是否可以通过其他不删帖的措施（如警告界面、添加标签或降低内容的排名）在保护表达权的同时降低风险。

除了上文提到的三步测试法，我们还将平等和歧视原则融入到决策过程中，认识到不同人可能会在不同情况下因为 Meta 的内容审核方法受到不同的影响。因此，我们才建议 Meta 在受冲突影响的地区用阿拉伯语和阿姆哈拉语评估人权影响，并呼吁其改善对缅甸语内容的审核。委员会在乳腺癌症状和裸露决定中对平等和非歧视问题的关注也表明，Meta 对自动化技术的依赖导致其错误地删除了女性特有的健康信息，因而会对用户产生不均衡的歧视性影响。出于相似的原因，我们也在关注不以政治观点为由进行歧视的重要性。

权利与规定 — 俄罗斯支持纳瓦尔尼的抗议活动

在我们的多数决定中，我们的人权分析与我们对 Meta 规定的解读得出了相同的结论。然而，在俄罗斯支持纳瓦尔尼的抗议活动决定中，尽管删帖行为符合 Meta 的规定，但却首次基于 Meta 的人权责任撤销其决定，这也成为委员会的一个里程碑事件。

2021 年 1 月 24 日，一位俄罗斯用户发布了几张照片、一段视频和一段文字，描述了前一天在圣彼得堡、莫斯科和该国其他地方爆发的声援反对派领袖亚历克西·纳瓦尔尼 (Alexei Navalny) 的抗议活动。另一名用户（即“抗议活动批评者”）回应了该帖，称虽然他们不清楚圣彼得堡到底发生了什么，但莫斯科的抗议者都是思维“迟钝”的小学生，遭到了“无耻地利用”。然而，另一位自称参加了圣彼得堡抗议活动的用户在评论的最后称这名抗议活动批评者是“懦弱的网络机器人”。随后，抗议活动批评者以存在欺凌和骚扰为由，向 Meta 举报了该抗议者的评论。

Meta 以违反“欺凌和骚扰”守则为由删除了那篇帖子。根据该守则，普通个人可以向 Meta 申请撤下包含“对其人格做出负面评论”的帖子。委员会决定支持那位抗议者恢复帖子的要求，并观察到：

此案件凸显了保护人们免受欺凌和骚扰的政策与保护言论自由的需要之间的紧张关系。在涉及政治抗议活动，且有可信投诉表明此类活动发生的国家/地区缺少有效的人权保护机制时，考虑上述矛盾尤其重要。

我们的决定认为，尽管删除内容的行为符合 Meta 的“欺凌和骚扰”社群守则，但在国际人权标准下，这种做法却是对言论自由不必要、不相称的限制。

以维护用户的表达权为出发点，委员会发现，虽然 Meta 删除该内容的原始决定“可能与严格执行社群守则相符”，但过于严格地执行这些守则会导致“未能考虑到更

广泛的背景，对言论自由施加了不相称的限制。”

该案件表明，当 Meta 的内容政策与其人权责任发生冲突时，委员会将优先考虑人权。

针对具体标准的案件研究

委员会的首批 20 项决定考虑了众多国际人权标准。2021 年公布的每一项决定都参考了《公民权利和政治权利国际公约》第 19 条、《联合国商业与人权指导原则》和联合国人权委员会关于意见和言论自由的第 34 号一般性意见。将这些标准和其他国际人权标准应用于我们的决定，使我们可以更加深刻地理解和分析推动许多内容审核决定的根本问题。这也帮助我们得出了更合理的结论。虽然每项决定都参考了多项人权标准，但我们还要在下文举例说明对我们的决定有重要影响的标准。

《联合国商业与人权指导原则》

Meta 曾经删除了一个配有贝壳串珠带（一种北美原住民艺术形式）图片的帖子，标题是“杀死印第安人/拯救人类”。我们在关于此案的决定中敦促 Meta 根据《联合国商业与人权指导原则》的第 17 条，履行对边缘化社群的人权责任。美国印第安人事务协会的一条公众评议意见揭示了这句引文的文化背景。该协会确认，引文的作者名叫理查德·亨利·普拉特 (Richard Henry Pratt)，是美国第一所联邦印第安人寄宿学校的创始人。结合上下文来看，引文很明显属于“反驳言论”，即故意引用仇恨言论来抵制（而非支持或宣扬）压迫和歧视

在回应本案件时，我们敦促 Meta 根据《联合国商业与人权指导原则》的第 17 条，履行对边缘化社群的人权责任。这样便要求了 Meta 开展人权方面的尽职调查，以确保系统

公平运作，不会加剧过去和现在仍在持续的压迫手段。其中包括识别、避开和减轻对原住民反对歧视的言论进行内容审核造成的不利影响。最后，我们提出了自己的期望，即 Meta 应优先考虑边缘化群体面临的风险，并就 Meta 如何改进提供证据。

第 34 号一般性意见

缅甸网络水军案件关注的是在该国发生非民主政变后发布的内容，在该案件中，我们向 Meta 澄清了必须将用户对政府的批评言论与用户基于某些人的国籍而发表的歧视性言论区分开来。在这种情况下，委员会借鉴了联合国人权事务委员会第 34 号一般性意见，该意见为各国如何解释用于保护全球范围内言论自由的《公民权利和政治权利国际公约》第 19 条提供了指导。针对缅甸网络水军案件的决定具体参考了一般性意见中涉及政治言论重要性的部分，特别是允许公共机构进行不受限制的讨论的价值观。虽然言论自由可能应该在某些情况下受到限制，以保护个人免受导致特定伤害的歧视性仇恨言论的影响，但人们应该自由地表达他们对民族、国家和公共机构的看法。

获得补救的权利 (ICCPR, 第 2 条)

奥贾兰的单独监禁案件的决定援引了联合国言论自由问题特别报告员的声明，该声明称补救程序“应包含一个透明且可获得的申诉流程，方便用户就平台决定提出申诉，然后由公司作出合理的公开回应。”

然而，在这个案件中，Meta 的做法并没有达到这些标准。起初，Meta 告诉该用户，受新冠疫情影响，他们无法审核向他们提出的申诉。我们也表达了对“什么情况下可能会有大量本不应删除的操作”的担忧，因为 Meta 在三年来未对用户允许发表的内容作任何明

确的政策指导。Meta 声称，在审核员未获得政策指导的情况下，因技术限制，无法确定删除了多少条内容。对于这些问题，以及其他一些问题，我们批评 Meta “没有尊重用户获得补救的权利，从而违反了其企业人权政策。”

一项艰巨但重要的任务

我们认识到，使用国际人权标准来解读社交媒体公司的人权责任是一项全新的具有挑战性的工作。

然而，尽管这是一项艰巨的任务，但却是一项重要的任务。在内容审核时使用国际人权标准可以帮助我们确保 Meta 在尊重其他人权的同时尊重言论自由，并为追究社交媒体公司责任提供统一的全球参照标准。

委员会的核心使命是打造可展现基于人权的内容审核做法的工作体系。2022 年，在我们继续确保尊重 Meta 平台用户人权的同时，也将着眼于早期决定所依据的原价值观。

委员会向 Meta 提出的帮助做出决定的问题

为了帮助我们做出决定，并迫使 Meta 尽可能保持透明，我们会在审议案件之前向 Meta 发送问题。通过提出具体的问题，并且在做出最终决定之前了解更多细节，我们可以向 Meta 追究责任，并向研究人员透露有关 Meta 运作方式的新信息。

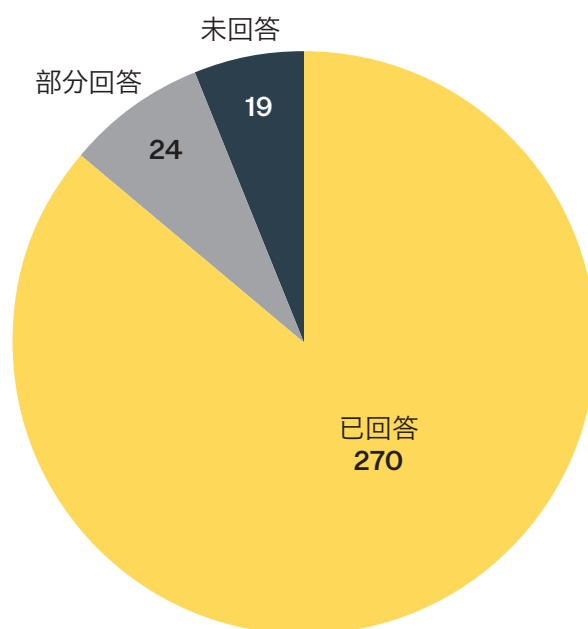
针对我们在 2021 年公布的 20 项决定，我们提出的问题涉及许多不同领域。在许多情况下，这些问题探讨了用户在向委员会发表的声明中强调的具体问题，以及通过公众评议提出的问题。其中许多问题都与 Meta 如何对待那些就内容提出申诉的用户有关，比如他们的内容被删除的原因，以及 Meta 是否审核了他们最初的申诉。对于其他情况，我们向 Meta 询问了关于规定、自动化系统和区域团队的更多细节。

为了帮助做出前面提到的 20 个决定，我们向 Meta 提了 313 个问题。其中，Meta 回答了 270 个，部分回答了 24 个，有 19 个未回答。为了透明度起见，我们提供这些数字，而对于 Meta 是否有充分的理由在任何情况下不回答这些问题，在此不做声明。

在委员会成立的第一年，Meta 普遍对我们的询问作出了回应，回答了绝大多数 (86%) 的问题。对于 Meta 未回答的问题，我们在季度透明度报告和某些决定中都予以了提出。在本节的后面部分，我们会对 Meta 未回答的问题及其给出的理由进行详细的说明。

在我们围绕早期案件提出的问题未能收到及时回应后，我们与 Meta 合作就分批制度达成一致意见，Meta 作出回应的期限可因委员会提出的问题数量而异。我们也感谢 Meta 能够亲自就与其规定相关的问题（包括交叉检查系统）作出简要的说明。

