

YouTube 如何保障社群免於仇恨言論與騷擾行為的侵擾

YouTube 的使命是賦予每個人發聲的機會。仇恨言論與騷擾行為會使人怯於向大眾分享自己的故事和經驗為此我們制定相關政策，禁止在 YouTube 上發布這類內容。我們會比照其他有害內容處置方式，迅速移除違反仇恨言論與騷擾行為處理政策的內容。



YouTube 政策的保護對象

我們是根據惡意內容的受害對象來區分仇恨言論與騷擾行為。

仇恨言論政策的保護對象是團體

仇恨言論政策禁止使用者上傳任何煽動暴力或激起仇恨的內容，避免理應受到保護的團體受害。

騷擾行為處理政策的保護對象是個人

我們的騷擾行為處理政策禁止使用者對可識別身分的人士進行惡意威脅或侮辱，攻擊其與生俱來的特質 (包括應受保護的特質或身體外貌)。

應受保護的特質

除了現行法令中定義的應受保護團體，我們也諮詢民權與仇恨言論方面的專家，納入更多種應受保護的特質。

- ✓ 種族
- ✓ 族裔
- ✓ 國籍
- ✓ 宗教
- ✓ 身心障礙
- ✓ 性與性別
- ✓ 性傾向
- ✓ 移民身分
- ✓ 年齡
- ✓ 種姓
- ✓ 退役身分
- ✓ 重大暴力事件的受害者或其親屬

YouTube 政策的防範重點

以下僅列舉幾種 YouTube 禁止的內容，完整資訊請上網查看[仇恨言論與騷擾行為處理政策](#)。

仇恨言論

- × 意圖將現實世界暴力正當化的陰謀論
- × 貶低人格
- × 煽動暴力
- × 劣等化
- × 優越主義
- × 威脅
- × 否定暴力事件存在

騷擾行為

- × 人肉搜尋
- × 模擬暴力
- × 跟蹤
- × 威脅
- × 非當事人同意的情慾內容
- × 對暴力事件受害者的騷擾
- × 求死的願望
- × 貶低人格
- × 嚴重侮辱
- × 針對性侮辱
- × 以弱勢個人為目標

YouTube 如何隨時因應仇恨言論與騷擾行為的演變趨勢？

有了適當的技術、專家與政策，現在我們能比以前更快移除含有仇恨言論與騷擾內容的影片、頻道與留言。

Intelligence Desk 是我們的內部團隊，負責關注新聞來源、社交媒體和使用者檢舉內容



我們與外部專家合作，針對暴力極端主義、優越主義、民權、網路霸凌及言論自由等議題交換意見




我們持續聽取各類創作者的意見，參考他們的經驗來制定及更新政策




YouTube 如何落實政策執行？


社群成員與機器學習系統都會協助我們偵測仇恨言論與騷擾行為

全球每分鐘上傳到 YouTube 的影片長度超過 500 小時，我們除了借助觀眾與創作者的力量，也斥資開發機器學習系統來大規模偵測不當內容。

 政府單位與非政府組織通常較能準確檢舉含有仇恨言論與騷擾行為的內容，可以透過認證檢舉人計畫協助我們


 創作者能利用留言管理工具封鎖特定字詞/使用者，避免相關字詞或使用者的留言出現在留言區


 觀眾可以檢舉影片與留言


 機器學習系統讓我們得以大規模偵測違規內容。我們的系統訓練有素，可以隨時偵測平台中的有害內容，在同樣的時間內以遠高於人工作業的速度來審查數十萬部影片

內容審查人員評估遭檢舉的內容

系統不會自動移除遭檢舉的內容。我們全球的专业審查團隊將根據仇恨言論與騷擾行為處理政策，決定要移除哪些內容。他們的判斷結果也會用來訓練及改進 YouTube 的機器學習系統。

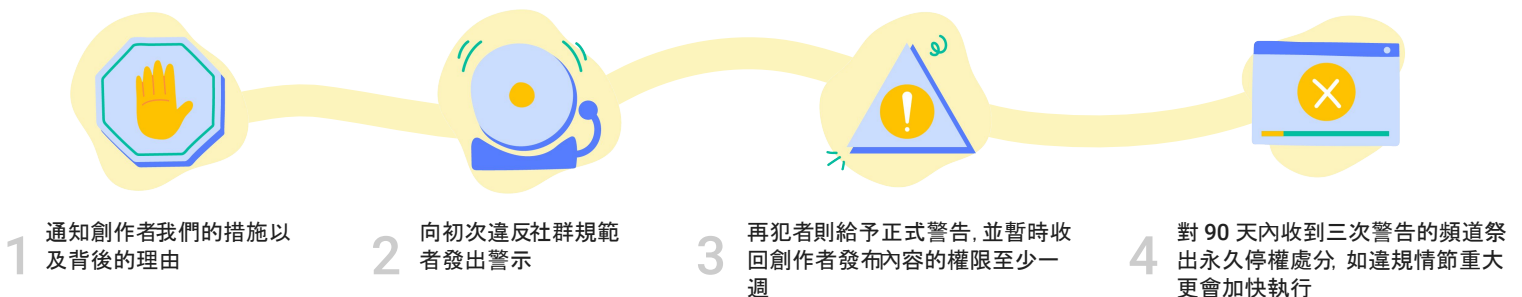
 全球團隊精通數十種語言，能審慎評估遭檢舉的內容

 評估技術受過專業訓練，能根據內容脈絡或語言去理解細部涵義

 公平落實政策規範，不受創作者身分、政治理念、立場或背景的影響

 移除違反政策的內容

我們會對發布仇恨與騷擾內容的創作者祭出處分



政策例外狀況

我們可能會破例允許以下類型的內容：具有明確教育、紀實、科學或藝術（簡稱 EDSA）意義的內容。舉例來說，如果影片中含有不雅或令人反感的用語，但觀眾知道影片純屬喜劇或諷刺性質，我們可能會允許發布這類內容。

減少散播違規邊緣內容

某些內容遊走在仇恨言論與騷擾行為處理政策的邊緣。系統會盡力減少在觀眾的首頁和「即將播放」清單中推薦這類違規邊緣內容，以降低傳播範圍。

YouTube 在移除仇恨言論與騷擾內容上的進展

得力於合適的技術、專家與政策，現在我們能比以前更快移除含有仇恨言論與騷擾內容的影片、頻道與留言。

仇恨言論*

移除的影片數量超過 **88,000** 部

移除的留言數量超過 **4,300 萬** 則

騷擾行為*

移除的影片數量超過 **322,000** 部

移除的留言數量超過 **1 億 3,500 萬** 則

70%

的違規影片在觀看次數累積未達 10 次前就遭移除*