

2024 年 5 月

打击针对女性的 政治暴力

使用 Facebook、Instagram、WhatsApp 和 Messenger 提供的各种安全工具



目录

01	前言	03
02	使用工具预防和遏制针对女性的政治暴力	05
	📄 举报内容	06
	💬 仇恨言论	08
	👥 欺凌和骚扰	10
	📄❗ 错误信息	13
03	预防针对女性的政治暴力	14



1 前言

1

前言

Meta 致力于保护用户自由表达的权利，让用户能够在安全的环境中与他人交流互动、分享内容。女性理应平等地享有互联网提供的所有经济机会、教育和社交联系，这就是为什么我们要采取一套全面的做法，将我们的技术打造成对女性更加安全的环境，为此我们付出的努力包括编写和制定政策、与专家交流以及开发尖端技术，从一开始就杜绝欺凌行为的发生。

我们的社群守则概述了 Meta 技术中允许和禁止的内容和行为。社群守则由我们的政策团队制定，其中就规定了禁止出现尤其会对女性造成严重影响的行为，例如未经他人同意而分享私密图像，这种行为在全球许多地方都是非法的。社群守则还包含禁止骚扰行为的规定，例如向已经明确表示不希望接收扰人消息的用户屡次发送骚扰消息。我们制定了各种程序，结合使用技术（例如人工智能）和人工团队，努力消除和减少有害内容。



除了通过主动检测、政策和技术来阻止针对女性的威胁行为，我们还提供一系列工具，帮助用户自主掌控他们在 Meta 的体验。同时，我们也在不断改进这些工具、完善各项政策、强化资源配置，持续适应线上线下发生的变化并采纳外部专家提供的反馈意见，例如我们的全球女性安全专家顾问提供的意见。