Meta 对监督委员会提问的回应情况



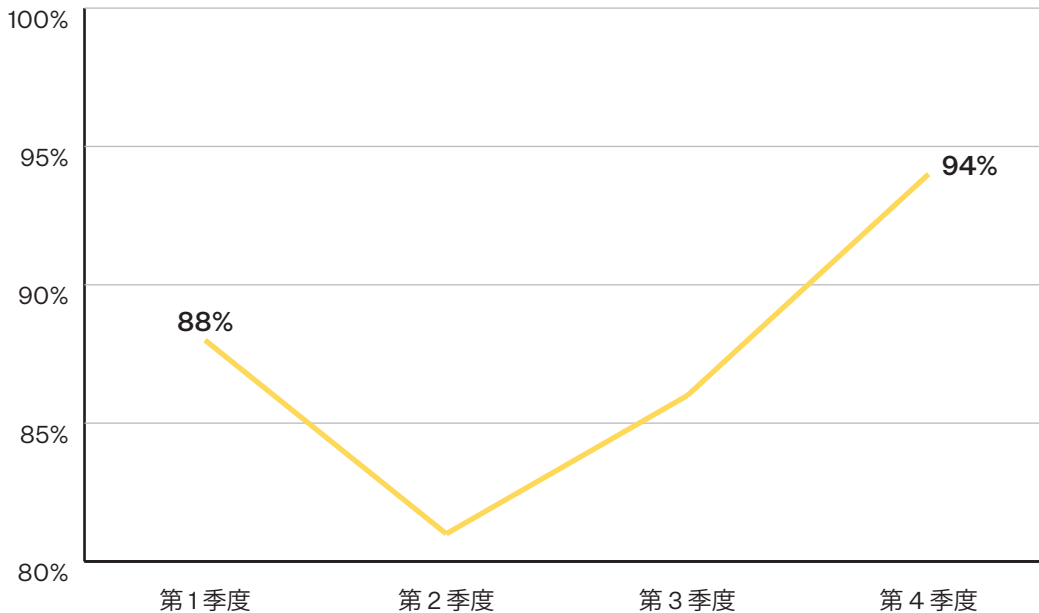
针对 2021 年第四季度公布的决定，Meta 回答的问题的比例有所上升

针对 2021 年第一季度至第三季度发布的决定，Meta 回答问题的比例在第一季度中达到 88%，而第二季度和第三季度分别为 81% 和 86%。然而，针对 2021 年第四季度发布的三项决定，这一比例上升到了 94%，Meta 回答了 54 个问题中的 51 个，部分回答了 3 个问题。在第四季度，Meta 首次做到了没有给出属于“未回答”的这一类回应。这一进步令人鼓舞，我们希望这一趋势可以持续到 2022 年。

Meta 没回答的问题的细节

为了帮助做出在 2021 年发布的那些决定，我们提出了 313 个问题，其中有 19 个问题 Meta 未回答。下面我们将就这些问题提供更

每个季度 Meta 已回答问题的占比情况



多细节，包括问题所涉及的案件，问题的具体内容，以及 Meta 给出的不回答的原因。

- Meta 未回答针对缅甸贬低穆斯林的帖子案件提出的两个问题。第一个问题是，Meta 之前是否对相关用户和群体执行过违反仇恨言论社群守则的处理。Meta 回应称，有关用户以前在平台上的行为的信息与委员会的决定无关。第二个问题是要求 Facebook 市场团队将帖子上的评论翻译成英文。Meta 回应称，在内容审核过程中，他们没有对评论执行自动翻译，也没有对评论进行人工翻译。
- Meta 未回答针对亚美尼亚人侮辱阿塞拜疆案件提出的一个问题。委员会询问的是，为何要删除该用户之前的一个帖子。Meta 回应称，有关用户以前在平台上的行为的信息与委员会的决定无关。
- Meta 未回答针对纳粹引语案件提出的一个问题。委员会询问的是，该用户是否有被禁止或暂停发帖的记录。Meta 回应称，有关用户以前在平台上的行为的信息与委员会的决定无关。
- Meta 未回答针对印度人抗议法国案件提出的一个问题。委员会提出的问题是，Meta 之前是否对该用户或群体执

行过违反暴力与煽动暴力社群守则的处理。Meta 回应称，有关用户以前在平台上的行为的信息与委员会的决定无关。

- Meta 未回答针对旁遮普语用户提出对印度 RSS 的担忧案件提出的两个问题。第一个问题是，内容中的哪些具体语言导致 Meta 以违反危险人物和组织社群守则为由删除了这个帖子。Meta 回应称，它无法识别是哪些具体语言导致了“内容已违反危险人物和组织政策”的这一错误结论。第二个问题包括询问，用户需要“违规”多少次才会让 Meta 施加帐户限制，以及需要违反危险人物和组织政策多少次才会被施以帐户级别限制。Meta 回应称，这些信息并非根据章程宗旨做出决定时需要的合理信息。
- Meta 未回答针对封禁前总统特朗普帐户案件提出的七个问题。Meta 未回答的问题是，Facebook 的动态消息和其他功能对特朗普先生发布内容的可见性产生了怎样的影响；Meta 是否研究过，或者计划研究与 2021 年 1 月 6 日事件相关的设计决定；以及有关特朗普先生帐户的粉丝发布的违规内容的相关信息。委员会还询问了有关封禁其他政治人物帐户和删除其他内容的问题；是否有政治官员或其工作人员就特朗普帐户被封一事联系过 Meta；帐户被封禁或删除是否会影响广告商瞄准

其粉丝帐户的能力。Meta 表示这些信息并非根据章程宗旨做出决定时需要的合理信息；由于技术限制而无法提供这些信息；这些信息受律师/委托人保密特权的保护；并/或由于法律、隐私、安全或数据保护方面的考虑而不能或不应该提供这些信息。

- Meta 未回答针对奥贾兰的单独监禁案件提出的三个问题。第一个问题是，Meta 是否可以确定在未执行政策时错误删除了多少条内容。第二个问题是，在过去的五年里，Meta 删除或保留了多少条涉及奥贾兰的内容。对于这两个问题，Meta 回应称，由于技术限制而无法确定此信息。第三个问题部分询问了，审核员为何认为该帖子呼吁人们采取行动支持奥贾兰和库尔德工人党 (PKK)。Meta 回应称，由于 Meta 不要求大规模审核内容的审核员记录他们对每个内容作出决定的理由，因此无法提供删除该帖子的依据的详细描述。

- Meta 未回答针对有关南非的诋毁性措辞案件提出的两个问题。第一个问题是，要求 Meta 提供帖子中使用的三个冒犯性词汇。Meta 回应称，由于技术限制，无法按要求提供相关信息。第二个问题是，Meta 在南非的公共政策团队是否包括来自南非的人。Meta 以委员会细则为由，拒绝按要求提供相关信息。

在 19 个 Meta 未回答的问题中，大约有三分之二（13 个）的问题与 Meta 根据危险人物和组织社群守则而删除内容的案件有关。其中包括纳粹引语案件、旁遮普语用户提出对印度 RSS 的担忧案件、封禁前总统特朗普帐户案件和奥贾兰的单独监禁案件。剩下的问题中，有 5 个与根据仇恨言论政策而删除内容的案件有关，1 个与根据 Meta 的暴力与煽动暴力政策而删除内容的案件有关。

公众评议

个人和组织可以提交公众评议，参与我们的决定制定流程。在委员会成立后的头一年里，公众评议让人们可以对决定的制定各抒己见，他们分享了语言、文化、政治和人权方面的重要专业知识。对于实现 Meta 改善世界各地用户和社群使用体验的这一目标而言，他们的意见至关重要。在许多情况下，公众评议会影响我们的决定和我们对 Meta 提出的建议。

针对我们 2021 年公布的 20 项决定，我们收到了来自世界各地的个人和组织发表的 9,986 条公众评议。其中的绝大多数 (97%) 是针对封禁前总统特朗普帐户案件的决定的；但必须要强调的是，在许多案件中，即使评议的绝对数量可能有限，我们也获得了许多优质且宝贵的专业知识。而且有些问题非常复杂，需要公众评议提供额外的专业知识。

96% 的公众评议来自美国和加拿大，这主要是因为美国针对封禁前总统特朗普帐户这一决定发表了大量评议。其中 95% 的公众评议来自于个人，仅 5% 来自于组织。

在发表的公众评议中，有三分之一的人匿名提交评议。有趣的是，许多人选择利用这个选项，分享他们专业知识的同时，保护自己的隐私权。在针对旁遮普语用户提出对印度 RSS 的担忧、“双按钮”表情包和哥伦比亚抗议这三项决定发表的评议中，有超过一半是匿名提交的，所有这些都与批评当前执政政府的内容有关。

公众评议如何影响我们的决定

对于委员会于 2021 年发布的 20 项决定，公众评议发挥了重要作用。评议表达了用户对案件的疑虑，并提供了可帮助做出决定的语言、文化和背景参考。2021 年期间，委员会发现提交的公众评议质量有所提高，在

2021 年收到最多公众评议的
监督委员会决定有：

1. 封禁前总统特朗普帐户
2. 亚美尼亚人侮辱阿塞拜疆
3. 巴西新冠疫情封城
4. 分享半岛电视台的帖子
5. 乳腺癌症状和裸露



该年早些时候公布的决定中明确参考或引用了更多评议意见。

“

通过公众评议，委员会可以和公民社会团体的声音站在一起，让大众听到这些声音；这一次，Meta 必须要有所回应。”

委员会成员 Afia
Asantewaa Asare-Kyei



提供语言专业知识

在多个案件中，公众评议提供了重要的语言背景。例如在缅甸网络水军案件中，公众评议认同委员会语言专家的观点，即在缅甸语中，那一个词可以同时用于指代某国家和该国的人民。这为委员会的结论提供了重要依据，因为帖子中的不雅语言攻击的是“中国”这个国家，而非“中国人”，应该恢复该帖子内容。

在有关南非的诋毁性措辞案件中，公众评议也帮助委员会了解了帖子中使用的诋毁性措辞在当地所代表的含义。这使我们得出结论，其中一个诋毁性措辞在南非是充满仇恨和伤害意味的字眼。这继而支持了我们拒绝在 Facebook 中恢复该帖子的决定。

提供更多的背景

在帮助我们判定巴西新冠疫情封城案件的过程中，公众评议着重指出了巴西新冠疫情防疫措施被政治化。这促使委员会提出了一项建议，即鉴于疫情，Meta 应考量当地情况，优先对公共机构分享的、疑似包含错误的卫生信息的内容进行事实核查。

表达用户之忧

公众评议也为人们提供了一个机会，让他们可以提出与案件有关的更广泛的问题。举例而言，有人在就分享半岛电视台帖子案件提交的评议中指控，Meta 在删除或降级处理展示巴勒斯坦用户的内容和阿拉伯语内容时未遵守“相称性”原则，与其对以色列境内威胁使用暴力反阿拉伯或反巴勒斯坦的帖子的处理方式形成鲜明对比。这些评议促使委员会就此案件提出建议，敦促 Meta 调查在对阿拉伯语和希伯来语的内容进行审核时是否存在偏见。

评估 Meta 的人权维护情况

委员会收到了来自联合国特别报告员以及世界各地其他人权专家的三份公众评议。在奥贾兰的单独监禁案件中，联合国人权和反恐问题特别报告员指出，Meta 的危险人物和组织社群准则和社群守则“不太符合国际法律，在实践中可能损害某些基本权利，包括但不限于表达自由、结社自由、参与公共事务以及无歧视的权利。”

经验教训

自我们 2020 年 12 月公布第一个案件以来，我们一直在倾听用户的意见，并不断改进公众评议流程。

作为对利益相关者反馈的回应，我们在案件总结中提供了更多的细节，并添加了“提示”问题，以着重指出特别欢迎他们发表意见的领域。为了帮助尽可能多的人和组织参与我们的工作，我们将最初 7 天的公众评议提交期限延长到 10 天，之后又延长到 14 天。我们还增加了页面限制，允许提交更详细的内容。虽然我们一直接受用户用英语以帖子的形式发表评议，但在 2022 年，我们还计划为用户提交评议提供更多语言支持选项。



