

## YouTube新功能：如果评论具有攻击性 将向播主发送通知

YouTube上的有毒和仇恨的评论一直是公司、创作者和用户的头疼问题。该公司以前曾试图通过引入一些功能来减少这种情况，如在发帖时向个人发出警告，以便他们发帖前能够思量再三。现在，这家流媒体服务机构正在推出一项新功能，将更积极地向播主提醒这些人的辱骂性评论，并采取更广泛的行动。



YouTube表示，它将向那些因违反平台规则而被删除辱骂性评论的人发送通知。如果用户在收到通知后仍继续发布辱骂性评论，该服务将禁止他们在24小时内发布任何评论。该公司表示，它在今天推出该功能之前进行了测试，发现通知和超时被证明具有实质性的成功。

目前，仇视性评论检测只适用于英语评论，但该流媒体服务的目标是在

未来包括更多语言，此前发帖前的警告可用于英语和西班牙语。

该公司说："我们的目标是既保护创作者不受试图通过评论对社区产生负面影响的用户的影响，又为那些可能因违反政策而被删除评论的用户提供更多的透明度，希望能帮助他们理解我们的社区准则。"

如果用户认为他们的评论被错误地删除，他们可以分享他们的反馈。不过，该公司没有说在看完反馈后是否会恢复评论。

此外，在一个论坛帖子中，YouTube表示，它一直在努力改善其人工智能驱动的检测系统。该公司声称，在2022年上半年，它已经删除了11亿条"垃圾"评论。它说，YouTube还加强了其系统，以更好地检测和删除实时聊天视频中的机器人。

YouTube和其他社交网络已经能够减少垃圾邮件和滥用内容，部分是依靠自动检测。然而，滥用者经常使用不同的俚语或拼错字来欺骗系统。更重要的是，用非英语语言发布仇恨性评论的人更难被发现。

该流媒体公司在最近几个季度测试了一系列工具，以减少平台上的仇恨性评论。这些测试包括默认隐藏评论，以及在用户的个人资料卡上显示用户的评论历史。

上个月，YouTube推出了一项功能，让创作者隐藏某个特定用户的评论。这种控制适用于整个频道，因此，即使该用户在另一个视频上发表仇恨性评论，也不会显示出来。

全球的平台都在努力解决遏制仇恨性评论传播的问题。当英格兰足球运动员布卡约-萨卡、马库斯-拉什福德和贾登-桑乔在去年的欧洲杯决赛中因罚失点球而受到骚扰时，Instagram就成了网暴者的温床。GLAAD和Media Matters的一份新报告指出，在埃隆-马斯克接管Twitter后，反LGBTQ的污言秽语急剧增加。虽然所有这些平台都推出了工具，将评论静音或隐藏，并限制对某些人的评论，但仇恨和辱骂的评论数量仍然是一个亟待解决的问题。

本文链接：<https://dqcm.net/zixun/16710644373180.html>