本指南介绍了相关的实用工具并提供了一些技巧建议，为您在我们旗下各平台上与社群成员安全交流互动保驾护航：



2

使用工具预防和遏制 针对女性的政治暴力

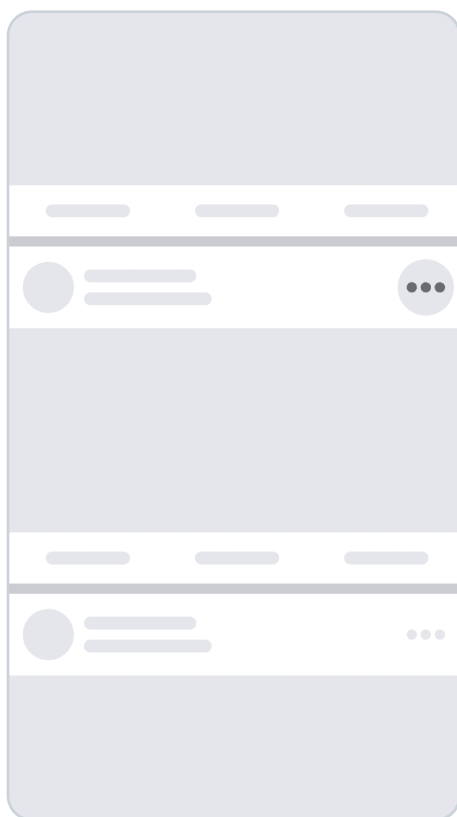
2

使用工具预防和遏制针对女性的政治暴力

针对女性的政治暴力具有各种各样的表现形式。下面这些工具可帮助预防或遏制数字空间内发生的一些暴力行为，是向主管当局报告问题等手段之外的辅助措施。

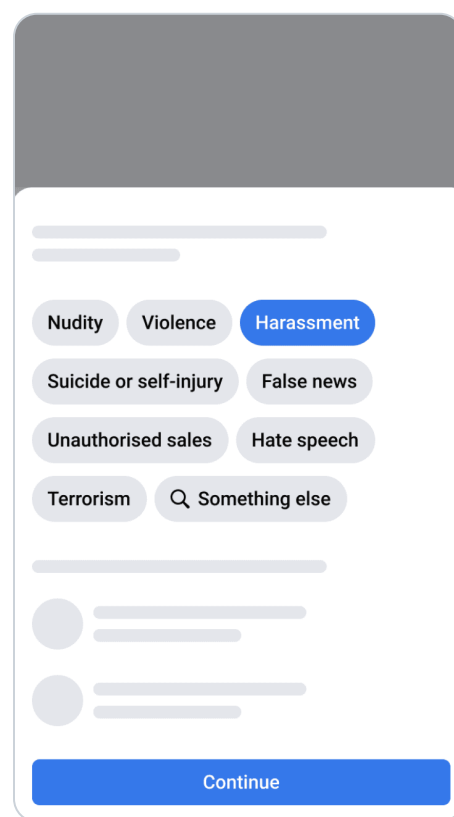


举报内容



01 通用入口：

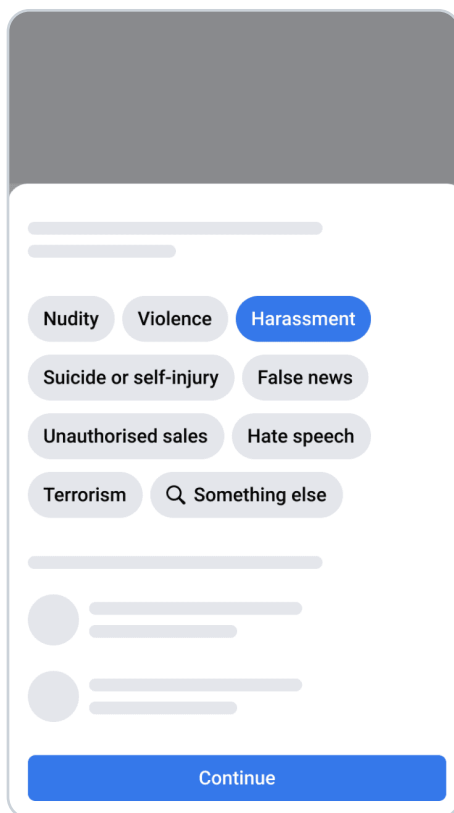
无论是帖子、评论、快拍、消息还是其他内容，我们都提供举报选项。



02 开始举报：

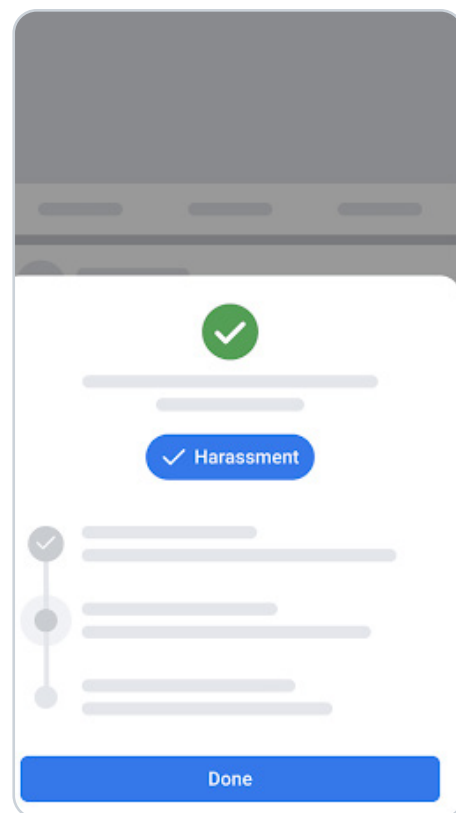
用户可以举报其认为不应该出现在我们技术中的内容。

举报内容（续）



03 选择问题：

我们会请用户详细说明出现了何种问题。这有助我们将举报移交至适当的处理渠道。



04 提交举报：

完成这些步骤后，便可提交举报。我们还会向用户说明接下来会发生什么。



如需详细了解如何举报内容，请前往 [Meta 政策及信息公示平台](#) 查看我们的政策。

2



仇恨言论

当人们觉得自己不会因身份而受到攻击时，才会以真实形象示人，更自由地与外界交流互动。所以我们禁止在旗下技术中发布仇恨言论。仇恨言论会营造恐吓和排斥的氛围，在某些情况下甚至可能助长现实世界中的暴力。我们对仇恨言论的定义是：针对他人受保护的特征，而非观念或习俗发起的直接言论攻击，这些特征包括：种族、族裔、原国籍、残疾、宗教信仰、种姓、性取向、性别、性别认同，以及严重疾病。