建议



2021 年向 Meta 提出了 86 条建议。

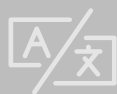
Meta 收到建议后，做出以下承诺：



在删除仇恨言论帖子时，为用户提供**更具体的信息**；



在某些地方提供新的提示信息，告知用户**自动审查或人工审查**是否会导致其内容被删除；



将其规则翻译成**4 亿多人使用的多种语言**；



采用新的**危机政策方案**，规范危机状况应对方式；



披露更多有关**政府要求**的报告；



本报告采用了一套**全新的数据驱动型方法**来跟踪落实情况，确保 Meta 落实对建议所做的承诺。



在我们提出的 86 项建议中，Meta 已证明落实或已汇报进度的建议达到 **2/3**，与提高**透明度**有关的建议多数都属于这种情况。



Meta 对建议的回应随着时间有所改善。

概述

在我们于 2021 年发布的 20 项决定中，我们就如何改进影响 Meta 数十亿用户的内容发布政策和程序向其提出了 86 项建议。

虽然我们的建议只是建议性质的，但 Meta 必须公开予以回应。2021 年，自我们发布决定之日起，Meta 有 30 天的时间对建议进行回应。这一期限已延长至 60 天。

我们的建议一再敦促 Meta 遵循一些有关透明度的核心原则：例如，将平台规定翻译成不同的用户语言，使其易于理解；尽可能清楚地公开平台是如何做出和执行决定的；以及如果有人违规，让他们知道是谁做错了。

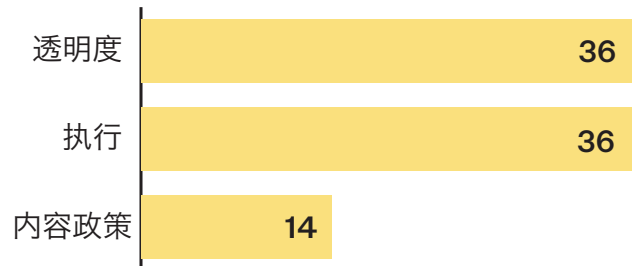
尽早兑现承诺，发挥越来越大的影响力

2021 年，我们见证了 Meta 承诺积极配合回应我们提出的建议，这一切努力很有可能切实改善世界各地 Facebook 和 Instagram 用户的使用体验。

公布规定，让用户知道他们的错处：

- 2021 年里，我们向 Meta 提出某一条建议的次数比其他任何建议都要多，有 6 项决定中反复提到了这条建议：那就是如果要删除用户的内容，应该告诉他们具体违反的是哪一条规定。Meta 积极回应了这条建议，所以，现在，每当 Facebook 英语用户发表的内容因违反仇恨言论政策而被删除，他们都会收到特定的通知。目前，Meta 正在测试这一特定通知功能，计划让它可以应用于阿拉伯语、西班牙语和葡萄牙语的内容，以及因违反欺凌和骚扰规定而被删除的帖子。
- 在我们发布的关于 Instagram 内容的三项决定中，我们敦促 Meta 告知用

委员会 2021 年提出的各种建议类型的数



户 Facebook 的社群守则同样适用于 Instagram。Meta 表示它“充分执行了”这项建议。

- 为了回应我们提出的建议，解释并举例说明危险人物和组织政策中的关键术语，Meta 补充了“赞扬”、“实质性支持”和“代表”等术语的定义和例子。这些说明对于用户了解什么是允许的，什么是禁止的至关重要。

向用户透明公开内容审核的依据：

- 为了回应我们在乳腺癌症状和裸露案件决定中的建议，Meta 在某些地方提供了新的消息提示，告知用户导致其内容被删除的是自动化还是人工审查。
- Meta 采纳了我们的另一项提议，同意在政府发布正式报告后，提供因违反社群守则而被删除的内容的信息。这将让政府如何对 Meta 施加压力以及 Meta 如何应对这种压力变得更加公开透明。
- 作为对另一项建议的回应，Meta 已经开始了必要的准备工作，计划在其按季度发布的社群守则执行报告中汇报允许保留具新闻价值内容的适用实例。Meta 有望在 2022 年年底之前落实这项建议。

公平对待用户，不管他们身在何处：

- Meta 遵循了另一项建议，将 Facebook 的社群守则翻译成了旁遮普语和乌尔都语。2022 年初，Meta 还承诺将 Facebook 的规定翻译成马拉地语、泰卢

固语、泰米尔语、古吉拉特语等语言。一旦落实，将会有超过 4 亿人可以用自己的母语读懂 Facebook 的社群守则，他们主要位于南亚地区。

- 在听到利益相关者声称 Meta 对巴勒斯坦用户帖子的删除违反了“相称性”后，我们敦促 Meta 调查他们在对阿拉伯语和希伯来语的内容进行审核时是否存在偏见。Meta 表示同意这样做。

改善执行规则的方式，一致地应用规则：

- 乳腺癌症状和裸露案件处理的是一篇可提高人们对该疾病的认识但却被错删的帖子，为了回应我们在该案件决定中的建议，Meta 更新了它的裸露内容自动删除模式，将卫生背景纳入考虑范围。
- 在我们的奥贾兰的单独监禁案件的决定中，Meta 首先删除了该内容，因为它声

称，该用户违反了有关“危险人物”的规定。然而，它随后披露，它发现了一项已经遗漏了三年的政策例外规定，而这项例外规定显然允许发布该帖子。该例外规定明确允许用户讨论被认定为危险组织成员的对象监禁条件。最终，为了落实我们针对此案提出的建议，Meta 表示其现已恢复该政策指导。这将意味着，在未来，用户将更清楚 Meta 在平台上审核危险人物相关内容时的审核标准。

- 为了回应我们在封禁前总统特朗普帐户案件决定中的建议，在经过与公司外部 50 多位全球专家的商议后，Meta 采用了新的危机政策方案，规范对危机状况的应对方式。这将为该公司应对危机状况的方式提供更加一致、透明的依据。

经验教训

虽然 Meta 在对我们最初建议的回应上存在问题，但 Meta 和委员会双方都已在 2021 年采取了改进建议流程的措施。尽管这项工作已经有所成效，但关键问题仍然存在。我们如何让我们的建议更有意义？我们如何在不影响独立性的前提下与 Meta 展开这方面的合作？至关重要的是，我们如何确保 Meta 以委员会可衡量以及全世界人民可感知的行动来兑现其承诺？针对这些问题，我们制定了一套更严格的新方法来监控 Meta 对我们建议的落实情况，本章稍后将对此进行说明。

Meta 对委员会建议的回应情况

下面的图表列出了 2021 年 Meta 对委员会提出的建议的回应情况：

Meta 承诺落实我们的大部分建议，在 Meta 共计 87 次回应中，有 55 次可以归入“充分执行”或“部分执行”类别。

对于 8 条建议，Meta 回应称它“正在评估可行性”，同时它还声称有 14 条建议反映的是它已经开展的工作。Meta 还表示有 10 条建议，它不会采取任何进一步措施来进行回应。这些数据反映了 Meta 于 2022 年 3 月发布的关于委员会的第四季度更新中列出的回应。由于对委员会最初提出建议的统计存在差异，Meta 给出的回应次数 (87) 与我们在此汇报的建议总条数 (86) 不匹配。

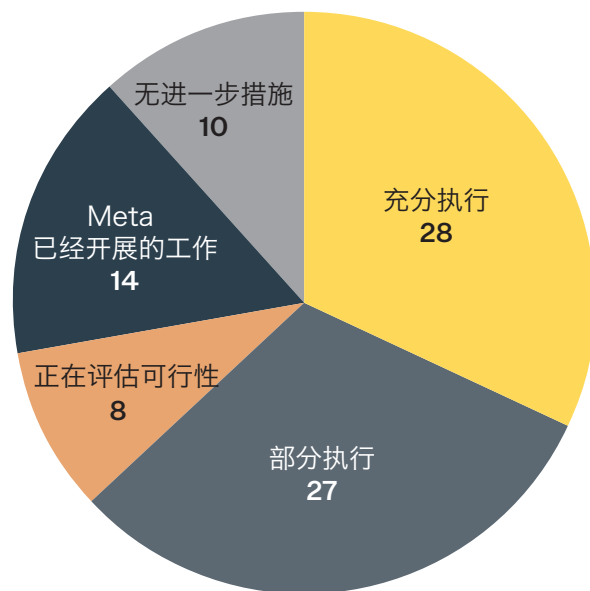
改进建议流程

在 2021 年，我们见证了建议流程的改进。委员会在提出建议方面有所进步，而 Meta 对建议的回应也有所进步。尽管 Meta 在对我们最初建议的回应上存在一些问题，但双方都迭代更新了他们的方法。在 2022 年，我们会更加了解建议流程的步骤以及如何改进该流程。

2021 年 2 月，Meta 回应了我们在最初提出的 5 项决定中的建议。虽然 Meta 承诺会采纳我们的大多数提议，但有时候，它也会修改或重订这些建议。其他时候，Meta 只回应了我们提议的某些部分，或者似乎没有完全理解我们试图表达的观点。

在委员会成立后第一年的工作中，我们还改变了流程，以加强建议的落实。

在扮成“黑彼得”案件的决定中，我们编撰了新的“政策咨询声明”部分，对我们的建



议进行了明确的编号，并声明“监督委员会要求 Facebook 对起草的每项建议提供单独的回复”。从 10 月起，我们也已开始使用季度透明度报告指出 Meta 歪曲或误解我们的建议的实例。

我们也认识到，在 2021 年，Meta 还采取了措施，来改进对我们建议的回应方式。

随着时间的推移，Meta 回应的类别更加明确，举例来说，从“承诺采取行动”变为“充分执行”或“部分执行”。Meta 还请 BSR（商务社会责任国际协会）研究它该如何与我们委员会互动来落实建议。

总体而言，与年初相比，年底，Meta 对我们的建议提供了更全面的回应。

对 2021 年底提出的建议作出回应的两个例子证实了这一进展。在对贝壳串珠带案件决定中的建议作出回应时，Meta 描述了它在与我们的提议相关的领域进行的两个实验。它表明了提出建议和回应建议的过程是如何增加透明度的，也揭示了 Meta 内部流程的新信息。在单独回应关于死藤水案件决定中的建议时，Meta 指出，它会在其发布了讨论记



关于 Meta 落实和回应我们建议的及时性的报告

- 根据 2021 年实施的细则，从委员会发布决定开始，Meta 必须在 30 天内对建议进行公开的回应。
- 2021 年，Meta 按时回应了我们的建议。

录的产品政策论坛上评估可行性。我们欢迎 Meta 提供更多的这类回应。

Meta 在提供影响的实证证据方面也对委员会更加透明公开。在最近的可行性评估中，Meta 了解到，通知用户哪些违反仇恨言论的子类型导致了他们的内容被删除，会让用户更好地理解这些规定是如何实施以及为什么实施的。Meta 透露，它发现，在多个市场上，人们对它的透明度和合法性的认知都有显著提高。鉴于此情况，Meta 现在正在尝试为用户提供更具体的欺凌和骚扰违规通知。

2021 年，Meta 多次就响应我们的建议所需的资源水平提出了问题。在 2021 年 7 月发布的第一份关于委员会的季度透明度报告中，Meta 表示：

“委员会建议的规模和范围超出了我们在设立委员会时最初预计的政策指导，落实其中一些建议需要投入数月或多年的时间。”

