

共建 **更安全** 網上世界： API 的功用

為保障使用者及其他機構免受壞份子的侵害，我們致力構建和分享各種可偵測和阻擋網上威脅的應用程式設計介面 (API)，令大眾可以享有更安全的互聯網環境。



構建

我們構建各種安全 API，為三大領域提供安全保障。
這些 API 包括：

兒童安全

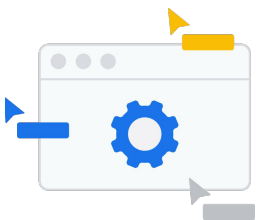
- Content Safety API
- CSAI Match API
- Hash Matching API

安全性

- Safe Browsing API
- Project Shield API
- VirusTotal API

資訊品質

- Perspective API
- Vision API
- FactCheck API
- Civics Information API
- Text Moderation



擴展影響力

為擴展影響力，我們與眾多合作夥伴分享各種安全 API：

4,600 萬

2022 年，Cloud Armor API 成功阻止歷來其中一次最大規模的第 7 層 DDoS 攻擊，這次攻擊的網絡請求量多達每秒 4,600 萬個。

20 億

超過 1,000 個合作夥伴每天存取 Perspective API 接近 20 億次。

40 億

「兒童安全工具包」(Content Safety API + CSAI Match API) 在過去 30 天處理了超過 40 億張圖片和影片。



限制

為確保只有值得信賴的合作夥伴可以存取 API，並以安全合規的方式使用，我們採取以下三大措施：



我們只允許經過嚴格審查的合作夥伴存取 API、限制在未經授權下使用這些工具，並嚴格執行服務條款政策。我們為值得信賴的合作夥伴提供安全 API，藉此互相合作，為大眾共建更安全的互聯網環境。

Google 的安全 API



兒童安全 API

兒童安全工具包

Content Safety

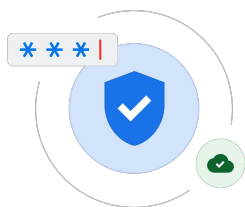
Google 的專有模型可協助合作夥伴分類數十億計兒童性虐待內容圖像，並加以優先處理。此模型會運用機器學習分類器，識別從未見過的兒童性虐待內容圖像，以便加快審核、移除和舉報流程。

CSAI Match

YouTube 的專有技術 (也是首項採用雜湊配對方法的技術) 可識別並系統性地標記含兒童性虐待圖像 (CSAI) 的影片，以供合作夥伴作審核、確認、舉報和採取行動。

Hash Matching

我們為 NCMEC 提供 Google Hash Matching API，協助他們更有效率地優先處理和審核 CyberTipline 報告，以鎖定涉及急需協助兒童的報告。



安全性 API

Google WAAP 解決方案

Cloud Armor

保護網站、服務和 API 免受針對性和自動化的進階分散式阻斷服務 (DDoS) 攻擊，以免資訊被遮蔽。

reCAPTCHA

保護網站免受欺詐活動、垃圾內容和其他濫用行為侵害。reCAPTCHA Enterprise 採用自動調節風險分析引擎，防止壞份子以自動化軟件從事濫用行為。

Apigee

Google Cloud 提供完整生命週期 API 管理平台，有助企業設計、保護、部署、監察和擴展應用各種 API。

Safe Browsing

Safe Browsing API 讓用戶端應用程式可存取我們定期更新的不安全網絡資源清單，以檢查網址安全性。這項工具每天保護 50 億部裝置免受惡意或垃圾軟件侵害。如工具發現網站包含這些內容，便會向使用者發出警告。

VirusTotal

讓開發人員提交檔案或網址以作分析，並接收報告，以判斷檔案或網址是否受惡意軟件感染。



資訊品質 API

Perspective

這項開放原始碼 API 會採用機器學習技術識別「惡意」留言，不但有助協導人與人之間的互動，還可提升人類與大型語言模型和生成式 AI 的互動品質。

Vision

讓開發人員輕鬆將視覺偵測功能整合至應用程式，包括標籤圖片、偵測面孔和地標、光學字元辨識 (OCR)，以及標記露骨色情內容。

FactCheck

旨在協助事實查核人員、記者和研究人員完成事實查核工作，確認哪些內容已經由全球核實。運用這項工具，使用者可輕鬆搜尋有聲譽發佈者提供的事實查核聲明，數量多達 30 萬則。

Civic Information

Civic Information API 讓開發人員可以製作應用程式，協助美國公民和選民瞭解政治代表的資料、選舉資訊和投票地點。在 API 支援的選舉期間，選民可輕鬆搜尋投票地點、提前投票和票箱位置，以及候選人資訊。

Text Moderation

Google 透過 Cloud Natural Language API 提供 Text Moderation 工具，協助機構掃描敏感或有害內容。這項工具可識別多種有害內容，包括仇恨言論、欺凌行為和性騷擾等。

每天都可更
安全地使用
Google

我們投資構建及分享可偵測和阻擋網絡威脅的 API，確保您**每天都可更安全地使用 Google**。歡迎瀏覽我們的[安全中心](#)，深入了解我們如何讓更多人更安全地使用互聯網。