预防仇恨言论可使用的工具



屏蔽骚扰消息

详细了解如何在 [Facebook](#)、[Instagram](#)、[Messenger](#) 和 [WhatsApp](#) 执行此操作



审核辅助工具：防止显示冒犯性评论。

详细了解如何在 [Facebook](#) 和 [Instagram](#) 执行此操作



禁用评论：您可以关闭评论功能，禁止用户对个别帖子发表评论。关闭评论功能时，帖子当前已有的评论将隐藏起来。

详细了解如何在 [Facebook](#) 和 [Instagram](#) 执行此操作



置顶正面评论：鼓励积极的行为。

详细了解如何在 [Instagram](#) 执行此操作

仇恨言论（续）

遏制仇恨言论可使用的工具

- ✔ **举报冒犯性评论和威胁性消息：**识别并移除违反我们[社群守则](#)的内容。
详细了解如何在 [Facebook](#) 和 [Instagram](#) 执行此操作
- ✔ **拉黑个人主页、公共主页、小组和应用：**当您拉黑某人后，对方将无法再看到您的帖子、标记您的公共主页、邀请您的公共主页参加活动或加入小组，或者与您的公共主页发起对话。对方如果退出了被拉黑的账户，则仍可看到您的公开内容。
详细了解如何在 [Facebook](#)、[Instagram](#) 和 [WhatsApp](#) 执行此操作
- ✔ **限制评论：**在不通知对方的情况下限制用户发布评论。
详细了解如何在 [Facebook](#) 和 [Instagram](#) 执行此操作
- ✔ **管理垃圾评论：**滥发垃圾信息是指向他人发送骚扰性质的内容或请求。这包括群发消息、在他人的时间线上发布大量的链接或图片。
详细了解如何在 [Facebook](#) 和 [Instagram](#) 执行此操作

2023 年第 4 季度，Facebook 上仇恨言论的传播度为 0.01% - 0.02%。换言之，根据估算，

每 10,000 次

内容观看中，仅 1-2 次可能包含仇恨言论²

2



欺凌和骚扰

欺凌和骚扰可能发生在很多地方，形式多样。言语威胁、公布个人身份识别信息、发送恐吓信息以及扰人的恶意联系，这些都属于欺凌和骚扰的范畴。我们不会容忍这类行为的发生，因为它会使 Meta 技术中的用户感到不安全和不受尊重。

我们鼓励用户积极交流讨论，但这些讨论经常涉及对新闻人物或知名人物的批判性评论，因此在政策中，我们对公众人物和普通个人进行了区分。对于公众人物，我们会移除针对他们的严重攻击内容，以及在帖子或评论中直接标记他们的某些攻击内容。我们对公众人物的定义是：州和国家级政府官员、这些职位的政治候选人、在社交媒体上的粉丝人数超过 100 万的人物，以及得到大量新闻报道的人物。

预防欺凌和骚扰可使用的工具



评论警告：当有人试图发布潜在冒犯性评论时，我们将显示一条警告消息，提醒他们注意遵守我们的社群守则并告知他们如果继续发布此评论，我们会将其移除或隐藏。



标记和提及控制选项：您可以选择是允许所有人、仅您关注的用户，亦或是禁止任何人在评论、帖子标题或快拍中标记或提及您。

欺凌和骚扰（续）

预防欺凌和骚扰可使用的工具

- ✔ **隐藏字词：**某些评论和陌生消息并未违反我们的社群守则，但可能被视为不当或具冒犯性，这种情况下，您可以将其过滤筛除。
- ✔ **限制互动：**自动隐藏来自非粉丝用户发布的评论和陌生消息，在一段时间内限制扰人的互动。
- ✔ **限制评论：**当您限制某位用户发布评论后，对方在您的帖子中发布的评论只对其本人可见。
- ✔ **锁定个人主页：**锁定个人主页功能目前仅在部分地区可用。此功能可防止您的照片、帖子及关于您的信息被您认为不适当的用户看到。
- ✔ **处理骚扰私信时可使用的限制功能：**新推出的功能，旨在帮助更好地保护我们的社群远离通过私信发送的骚扰图片和视频。