Meta 在 2021 年 11 月发布的下一份关于委员会的季度透明度报告中阐述了这些担忧：

“建议的节奏快且建议量大，这使我们没有足够的时间在一开始就评估和落实这些建议。平均而言，我们的团队能在任意一个时间点及时评估并回应 5 到 35 条建议。这些建议中的大多数需要十几个人来评估可行性，我们不可能在 30 天内轻松完成。我们需要将建议纳入现有的计划和优先事项，例如让建议

在我们的职业操守和相关产品团队的‘路线图’中得到体现，这使得这一困难进一步加剧了。”

2021 年，我们的建议推动 Meta 对用户的 Facebook 和 Instagram 体验方式做出了重大改变，其中可能需要进行巨大的技术变革。我们知道，Meta 在全面响应这些变革之前，首先需要时间来评估它们是否可行。为了解决这些问题，我们随后同意将 Meta 对建议作出回应的期限从 30 天延长到 60 天。

除了 Meta 需要专门详细政策指南的封禁前总统特朗普帐户和奥贾兰的单独监禁案件，在 2021 年公布的决定中，平均每个决定包含大约 3 条建议。因此，我们力求确保委员会的建议具有可执行性且真正得到落实，并与 Meta 就采取行动所需的资源进行了商议。

改进 Meta 回应建议的方式

本报告分析了 Meta 对我们委员会在 2021 年提出建议的回应情况。“回应”的评估结果评估的是 Meta 对委员会的建议是否作出了全面回应。这些数据反映的是 Meta 发布的截至 2021 年第三季度的委员会季度更新信息。

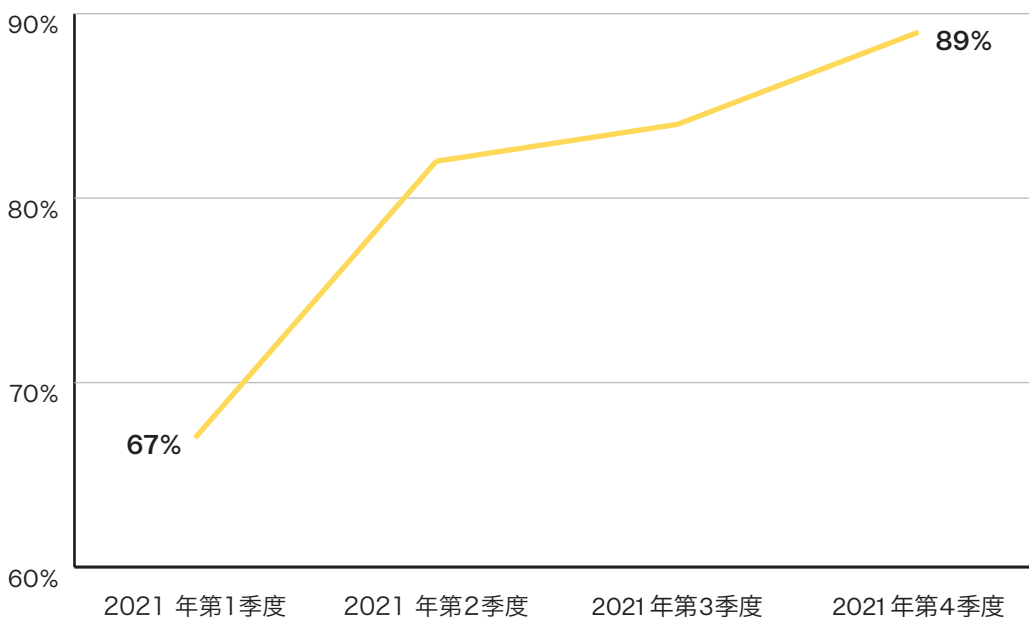
在对每一次回应进行评估时，我们通常会提三个问题：(1) 回应是否有涉及到建议的各个部分？(2) 回应是否有承诺采取行动？(3) 回应是否给出了时限？如果 Meta 满足其中一个条件，我们认为它提供了“比较全面”的

回应，而如果它满足至少两个条件时，则视为它提供了“全面”的回应。

当我们研究 Meta 对其他季度发布的建议的回应情况时，可以看出，被认为是“全面”或“比较全面”的回应的份额一直在随时间推移而增加，从 2021 年第一季度的 67% 上升到第四季度的 89%。

虽然仍有改进的空间，但这似乎表明，之所以 Meta 作出了更全面的回应，是源于 2021 年建议流程的变化。

Meta 对我们的建议回应“全面”或“比较全面”的频率情况



从承诺到行动：为用户造福



重要的是，要确保 Meta 做出的承诺不会被搁置或遗忘，其中一些承诺需要时间来落实。”

监督委员会行政管理总监
Thomas Hughes。



Meta 承诺落实我们在 2021 年提出的大多数建议，而我们的下一个任务是确保其恪守承诺。委员会的首要任务是确保 Meta 将承诺转化为行动，为世界各地的用户改善使用体验。

由于 2021 年建议数量增多，我们跟进落实的工作量也随之增加。本报告中列出了一套严格的数据驱动型方法，用于跟踪 Meta 如何落实我们提出的建议，以及更明确的评估这些建议对用户的影响。

培养责成 Meta 跟进落实的能力

从 2021 年年中起，我们开始强化对 Meta 建议落实情况的监督能力。7 月，我们成立了案件落实和监督小组，并配置了相关人员来支持委员会的工作。这个小组现已扩大规模，他们负责监督和评估 Meta 采取的回应措施和行动情况，以了解委员会建议对 Facebook 和 Instagram 用户的影响。

2021 年年中，我们还成立了一个由委员会成员和 Meta 高层员工组成的落实工作组。该工作组已多次会面以讨论改进建议流程，会上 Meta 回答了我们就提议落实流程提出的问题。该年年底，我们还成立了一个落实委员会，由五名委员会成员组成，与我们的案件甄选和会员委员会一起工作。这代表了一

个明确的选择，将落实与我们组织最关键的职能保持同步。落实委员会一直在努力优化委员会提出的建议，确保这些建议侧重于具体的、可衡量的影响。

用于追踪落实情况的全新数据驱动型方法

在帮助我们责成 Meta 跟进落实的过程中，这份报告也可以起到标志性的作用。在本报告的附件中，我们首次评估了 Meta 对我们在 2021 年提出的 86 项建议的落实情况。这些数据反映的是 Meta 发布的截至 2021 年第三季度的更新信息。

为了评估 Meta 的落实进度，我们调查了特定建议的落实是否达到了特定标准。具体情况因建议内容而异。例如，如果我们建议 Meta 在其面向公众的社群守则中添加额外的细节，或者就某个主题发布一份人权尽职调查报告，那么只要对外发布这些内容就能证明已落实这些建议。而对于其他一些建议，要想证明已落实，Meta 需要发布不能公开的信息。这意味着可能需要向用户发布通知，以证实它指出了用户违反的特定规定。接下来，我们会评估 Meta 对委员会建议的落实情况，将结果分为四个类别，按季度更新我们的评估报告：

- “已证明落实” – Meta 已经向我们提供了足够的证据证明已经落实了这个建议。
- “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 表示其已落实此建议，但未向我们提供足够的证据来验证这一点。
- “建议已遭到忽略、拒绝或修改” – Meta 将不会对我们的提议采取任何进一步的措施。

使用这种全新的数据驱动型方法意味着，我们对 Meta 是否落实了建议的评估结果有时可能不同于公司自己报告的结果。但是，我们相信，这种单独的验证方式对于责成 Meta 跟进落实以及确保让用户感受到建议落实的作用至关重要。下方图表便对 Meta 落实我们建议的情况进行了详细的分析评估。

这表明，Meta 证明了它在 2021 年落实了我们提出的其中 14 条建议。由于委员会发布建议的时间还不长，再加上我们在成立第一年提出的建议雄心勃勃、充满挑战，因此 Meta 需要时间来落实我们的建议，并证明他们已经在采取行动。总共有 42 项建议属于“已汇报进度”类别，其中包括本节导言中列出的许多有重要影响力的建议。我们预计，在 2022 年，之前属于这一类别的许多建议会归入“已证明落实”类别。

对于其中 21 条建议，Meta 报告称已落实我们的提议或表明已完成相关工作，但未向我们公布足够的信息来验证这一点。接下来，我们会鼓励 Meta 分享更多数据，并提供更多证据来支持其已落实我们建议的这一说法。我们也在敦促 Meta 进一步放开公司数

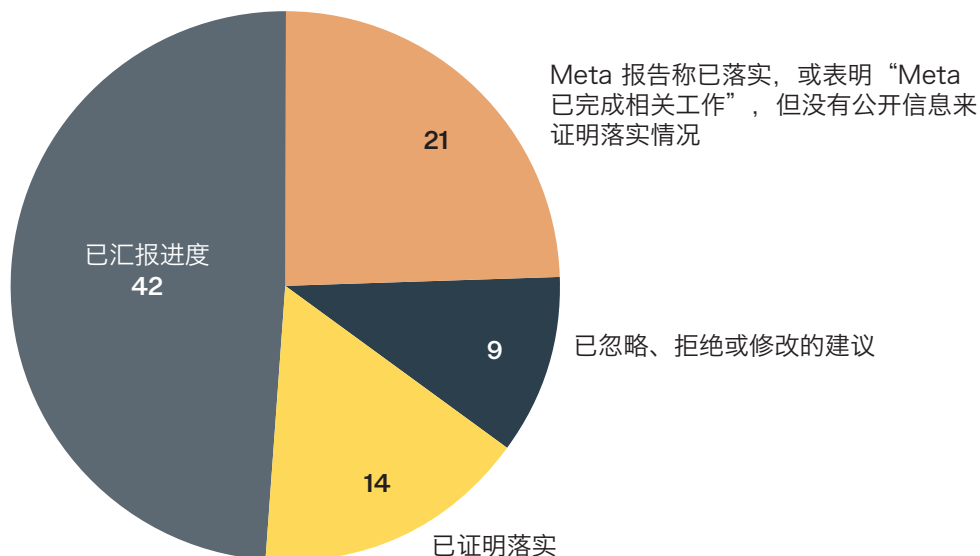
据的访问权限，以支持进度的验证工作。有 9 条建议属于“建议已遭到忽略、拒绝或修改”类别，这表示 Meta 将不会对这些提议采取进一步的措施。

如果我们分析被评估为“已证明落实”或“已汇报进度”的各种建议类型，就会发现与透明度和内容政策有关的建议中分别有 79% 和 67% 属于这种类别，而与执行有关的建议则只有 58%。与执行有关的建议比例较低可能反映了这样一个事实，即这些提议往往要求 Meta 对 Facebook 和 Instagram 做出更广泛的改变。

为 Facebook 和 Instagram 用户造福

虽然验证 Meta 对委员会建议的落实情况是一项非常复杂的任务，但我们选择优先处理这项任务，并在这方面投入了大量的资源。这样做的原因很简单：只有帮助 Meta 将承诺转化为行动，我们才能从长远的角度为 Facebook 和 Instagram 用户改善使用体验，并履行它尊重人权的责任。

