攜手建立更安全的網路：API 扮演的角色

為了保障使用者和其他機構的安全，免於惡意行為者的侵害，我們建構並分享應用程式設計界面 (API) 來偵測和防堵網路威脅，協助打造更安全的網路環境。

建構

為了在三個重要領域提供保障，我們打造了安全 API，這類 API 包括：

兒童安全

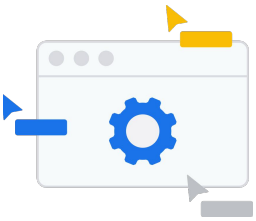
- Content Safety API
- CSAI Match API
- Hash Matching API

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



擴大影響力

我們將安全 API 與合作夥伴分享，以擴大影響力：

4,600 萬
2022 年，Cloud Armor API 阻止了史上規模最大的第 7 層 DDoS 攻擊之一，此次攻擊每秒發起 4,600 萬次請求。

20 億
1,000 多名合作夥伴每天呼叫 Perspective API 近 20 億次。

40 億
兒童安全工具包 (Content Safety API + CSAI Match API) 在過去 30 天內處理了超過 40 億個圖片和影片。

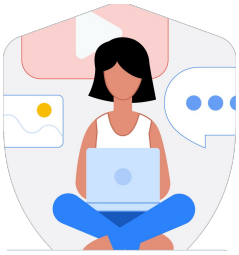
限制

為了確保只有值得信賴的合作夥伴可存取我們的 API，並以安全合規的方式使用，我們採取了三項重要措施：

審核 → 限制 → 執行

我們僅允許經過仔細審核的合作夥伴存取我們的安全 API、限制未經授權的使用行為，並嚴格執行我們的服務條款和政策。與值得信賴的合作夥伴分享安全 API，我們就能一同努力，為所有人打造更安全的上網環境。

Google 的安全 API



兒童安全 API

兒童安全工具包

Content Safety

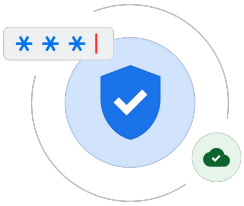
Google 擁有的模型可協助合作夥伴，將數十億張兒少性虐待圖像分類並排定審查優先順序。此模型使用機器學習分類器，識別新的兒少性虐待圖像，以加快審查、移除和檢舉的速度。

CSAI Match

YouTube 擁有的技術 (也是第一個使用雜湊比對的技術) 可識別並有系統地標記出現兒少性虐待圖像 (CSAI) 的影片，以供合作夥伴審查、確認、檢舉和採取行動。

Hash Matching

我們將 Google Hash Matching API 提供給 NCMEC，協助他們有效安排 CyberTipline 報告的審查順序，若報告提及的兒童急需幫助，便能優先處理。



安全性 API

Google 的 WAAP 解決方案

Cloud Armor

防範鎖定目標的自動分散式阻斷服務 (DDoS) 進階攻擊，避免攻擊者審查網站、服務和 API 的資訊。

reCAPTCHA

保護網站免於詐欺活動、垃圾內容和其他不當行為的侵擾。reCAPTCHA Enterprise 使用會自動調整的風險分析引擎，防止有心人士使用自動化軟體從事不當行為。

Apigee

Google Cloud 提供完整生命週期 API 管理平台，有助於企業設計、保護、部署、監控和擴展 API。

Safe Browsing

Safe Browsing API 可讓用戶端應用程式檢查網址，將其與 Google 持續更新的不安全網路資源清單比對。這項工具每天保護 50 億部裝置，如果發現網站含有惡意軟體或垃圾軟體，就會警告使用者。

VirusTotal

讓開發人員可以提交要分析的檔案或網址，並接收報告以瞭解是否感染惡意軟體。

資訊品質 API

Perspective

這款開放原始碼 API 使用機器學習技術識別惡意評論，不僅可促進人類之間的互動，還有助於提升人類與大型語言模型和生成式 AI 之間的互動品質。

Vision

讓開發人員可以輕鬆地將視覺偵測功能整合到應用程式中，包括為圖片加上標籤、偵測臉部和地標、光學字元辨識 (OCR) 以及標記煽情露骨內容。

FactCheck

這款 API 可協助事實查核人員、記者和研究人員，確定哪些內容已經或尚未在全球各地驗證。使用此工具，使用者可以從信譽良好的發布商提供的 30 萬項事實查核聲明中搜尋。

Civic Information

Civic Information API 讓開發人員可以建構應用程式，幫助美國公民和選民瞭解政治代表、選票資訊和投票地點。在 API 支援的選舉期間，選民可以搜尋投票地點、提前投票地點和投遞箱位置，以及候選人資訊。

Text Moderation

Cloud Natural Language API 提供的 Google Text Moderation 工具可協助機構掃描敏感或有害內容，可識別的有害內容類別相當廣泛，包括仇恨言論、霸凌和性騷擾。

Google 致力於鞏固使用者的線上安全

Google 持續建構並分享 API 來偵測和防堵網路威脅，致力於**鞏固使用者的線上安全**。請造訪我們的安全中心，進一步瞭解我們如何優於業界，為更多人維護線上安全。