邀请模式：向用户发送一条消息，以此邀请对方继续聊天。只有当对方接受您的聊天请求后，您才能继续向对方发送私信。

纯文本：邀请消息只能是纯文本格式。这意味着，如果是非粉丝用户，只有当对方接受您的聊天请求后，您才能向他们发送图片、视频或语音消息。

详细了解我们如何预防 Facebook 和 Instagram 上的欺凌和骚扰行为。



2

欺凌和骚扰（续）

遏制欺凌和骚扰及预防图像骚扰行为时可使用的工具

 **举报：**如果您看到违反我们社群守则的内容，请举报，以便我们的团队可以审核内容并采取行动。我们的团队全天候运作，覆盖 70 多种语言，并且会对违规内容和行为采取行动。

详细了解如何在 [Facebook](#) 和 [Instagram](#) 执行此操作

 **拉黑：**拉黑某用户后，对方将无法看到您的个人主页、帖子或快拍。

详细了解如何在 [Facebook](#) 和 [Instagram](#) 执行此操作

 **向我们举报有关您的裸露和私密照片或视频，**或向您发出将这些照片和视频分享到我们技术中的威胁行为，防止这些照片和视频被转发出去。

详细了解如何在 [Facebook](#)、[Instagram](#) 和 [Messenger](#) 执行此操作

 **使用 [StopNCII.org](#) 防止您的裸露或私密照片或视频被分享到网上：**

➤ 如果您有满 18 周岁后拍摄的私密图片或视频，并且您担心这些图片或视频会在网上被分享或转发，您可以通过 [StopNCII.org](#) 帮助防止其被进一步传播。

➤ 此工具使用散列生成技术为图像分配一条专属的散列值（数字代码），从而创建一个安全的数字指纹。加入 [StopNCII.org](#) 的科技公司会收到这个散列值，并可使用此散列值检测他们的平台中是否有人分享或试图分享这些图像。

在 Instagram，有

95.3%

的违规内容是我们先于用户举报发现并处理的³

2023 年第 4 季度，Facebook 上仇恨言论的传播度为 0.01% - 0.02%。换言之，根据估算，

每 10,000 次

内容观看，仅 1-2 次可能包含仇恨言论³

2



错误信息

Meta 致力于防止错误信息的传播。我们采用政策执行技术、人工审核以及独立核查机构相结合的方式发现、审核和处理这类内容。



移除

我们重视表达自由，同时又致力于保障用户的安全，因此在一些特定情况下，我们会移除 Meta 技术中的错误信息：

- ④ 错误信息有可能诱发紧迫的人身伤害。
- ④ 错误信息有可能干预或压制投票。
- ④ 视频以常人难以察觉的方式被编辑（会让大众误以为视频中的人说了他们本没有说过的话），或者视频属于人工智能的产物，通过内容的组合、替代或者叠加制作成一个看似真实的视频。



移除违规内容



减少传播

一旦某条内容被我们的某个事实核查合作机构评定为虚假，我们会减少其在动态及其他地方的传播，避免让更多人看到。对于屡次分享这类内容的公共主页和小组，我们不仅会减少它们动态中所有帖子的传播，还会取消它们上推荐位的机会。



减少传播问题内容



警告与提示

我们会给已经过事实核查机构评定的帖子贴出提示，并给帖子发布者发送通知。这样，用户便能看到事实核查机构的结论，并据此判断自己是否要阅读、信任或分享某条内容。另外，我们还与世界各地的组织开展合作，共同努力提高人们的新闻素养。



为敏感内容或误导性内容提供背景信息

03

预防针对女性的 政治暴力

3

预防针对女性的政治暴力

我们坚定不移地践行保障女性安全的承诺。近些年来，我们通过寻求相关领域专家的协助，确保我们的技术对女性是安全可靠的，我们在这方面的成果包括：

- ④ 制定健全的政策，保护女性远离网络欺凌。
- ④ 开发工具，让用户可以掌控自己的线上体验。
- ④ 发布技术，打击未经他人同意而分享私密图像的行为。

您可以如何帮助预防针对女性的政治暴力



提高认知：

一直以来，针对女性的政治暴力问题从未得到应有的关注度，相关的统计数据仍然匮乏。



使用我们的技术曝光暴力行为：

勇敢发声并分享亲身经历的政治暴力事件对于打击这种行为非常重要。



熟悉我们的社群守则：

我们有责任确保我们的服务远离欺凌行为。为此，我们制定了一套社群守则，规定我们的技术中允许和禁止的内容和行为。



在选举周期内与 Meta 合作保护自身安全：

选举周期内，政治账户可能面临更多威胁，因此可能需要更多保护。我们推出了 Facebook 账户保护计划，旨在让政治候选人、民选官员、联邦和州级部门和机构、政党委员会及其工作人员能够更好地保护他们的 Facebook 账户。