Meta 对委员会建议的落实情况



责成 Meta 回应针对交叉检查提出的问题

2021 年 9 月 13 日，《华尔街日报》报道称，Meta 交叉检查计划的使用程度和范围远远超出公司之前所宣称的那样。一系列文章的内容基于从举报人处获得的公司文件，第一篇文章指出：

马克·扎克伯格曾公开表示，[Meta] 允许其 30 多亿用户与政治、文化和新闻精英平等对话，其行为标准适用于每个人，无论其地位或名望如何。[但]在私下里，该公司建立了一个系统，可以让知名用户免受其部分或全部规定的约束。

这篇文章称，交叉检查已经从一个“最初旨在作为针对名人、政客和记者等知名帐户采取的行动进行质量控制的措施”演变成一个保护“数百万贵宾用户”不受公司正常执行程序约束的计划。根据举报人的文件，一些知名的博主被“列入了白名单”。这篇文章称，这样做的结果要么是使 Meta 员工能够将某些实体（如个人资料、页面）添加到列表中，这些列表据称可以无限期地保留“违规内容”，要么是“往往不需要等待 Facebook 员工审核。”

2021 年 9 月 21 日，我们发布了对这些爆料的回应：

一段时间以来，我们监督委员会一直在询问关于交叉检查的问题。在我们就美国前总统唐纳德·特朗普的帐户做出的决定中，我们警告称，缺乏关于交叉检查的明确公开信息，以及 [Meta] 的“允许保留具新闻价值内容的例外

规定”，可能会让人认为 [Meta] 受到了政治和商业因素的过度影响。

我们在该决定中向 Meta 提交了建议，其中一个建议敦促它“报告与正常执行程序相比，通过交叉检查程序做出的决定的相对错误率和主题一致性。” Meta 回应称拒绝落实这一提议，理由是“无法追踪这一信息。”它还提供了 2018 年发布的一篇博客帖子的链接来支持这一说法：“无论谁发布的内容违反了我们的守则，我们都会将其从 Facebook 上删除。”然而，《华尔街日报》报道称，Meta 在 2019 年的内部审核中发现，这篇博客帖子“具有误导性”。

委员会总结道：“鉴于最近的事态发展，我们正在调查 [Meta] 在交叉检查方面的回应中是否完全坦诚，包括白名单做法。委员会已与 [Meta] 联系，以进一步阐明之前与我们分享的信息。我们预计将在未来几天内收到简报。”

在该简报中，Meta 团队承认，在回应我们在与美国前总统特朗普相关的案件中提出的建议时，“不应该说交叉检查只适用于‘少数决定’”。2021 年 9 月 28 日，也就是我们 9 月 21 日发布博文的一周后，Meta 以征求政策咨询意见的形式对委员会的问题做出回应。Meta 要求我们在意见中列入关于如何改进交叉检查的具体建议。在宣布这一提及时，Meta 指出：

最近，委员会表示有兴趣更深入地了解我们的交叉检查系统。该提请超出了我

7 “Facebook 表示其规则适用于所有人。公司文件揭示了一个被豁免的秘密精英”，Jeff Horwitz，《华尔街日报》，2021 年 9 月 13 日

8 同上。

9 同上。

们已经提供的简报的范畴。我们正在通过正式且透明的流程主动征求委员会的意见……具体来说，我们将要求委员会就我们用于通过交叉检查确定二次审查优先事项的标准以及我们如何管理该计划提供指导。

10月初，在发布委员会的第一份季度透明度报告之前，我们宣布将与之前受雇于 Meta 的举报人 Frances Haugen 会面，讨论她披露的关于 Meta 内容审核程序和政策的重大问题。

在过去几周里，由于 [Meta] 前员工 Frances Haugen 的行为，关于 [Meta] 内容审核方法的新信息被曝光。鉴于 Haugen 女士对 [Meta] 提出的严重指控，我们已邀请她在未来几周内向委员会发表讲话，她已接受。委员会成员很高兴有机会讨论 Haugen 女士的经历，并收集执行信息，以帮助推动 [Meta] 通过我们的案件审理决定和建议提高透明度和责任感。

我们的博客总结道：

[Meta] 之类公司所做的选择对世界各地数十亿人的言论自由和人权有着现实影响。在这种情况下，规则的透明度至关重要。

作为委员会，我们将继续向 Facebook 提出尖锐的问题，并推动该公司致力于提高透明度、责任感和公平性。归根结底，只有这样才能让用户相信自己将受到公平对待。

在我们当月早些时候发布的第一份季度透明度报告中，我们指出，在我们看来，Meta 负责提供信息的内部团队并没有对交叉检查“完全坦诚”。“在某些情况下，”我们阐述道，“[该团队] 未能向委员会提供相关信

息……[而且] 在其他情况下，它所提供的信息并不完整。”有鉴于此，我们宣布委员会未来与 Meta 的工作关系将取决于我们是否能够“相信 Meta 提供的信息准确且全面，并描绘了当前主题的全貌”。为了实现这一目标，我们于 2021 年 10 月 21 日宣布，我们已接受 Meta 的请求，将制定政策咨询意见，审查其交叉检查计划并就如何改进提出建议。

10 “要求监督委员会就我们的交叉检查系统提供指导”，Meta 博文，作者：全球事务副总裁 Nick Clegg，2021 年 9 月 28 日（2021 年 10 月 21 日更新）

11 “监督委员会将与 Frances Haugen 会面”，监督委员会博文，2021 年 10 月

交流和宣传

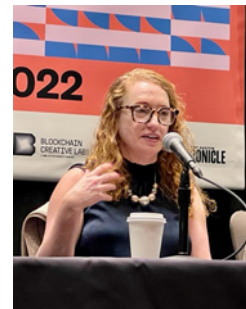
在第一年，我们与世界各地的个人和组织交流，解释委员会的工作，鼓励他们对我们的决定提出意见，并讨论为什么有原则的全球性内容审核方法对用户很重要。



Julie Owono



Nighat Dad



Suzanne Nossel

鼓励人们分享自己的观点

我们的大部分外部互动都鼓励个人和组织参与我们的工作，并为正在处理的案件提交公众评议。我们举办了 19 场“办公时间”会议来介绍如何提交公众评议，有 500 多名利益相关者参加了这些会议。为了鼓励世界各地人员参与，这些会议在多个时区举行，并提供阿拉伯语、希伯来语、葡萄牙语和西班牙语等语言的口译服务。在宣布封禁前总统特朗普帐户案件后的两周内，我们还举办了 12 场活动和圆桌会议，以鼓励公众就该案件发表评议。有 1,000 多人参加了这些活动。

“

在全球北方提出的建议
会损害全球南方的社
群。”

委员会成员 Nighat Dad



推动关于内容审核和人权的讨论

我们将人权应用于内容审核的方法不仅由委员会成员之间的内部审议决定，还受到学术界、新闻界、民间团体和国际机构代表的公开讨论的影响。

6 月，我们在 RightsCon 举办了多场活动，包括一场关于让内容审核与人权相一致的会议。9 月，在联合国大会期间，委员会成员 Julie Owono 加入了两个旨在打击虚假信息 and 仇恨言论以及提高透明度的联合国小组。12 月，委员会成员和前联合国特别报告员 Maina Kiai 在一次活动中发表讲话，讨论了委员会在人权保护和数字平台治理方面获得的经验教训。Maina Kiai 与联合国人权事务负责人 Peggy Hicks 一起发言时说：“在监督委员会，我们明白我们只是在解决一部分问题，许多其他领域的专家也必须采取行动。”我们还在 12 月的互联网治理论坛上举办了一场活动。在该活动中，委员会主任 Thomas Hughes、委员会成员 Afia Asantewaa Asare-Kyei 和委员会受托人 Cherine Chalaby 讨论了有关数字权利和自我监管的问题。

在全球展开关于内容审核的讨论

社交媒体影响着世界各地的事件和数十亿人。然而，围绕内容审核的讨论往往只考虑欧洲和北美的用户。为了让所有人都能参与这场讨论，我们优先安排了一些活动，研究全球南方国家面临的内容审核挑战。

5月，委员会成员 Catalina Botero-Marino 和 Jamal Greene 参加了“美洲对话”活动，讨论了委员会的决定对全球南方，特别是拉丁美洲的意义。8月，委员会成员 Endy Bayuni 参加了亚洲论坛举办的以人为

本的网络安全活动。9月，委员会成员 Julie Owono 在塞内加尔个人数据保护委员会参加的活动中，讨论了西非和中非在内容审核方面的经验教训。

最后，在11月，委员会成员 Nighat Dad、Ronaldo Lemos 和 Maina Kiai 在巴黎和平论坛上发表讲话，Nighat Dad 指出“我们在全球南方看到的一些规则是从全球北方复制粘贴过来的，没有考虑我们当地的背景和现实。在全球北方提出的建议会损害全球南方的社群。”



展望未来：2022 年及以后

2021 年，监督委员会向 Meta 提交了 20 项决定和 86 项建议。综合而言，它们表明我们致力于让公司承担责任并引导其提高透明度。

随着我们进入 2022 年，除了增加新的委员会成员之外，我们还希望深化我们的影响，改善 Meta 对待世界各地的用户和社群的方式。

拓宽我们的业务范围

我们正在与 Meta 就扩大委员会的业务范围展开对话，包括在群组和帐号等方面审核用户对公司决定提出的申诉，而不再局限于个别帖子，并希望在明年报告这方面的进展。我们也在审视自己在评估 Meta 针对新兴“

元宇宙”制定的内容审核计划中扮演的角色。2021 年，该公司在“元宇宙”方面花费了 100 亿美元，表明这将是未来几年的一个主要关注领域。

发挥全球影响力

由于 Meta 的影响遍及全球，委员会的工作必须具有全球性。2021 年，超过 2/3 的用户申诉来自欧洲、美国和加拿大，其中美国和加拿大占 49%，欧洲占 20%。相比之下，仅 14% 的申诉来自拉丁美洲和加勒比地区；9% 来自亚太和大洋洲地区；4% 来自中东和北非地区；2% 来自中亚和南亚地区；以及 2% 来自撒哈拉以南非洲地区。这一分布情况反映了我们正在努力解决的一种不平衡现象。

2022 年，我们将扩大宣传和利益相关者的参与，与亚洲、拉丁美洲、中东和非洲的政策制定者和民间团体领袖合作。我们将建立一个全球性的区域顾问网络，鼓励人们在各自的区域内提出申诉和提交公众评议。我们还计划接受更多语言的公众评议，而不仅仅是英语和所涉及内容的语言。

发布政策咨询意见

除了审核各个案件以删除或恢复内容，委员会还可以接受 Meta 的政策咨询意见请求。这让我们能够审核公司在特定领域的政策，并就如何改进这些政策提出建议。2021 年，我们接受了 Meta 的两项关于共享私人

住宅信息和公司交叉检查系统的请求。我们关于共享私人住宅信息的意见已于 2022 年 2 月发布，而关于交叉检查的意见将于今年晚些时候发布。

获取对数据的更大访问权限

2022 年，作为我们确保 Meta 信守其承诺工作的一部分，委员会正在寻求访问 Facebook 数据和研究。这将更好地说明委员会甄别的案件、我们提出的建议以及我

们产生的影响。在努力了解和提高我们的影响力的过程中，我们还计划聘请数据科学家。

分享我们迄今为止所积累的经验

在第一年，我们通过应用基于国际人权标准的框架，制定、完善并落实了一个全新的内容审核方法。在此过程中，我们开发出一套最佳实践，并积累了丰富的经验。

2022 年，我们期待着分享我们在共同努力改进社交媒体的过程中所获得的经验。在不

组建一个面向长远成功的机构

最后，联合主席和委员会成员、受托人以及行政管理机构将继续按照多年的战略规划，建立一个高绩效且可持续发展的机构。这将包括委员会成员识别、选择和入职以及

断变化的内容审核和监管格局中，我们还将进一步发挥作为自我监管最佳实践来源的作用。我们将在技术政策对话中分享我们关于自我监管解决方案的专业知识，并与潜在的技术行业合作伙伴就内容审核创新展开讨论。

绩效和行为评估的自主流程。我们也会继续确保对委托资源进行有效和负责任的管理，包括制定长期融资模式。

总结

我们第一份年度报告中的见解表明，监督委员会已经对 Meta 的工作方式和服务用户的方式产生了切实的影响。我们做出的每一个决定、获得的经验以及得出的结论，都代表着漫长旅程中的第一步。

“内容审核”这一平淡无奇的术语掩盖了这样一个事实：在社交媒体上发布和分享的内容会对人们的线下和线上生活产生重大影响，无论是积极还是消极。虽然社交媒体在很大程度上实现了其早期承诺，将数十亿人聚集在一起，但它也为人们创造了伤害他人的新方式。

除了任何具体内容决策的结果，用户将越来越多地感受到委员会的影响，他们将从 Meta 明确、公平且透明地应用自己的政策中受益。我们希望委员会的工作为世界各地的其他社交媒体公司和平台出版商、监管机构和政策制定者提供一个可靠的框架，以应对内容审核固有的紧张局势和挑战。

在监督委员会会议上，我们提交了这份首次年度报告，准确地说明了这项工作。我们为迄今为止取得的成就感到自豪，并对未来的大量工作有清醒的认识。这份文件首先反映了我们对推进透明度、问责制和责任感等价值观的集体承诺，我们希望这将使社交媒体平台成为用户消磨时间的更好场所。我们将继续让 Meta 承担起责任，改善对待世界各地用户和社群的方式。

附件

Meta 对我们的建议的回应和落实情况¹²

建议标识	监督委员会建议	类别	委员会对 Meta 回应的评估	委员会对落实情况的评估
亚美尼亚人侮辱阿塞拜疆				
1	确保始终告知用户对其执行任何社群守则的原因，包括 Facebook 正在执行的具体规则。这样做将使 Facebook 能够鼓励符合其社群守则的表达，而不是对用户采取敌对姿态。在该案件中，用户被告知帖子违反了关于仇恨言论的社群守则，但未被告知帖子因包含针对国籍的诋毁性言论而违反守则。在这种情况下，Facebook 满足了合法性原则，但由于 Facebook 缺乏透明度，其决定很容易被误认为是因为用户触及有争议的主题或表达了 Facebook 不认同的观点而被删除帖子。	执行	●	
乳腺癌症状和裸露				
1	改进对有文字覆盖的图片的自动检测，以确保旨在提高对乳腺癌症状的认识的帖子不会被错误地标记为需要审核。	执行	●	
2	确保始终告知用户对其执行内容政策的原因，并提供 Facebook 所依据的社群守则中的具体规则。	执行	✘	
3	在使用自动化方式对用户内容采取执行措施时，需告知用户，并提供易于理解的关于其含义的说明。	执行	●	
4	确保用户在其内容被发现违反 Facebook 关于成人裸露和性行为的社群守则时，可以对自动系统做出的决定提出申诉以待人工审查。在试图防止儿童性剥削或传播未经同意的亲密图像时，Facebook 应该根据其关于成人性剥削以及儿童性剥削、虐待和裸体的社群守则来执行，而不是依靠过度执行关于成人裸体的政策。在这些情况下，仍然应该可以提出申诉，以便在误删允许的自愿成人裸体图片的情况下撤销该操作。	执行	✘	
5	落实内部审核程序，以持续分析具有统计代表性的自动内容删除决策样本，以撤销执行错误并从中吸取教训。	执行	✘	
6	扩展透明度报告，披露按照社群守则做出的自动删除决定的数量，以及在人工审查后这些决定被撤销的比例。	透明度	●	
7	修改 Instagram 社群准则的“简短”解释，以阐明禁止成人裸体并不是绝对的。	内容政策	✘	
8	修改 Instagram 社群准则的“长篇”解释，以阐明可以展示可见的女性乳头以提高对乳腺癌的认识。	内容政策	●	
9	阐明 Instagram 社群准则的解读与 Facebook 社群守则一致，如果存在不一致之处，以后者为准。	内容政策	●	
纳粹引语				
1	确保始终告知用户对其执行任何社群守则的原因，包括 Facebook 正在执行的具体规则（例如，支持仇恨组织）。	执行	●	
2	解释并举例说明“危险人物和组织”政策中使用的关键术语的应用，包括“赞扬”、“支持”和“代表”的含义。这些应与 Facebook 的内部实施标准中使用的定义相一致。社群守则应为用户提供更明确的指导，说明在讨论被指定为危险实体的人物或组织时如何使他们的意图显而易见。	内容政策	●	
3	提供一份根据危险人物和组织社群守则认定为“危险实体”的组织和人物的公示名单。至少应提供说明性的例子。这将有助于用户更好地理解政策并据此行事。	内容政策	●	
宣称新冠肺炎疗法				
1	委员会建议 Facebook 针对健康错误信息制定明确且易于理解的社群守则，在一个位置整合和澄清现有规则（包括定义“错误信息”等关键术语）。这一规则的制定应附有“详细的假设，说明 [这些] 规则的解读和应用中的细微差别”，以便为用户提供进一步阐述（请参阅报告 A/HRC/38/35，第 46 段（2018 年））。在修改规则的过程中，Facebook 应与相关利益相关者进行人权影响评估（UNGP，第 18-19 条原则）。	内容政策	●	

委员会对 Meta 回应的评估 ● 全面, ● 比较全面, ✘ 不全面

委员会对落实情况的评估

- 通过公布的信息证明落实情况
- 已汇报进度

- 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
- 已忽略、拒绝或修改的建议

建议标识	监督委员会建议	类别	委员会对 Meta 回应的评估	委员会对落实情况评估
宣称新冠肺炎疗法				
2	为确保针对健康错误信息的执行措施是对保护公众健康的侵扰最少的方法，委员会建议 Facebook：阐明寻求预防的特定伤害，并就如何评估特定内容的潜在伤害提供透明度；对其处理健康错误信息的一系列现有工具进行评估；考虑进一步开发比删除内容侵扰更小的工具的可能性；公布社群守则中的一系列执行方案，并根据这些方案对言论自由的侵扰程度，从高到低进行排序；阐明平台在执行社群守则以保护公众健康时，将使用哪些因素（包括基于证据的标准）来选择侵扰性最小的方案；在社群守则中明确规定适用于每条规则的执行方案。	执行	●	
3	如果用户发布的关于新冠肺炎治疗的信息与卫生部门的具体建议相矛盾，并且发现有可能造成身体伤害但并非迫在眉睫，委员会强烈建议 Facebook 采取一系列侵扰程度较低的措施。这可以包括添加标签，以提醒用户注意帖子内容的争议性，并提供指向世界卫生组织和国家卫生部门的观点的链接。在某些情况下，可能需要在帖子中引入额外的限制（例如，通过阻止互动或分享），以减少有机和算法驱动的放大效应。也可以考虑降低内容的排名，以防止在其他用户的新闻源中出现。所有执行措施，包括添加标签或其他引入限制的方法，都应明确告知用户，并支持提出申诉。	执行	✘	
4	发布一份关于在新冠疫情全球健康危机期间如何执行社群守则的透明度报告。该报告应包含：关于删帖次数的绝对值和百分比的数据，以及关于其他执行措施和所依据的具体社群守则的数据，包括完全依赖自动化方式的占比；按所执行内容类型进行的细分（包括个人帖子、帐号和群组）；按检测来源所做的细分（包括自动化、用户标记、可信的合作伙伴、执法机构）；按区域和语言所做的细分；关于侵扰性较低的措施有效性的指标（例如，添加标签或降低排名的影响）；关于整个危机期间申诉可用性的数据，包括申诉被完全撤回的案件总数，以及自动申诉所占的百分比；结论和经验教训，包括关于 Facebook 为确保进一步遵守其人权责任而做出的任何改变的信息。	透明度	✘	
印度抗议法国				
1	为确保用户清楚了解允许的内容，委员会建议 Facebook 向用户提供有关本社群守则的范围和执行情况的更多信息。执行标准应该公开，并与 Facebook 的内部执行标准保持一致。具体而言，Facebook 的标准应涉及意图、用户和受众的身份以及背景。	内容政策	●	
扮成“黑彼得”				
1	Facebook 应将仇恨言论社群守则中禁止扮演黑人的规则与公司制定该规则的理由联系起来，包括试图防止的伤害。	内容政策	●	
2	根据委员会在 2020-003-FB-UA 案件中的建议，Facebook 应“确保始终告知用户对其执行任何社群守则的原因，包括 Facebook 正在执行的具体规则”。在这种情况下，向用户发出的任何通知都应指明关于扮演黑人的规则，并链接到上述资源，解释这一规则所要防止的伤害。Facebook 应提供一份详细的最新资料，说明其对委员会先前关于这一主题的建议的“可行性评估”，包括任何技术限制的具体性质以及如何克服这些限制。	执行	◐	
旁遮普语用户提出对印度 RSS 的担忧				
1	Facebook 应该将其社群守则和内部实施标准翻译成旁遮普语。Facebook 应该致力于使其社群守则能够以其用户广泛使用的语言进行访问。这将使用户充分了解在使用 Facebook 的产品时必须遵守的规则。它还将使用户更容易就可能侵犯其权利的内容与 Facebook 互动。	内容政策	●	
2	根据委员会在 2020-004-IG-UA 案件中的建议，公司应尽快将人工审查和人工申诉流程恢复到疫情发生前的水平，同时充分保护 Facebook 员工和承包商的健康。	执行	●	
3	Facebook 应改进透明度报告，以增加有关错误率的公开信息，并支持针对每个社群守则按国家/地区和语言查看这些信息。委员会强调，更详细的透明度报告将帮助公众发现比较常见的错误领域，包括对少数群体的潜在具体影响，并提醒 Facebook 纠正这些错误。	透明度	◐	
封禁前总统特朗普帐号				
1	委员会认为，明确区分政治领导人和其他有影响力的用户并不总是有益的。重要的是要认识到，拥有大量受众的其他用户也可能造成严重的伤害风险。同样的规则应适用于平台的所有用户；但在评估因果关系以及伤害的可能性和紧迫性问题时，背景情况很重要。重要的是一个用户对其他用户的影响程度 [...] 对于有影响力的用户发布的帖子，Facebook 必须根据背景情况和帖子可能被理解的方式进行评估，即使他们的煽动性信息是用旨在避免责任的语言来表达的，例如表面上鼓励和平或合法的行动。Facebook 在本案中使用了拉巴特行动计划中的六个背景因素，委员会认为这是评估潜在有害言论的背景风险的有用方法。委员会强调，在这种情况下，时间至关重要；在有影响力的用户造成重大伤害之前采取行动应优先于新闻价值和其他政治传播价值。	内容政策	◐	
2	当有影响力的用户发布的帖子很有可能造成迫在眉睫的伤害时（根据国际人权标准进行评估），Facebook 应迅速采取行动执行其规则。	执行	◐	
3	虽然所有用户都应该遵守相同的内容政策，但在评估政治领导人的言论时，必须考虑一些独特的因素。国家元首和其他高级政府官员可能比其他人更有能力造成伤害。Facebook 应该认识到，国家元首和其他政府高级官员的帖子可能会增加鼓励、合法化或煽动暴力的风险，要么是因为他们的显赫地位使他们的言辞具有更大的力量和可信度，要么是因为他们的追随者可能会推断他们可以逍遥法外。同时，保护人们听取政治言论的权利也很重要。尽管如此，如果国家元首或政府高官多次发布可能造成伤害的信息（根据国际人权准则），那么 Facebook 应该封禁该帐号，并确定一个足以防止迫在眉睫的伤害的期限。	内容政策	●	

委员会对 Meta 回应的评估 ● 全面，◐ 比较全面，✘ 不全面

委员会对落实情况的评估

- 通过公布的信息证明落实情况
- 已汇报进度

- 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
- 已忽略、拒绝或修改的建议

建议标识	监督委员会建议	类别	委员会对 Meta 回应的评估	委员会对落实情况评估
封禁前总统特朗普帐户				
4	封禁时间应该足够长，以阻止不当行为；在适当的情况下，还可以删除帐号或页面。	内容政策	●	
5	对言论的限制通常由有影响力的国家行为者或在其要求下针对不同的政见和政治反对派成员实施。Facebook 必须抵制政府为让其政治反对派噤声而施加的压力。在评估潜在风险时，Facebook 应特别谨慎地考虑相关的政治背景。	执行	●	
6	在评估极具影响力的用户的政治言论时，Facebook 应迅速将内容审核流程升级到熟悉语言和政治背景、不受政治和经济干扰以及不当影响的专业人员。该分析应检查极具影响力的用户在 Facebook 和 Instagram 平台以外的行为，以充分评估潜在有害言论的完整相关背景。此外，Facebook 应确保投入足够的资源和专业知识来全局评估有影响力的帐号的伤害风险。	执行	✘	
7	Facebook 应该公开说明在对有影响力的用户实施帐号级别制裁时使用的规则。这些规则应确保，当 Facebook 对有影响力的用户的帐号实施限时封禁以降低重大伤害的风险时，它将在封禁期限到期之前评估风险是否已经消退。如果 Facebook 确定用户当时存在煽动迫在眉睫的暴力、歧视或其他违法行为的严重风险，并且有必要实施与风险相称的封禁措施以保护公共安全，则再次实施有时限的封禁。	内容政策	●	
8	当 Facebook 实施适用于有影响力的用户的特殊程序时，这些程序应该有据可查。目前尚不清楚 Facebook 在本案中是否应用了不同的标准，而且委员会听到了许多关于允许保留新闻价值内容的潜在应用的担忧。Facebook 必须解决这种缺乏透明度及其造成的混乱的问题。	透明度	●	
9	Facebook 应该提供更多来帮助用户理解和评估应用“允许保留新闻价值内容”的过程和标准。Facebook 应该清楚说明“允许保留新闻价值内容”如何应用于有影响力的帐号，包括政治领导人和其他公众人物。	内容政策	●	
10	对于交叉检查审核，Facebook 应明确说明审核的理由、标准和流程，包括用于确定入选页面和帐号的标准。	内容政策	●	
11	Facebook 应报告与普通执行程序相比，通过交叉检查流程做出的决定的相对错误率和主题一致性。	透明度	✘	
12	当 Facebook 平台被有影响力的用户滥用且对人权造成严重不利影响时，Facebook 应该对事件进行彻底调查。Facebook 应评估自身所具有的影响，并评估可以做出哪些改变来识别、预防、减轻未来的不利影响并担责。	执行	✘	
13	Facebook 应该全面审查其对选举舞弊说法，以及加剧的紧张局势最终导致 2021 年 1 月 6 日在美国发生暴力事件的潜在影响。这应该是对 Meta 做出的可能使其平台被滥用的设计和政策选择的公开反思。Facebook 应开展尽职调查，实施一项计划以根据其调查结果采取行动，并就其如何应对所涉及的不利人权影响进行公开交流。	透明度	✘	
14	在 Facebook 或 Instagram 用户可能参与暴力犯罪或严重侵犯人权，以及 ICCPR 第 20 条所述的煽动暴力行为的情况下，删除内容和停用帐号虽然可能会降低伤害风险，但也可能破坏问责工作，包括删除证据。Facebook 有责任收集、保存并在适当情况下共享信息，以协助主管当局和问责机制调查，甚至可能起诉严重违反国际刑事、人权和人道主义法的行为。Facebook 的企业人权政策应明确公司在这方面制定的协议。该政策还应明确如何将以前在平台上公开的信息提供给进行符合国际标准和适用数据保护法的调查的研究人员。	透明度	●	
15	Facebook 应在其社群守则和社群准则中以明确、全面且易于理解的方式解释其针对限制 Facebook 和 Instagram 上的个人资料、页面、群组和帐号的打击和处罚流程。这些政策应向用户提供足够的信息，以了解何时实施打击（包括任何适用的例外或允许情况）以及如何计算罚金。	内容政策	●	
16	Facebook 还应向用户提供获得违规、打击和惩罚评估结果的次数信息，以及未来违规的后果。	执行	●	
17	在透明度报告中，Facebook 应提供个人资料、页面和帐号限制的次数，包括采取执行行动的原因和方式，并按地区和国家划分信息。	透明度	✘	
18	Facebook 应制定并发布一项政策，以在常规流程无法防止或避免迫在眉睫的伤害的情况下，管理其对危机或新情况的回应。虽然这些情况并非总能预料到，但 Facebook 的指南应为此类行动设定适当的参数，包括要求在固定时间内审核其决定。	内容政策	●	
“双按钮”表情包				
1	做好技术方面的安排，确保向用户发出的通知参考公司执行的社群守则。如果 Facebook 认为 (i) 该内容没有违反用户的社群守则，但 (ii) 该内容违反了其他社群守则，则应适当通知用户，并给予用户再次申诉的机会。在联系委员会之前，他们应该始终能够访问正确的信息。	执行	●	
2	在仇恨言论社群守则的公共语言中包含当前未传达给用户的讽刺作品例外情况。	内容政策	●	
3	确保有适当的程序来正确评估讽刺内容和相关背景。这包括为内容审核员提供以下支持：(i) 联系 Facebook 的当地运营团队，以收集相关的文化和背景信息；以及 (ii) 有足够的时间咨询 Facebook 的当地运营团队并进行评估。Facebook 应确保其针对内容审核员的政策能够鼓励内容审核员在不确定表情包是否具有讽刺意味的情况下进一步调查或升级。	执行	●	

委员会对 Meta 回应的评估 ● 全面, ● 比较全面, ✘ 不全面

委员会对落实情况的评估

- 通过公布的信息证明落实情况
- 已汇报进度

- 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
- 已忽略、拒绝或修改的建议

建议标识	监督委员会建议	类别	委员会对 Meta 回应的评估	委员会对落实情况的评估
“双按钮”表情包				
4	让用户在其申诉中表明他们的内容属于仇恨言论政策的例外情况之一。这包括讽刺内容的例外情况以及用户分享仇恨内容以谴责或提高认识的情况。	执行	●	
5	确保基于政策例外的申诉优先进行人工审查。	执行	●	
支持纳瓦利内抗议活动				
1	Facebook 应修改和重新起草欺凌和骚扰社群守则，以解释政策理由和“禁止事项”之间的关系以及其他限制其内容的规则。	内容政策	●	
2	区分欺凌和骚扰，并提供区分这两种行为的定义。此外，社群守则应向用户清楚说明欺凌和骚扰与只会引起冒犯并可能受到国际人权法保护的言论有何不同。	内容政策	⦿	
3	明确定义针对不同目标用户类别的方法，并提供每个目标类别的说明性示例（即谁有资格成为公众人物）。按政策中当前列出的用户类别制定关于欺凌和骚扰的社群守则。	内容政策	●	
4	在欺凌和骚扰社群守则中包含违规和非违规内容的说明性示例，以阐明所划定的政策界线，以及这些区分如何基于目标的身份地位。	内容政策	●	
5	在评估包含针对无公职成年人的“负面性格声明”的内容时，Facebook 应修改社群守则，要求评估内容的社会和政治背景。Facebook 应该重新考虑在政治或公共讨论中执行此规则，因为删除内容会扼杀讨论。	内容政策	✘	
6	每当 Facebook 因负面性格声明而删除内容，但仅涉及较大帖子中的一个单词或短语时，应及时通知用户这一事实，以使用户可以重新发布没有负面性格声明的材料。	执行	●	
奥贾兰单独监禁				
1	立即将放错位置的 2017 年指南恢复到内部实施标准和已知问题（适用于内容审核员的内部指南），通知所有内容审核员该指南的存在并立即安排相关培训。	执行	●	
2	评估执行危险人物和组织政策的自动审核流程。如有必要，Facebook 应更新分类器，以排除先前因未能应用 2017 年指南而导致的执行错误中的培训数据。应增加新的培训数据，以反映本指南的恢复情况。	执行	✘	
3	发布正在进行的审查过程的结果，以确定是否丢失了任何其他策略，包括对所有丢失策略的描述、策略丢失的时间以及为恢复它们而采取的步骤。	透明度	⦿	
4	在危险人物和组织的“政策理由”中表明，尊重人权和言论自由，特别是公开讨论与恐怖主义和反恐工作相关的侵犯和践踏人权行为，可以推进“安全”的价值，而且平台必须为这些讨论提供一个空间。虽然“安全”和“言论”有时可能会出现矛盾，但政策理由应更详细地说明，当“言论”被压制时，政策所要预防和破坏的“现实世界的伤害”。	内容政策	●	
5	在危险人物和组织政策中添加关于“支持”不包括哪些内容的明确说明。用户应该可以自由讨论涉嫌侵犯和践踏指定组织成员人权的行。这不应仅限于被拘留的个人。它应包括对 Facebook 企业人权政策中引用的联合国人权公约保护的权利的讨论。例如，这应允许讨论关于酷刑或残忍、不人道或有辱人格的待遇或处罚、侵犯公平审判权以及法外处决、即决处决或任意处决、强迫失踪、特别引渡和取消公民身份使人无国籍的指控。要求对侵犯和践踏人权行为追究责任的呼吁也应受到保护。煽动暴力行为或招募人员加入或以其他方式向 Facebook 指定组织提供物质支持的内容应排除在保护范围之外，即使相同的内容也讨论了人权问题。用户的意图、他们发布的更广泛的背景以及其他用户如何理解他们的帖子，是确定此类帖子可能导致现实世界伤害的可能性的关键。	内容政策	●	
6	在社群守则中阐明用户可以如何让他们清楚地了解他们的帖子背后的意图。通过执行委员会现有的公开披露公司指定个人和组织名单的建议，将有助于实现这一目标（请参阅：案件 2020-005-FB-UA）。Facebook 还应提供说明性示例，以说明允许内容和禁止内容之间的界限，包括与阐明“支持”不包括哪些内容的规则的应用有关的内容。	内容政策	⦿	
7	通过 Facebook 的产品政策论坛（包括公开征求意见），确保利益相关者有意义地参与拟议的政策变更。Facebook 应跨地区以多种语言开展该互动，确保受该政策旨在防止的伤害影响最大的个人的有效参与。这种参与还应包括对这些伤害以及过度执行现有政策可能导致的伤害具有专业知识的人权、民间团体和学术组织。	内容政策	⦿	
8	确保向内容审核员提供有关任何新政策的内部指导和培训。应为内容审核员提供足够的资源以了解新政策，并使其在执行政策时有足够的时间做出决策。	执行	●	
9	确保用户在其内容被删除时收到通知。通知应指出，删除是由于政府的要求，还是由于违反了社群守则，或者由于政府声称违反了国家法律（以及任何删除的管辖范围）。	执行	●	

委员会对 Meta 回应的评估 ● 全面, ⦿ 比较全面, ✘ 不全面

委员会对落实情况的评估

- 通过公布的信息证明落实情况
- 已汇报进度

- 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
- 已忽略、拒绝或修改的建议

建议标识	监督委员会建议	类别	委员会对 Meta 回应的评估	委员会对落实情况的评估
10	根据案件 2020-004-IG-UA 中的建议，向 Instagram 用户阐明 Facebook 的社群守则适用于 Instagram 的方式与适用于 Facebook 的方式相同。	内容政策	●	
奥贾兰单独监禁				
11	提供 Facebook 收到的政府基于违反社群守则（而不是违反国家法律）的内容删除请求次数以及这些请求的结果的信息。	透明度	⦿	
12	提供更全面的信息，按地区和语言分类，说明执行关于“赞扬”和“支持”危险人物和组织的规则的错误率。	透明度	⦿	
缅甸机器人				
1	Facebook 应确保其内部实施标准以内容审核员审查内容的语言提供。如果有必要确定优先级，Facebook 应首先关注人权风险更为严重的情况。	执行	⦿	
巴西新冠疫情封城				
1	Facebook 应进行相称性分析，以确定一系列比删除内容侵扰性更低的措施。如果与新冠肺炎相关的内容曲解了国际卫生当局的建议，并且发现有可能造成身体伤害但并非迫在眉睫，则应根据需要使用侵扰程度较低的措施。建议的措施包括：(a) 针对内容添加标签，以提醒用户注意帖子内容的争议性，并提供指向世界卫生组织和国家卫生部门的观点的链接；(b) 在帖子中引入分歧以防止互动或分享；以及 (c) 降低内容的排名，以防止在其他用户的新闻源中出现。所有这些执行措施都应清楚传达给所有用户，并支持提出申诉。	执行	✘	
2	考虑到新冠疫情的背景，Facebook 应该做出技术安排，优先对公共机构分享的、引起公司注意的潜在健康错误信息进行事实核查，同时考虑当地情况。	执行	✘	
3	Facebook 应该在虚假新闻社群守则中提供更多关于内容何时会有资格进行事实核查的透明度，包括公共机构的帐号是否需要接受事实核查。	内容政策	●	
分享半岛电视台的帖子				
1	在危险人物和组织政策中添加标准和说明性示例，以增加对中立讨论、谴责和新闻报道例外情况的理解。	内容政策	●	
2	确保将社群守则的更新内容迅速翻译成所有可用语言。	内容政策	⦿	
3	聘请与巴以冲突任何一方无关的独立实体进行彻底检查，以确定 Facebook 的阿拉伯语和希伯来语内容审核（包括其对自动化的使用）是否已无偏见地应用。该审查不仅应审查对巴勒斯坦或亲巴勒斯坦内容的处理，还应审查煽动针对任何潜在目标的暴力的内容，无论其国籍、种族、宗教或信仰或政治观点如何。审查应着眼于位于以色列和巴勒斯坦被占领土内外的 Facebook 用户发布的内容。报告及其结论应公开。	透明度	●	
4	就其如何接收和回应政府提出的所有内容删除请求制定一个透明流程，并确保将其包含在透明度报告中。透明度报告应区分因违反社群守则而导致删除的政府请求与因违反当地法律而导致删除或地理封锁的请求以及导致不采取行动的请求。	透明度	●	
哥伦比亚抗议				
1	公布在仇恨言论社群守则下被指定为违反的侮辱性言语清单中的说明性例子，包括在某些情况下可能有害但在其他情况下没有的边缘性案例。	内容政策	●	
2	将社群守则介绍中提供的“允许保留具新闻价值内容”的简短说明与透明度中心关于如何应用该政策的更详细说明关联起来。公司应使用来自各种背景的说明性示例来补充这一说明，包括报道大规模抗议活动。	内容政策	●	
3	为内容审核员制定并公布明确的标准，以便对可能违反社群守则但可能符合“允许保留具新闻价值内容”的公共利益内容进行额外审核。这些标准应涵盖描绘对政治问题的大规模抗议的内容。	执行	✘	
4	通知所有举报内容被评估为违规但出于公共利益原因留在平台的用户，该帖子已应用“允许保留具新闻价值内容”。该通知应链接到透明度中心对“允许保留具新闻价值内容”的解释。	执行	●	
南非诋毁性措辞				

委员会对 Meta 回应的评估 ● 全面, ⦿ 比较全面, ✘ 不全面

委员会对落实情况的评估

- 通过公布的信息证明落实情况
- 已汇报进度

- 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
- 已忽略、拒绝或修改的建议

建议标识	监督委员会建议	类别	委员会对 Meta 回应的评估	委员会对落实情况的评估
1	根据案件审理决定 2020-003-FB-UA (亚美尼亚人侮辱阿塞拜疆) 和 2021-002-FB-UA (扮成“黑彼得”) 中的建议, 通知用户使用 Facebook 时使用的语言违反了仇恨言论社群守则中的特定规则。例如, 在这种情况下, 应该通知用户他们违反了诋毁性言论禁令。委员会注意到 Facebook 对案件审理决定 2021-002-FB-UA 中第 2 条建议的回应, 其中描述了一个新的分类器, 该分类器应该能够通知使用英语的 Facebook 用户他们的内容违反了诋毁性言论规则。委员会期待 Facebook 提供确认对英语用户的落实情况的信息, 以及有关对其他语言用户的落实时间表的信息。	执行	●	
串珠带				
1	向用户提供及时且准确的通知, 说明公司正在对其申诉内容采取的任何行动。在适用的情况下, 包括在类似本案件的执行错误案件中, 向用户发出的通知应承认该行动是监督委员会审查流程的结果。当委员会行动影响用户申诉的内容决定时, Meta 应共享发送的用户消息, 以证明已遵循此建议。应针对在委员会流程的资格认定阶段纠正的所有案件采取这些行动。	执行	●	
2	研究改进的二次审查方法对审查员准确性和吞吐量的影响。特别是, 当内容审核员被告知他们正在进行二次审查时, 委员会要求评估准确率, 这样他们就知道最初的决定存在争议。理想情况下, 该实验应包括让用户提供相关背景的机会, 以帮助审查员根据委员会先前的建议评估其内容。Meta 应与委员会分享这些准确性评估的结果, 并在其季度委员会透明度报告中总结结果, 以证明其已遵循此建议。	执行	●	
3	进行准确性评估, 重点是涵盖艺术表达和关于侵犯人权的表达的仇恨言论政策允许范围 (例如, 谴责、提高认识、自我参照使用、授权使用)。这项评估还应该具体调查审核员的位置如何影响审核员准确评估来自同一或不同地区的仇恨言论和反击言论的能力。委员会了解, 这种分析可能需要开发适当且准确标记的相关内容样本。Meta 应与委员会分享该评估的结果, 包括这些结果将如何为改进执行行动和政策制定提供信息, 以及是否计划对这些允许情况进行定期审查员准确性评估, 并在其季度委员会透明度报告中总结结果, 以证明其已遵守这一建议。	执行	●	
死藤水酿造				
1	委员会重申其在案件审理决定 2020-004-IG-UA 和 2021-006-IG-UA 中的建议, 即 Meta 应向用户说明公司在 Instagram 上执行 Facebook 社群守则, 但有几个特定例外情况。委员会注意到 Meta 对这些建议的回应。虽然 Meta 可能正在采取其他行动来遵守这些建议, 但委员会建议 Meta 在 90 天内更新 Instagram 社群准则 (“简短”社群准则) 的介绍, 以告知用户, 如果内容在 Facebook 上被视为违反, 则正如公司的透明度中心所述, 也将在 Instagram 上被视为违规, 但有一些例外情况。	执行	●	
2	委员会重申其在案件审理决定 2021-005-FB-UA 和 2020-005-FB-UA 中的建议, 即 Meta 应向用户准确解释他们违反了内容政策中的哪些规则。	执行	●	
3	为了尊重不同的传统和宗教表达和做法, 委员会建议 Meta 修改 Instagram 社群准则和 Facebook 管制商品社群守则, 以允许对有历史证据的非医疗药物的传统和宗教用途进行积极讨论。委员会还建议 Meta 公布所有允许情况, 包括现有允许情况。	内容政策	●	
在 Raya Kobo 涉嫌犯罪				
1	Meta 应重写 Meta 的“安全”价值观, 以反映网络言论除了存在威胁、排斥和沉默的风险, 还可能对人身安全和生命权构成风险。	内容政策	●	
2	Facebook 的社群守则应该表明, 在战争和暴力冲突的背景下, 未经证实的谣言会对生命权和人身安全构成更高的风险。这应该反映在审核过程的各个层面。	内容政策	✘	
3	Meta 应委托进行独立的人权尽职评估, 了解 Facebook 和 Instagram 如何被用来传播仇恨言论和未经核实的谣言, 从而加剧了埃塞俄比亚的暴力风险。该评估应审查 Meta 为防止其产品和服务在埃塞俄比亚被滥用而采取的措施是否成功。该评估还应该审查 Meta 为允许对埃塞俄比亚的人权暴行进行确证和公共利益报道所采取的措施是否成功。评估应审查 Meta 在埃塞俄比亚的语言能力, 以及这些能力是否足以保护其用户的权利。评估应涵盖 2020 年 6 月 1 日至今的时段。公司应在回应这些建议后的六个月内完成评估。评估情况应完整公布。	透明度	●	

委员会对 Meta 回应的评估 ● 全面, ● 比较全面, ✘ 不全面

委员会对落实情况的评估

- 通过公布的信息证明落实情况
- 已汇报进度

- 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
- 已忽略、拒绝或修改的建议

