Google

Apr 2025

共同打擊詐騙和詐欺

Google 白皮書

摘要

Google 的使命是彙整全球資訊，供大眾使用，使人人受惠。這項使命的核心精神，是為使用者呈現切合需求的高品質資訊。因此，我們非常重視反網路詐騙和詐欺以及提供可靠資訊和內容的責任。

金融詐欺對消費者和正當企業都有害，還會破壞使用者對數位平台的信任。尤其網路詐騙問題在全球日益猖獗，可能會嚴重危害個人和企業，並波及世界各地、各年齡背景的使用者。此外，詐騙分子還會趁機利用新技術和社會趨勢，手法持續演變。COVID-19 疫情更加速了這種現象。隨著對數位服務的依賴加深，網路詐騙和詐欺案例也大幅增加。因此，我們決心並承諾持續創新，投注大量心力來揪出和打擊詐欺行為。

這份政策白皮書聚焦金融詐欺，以通常為跨國集團主導的組織犯罪案為主，當中詳述Google 如何因應這項挑戰並提供建議，幫助我們提升生態系統裡彼此之間的合作，發揮團結力量大的精神，攜手對抗危害社會和經濟的詐騙犯罪。

Google 防範網路詐騙和詐欺的措施

Google 打擊詐騙與詐欺的方式主要有三種，分別是主動採取措施保護使用者、提供可靠資訊，以及與各界聯手打造安全的網路環境。我們的具體做法是制定政策，並採用內建的技術保護功能，以防範、偵測和處置有害及非法的內容。我們還積極與相關專家和機構 (例如國家反詐騙機構) 合作與對話，讓業界領先的措施造福更多人，保障使用者的上網安全。

我們的每項產品都有量身打造的防護機制。舉例來說，Android 的行動作業系統採用多層保護措施來偵測和防範詐騙，像是 Google 電話就透過 AI 技術，過濾可能是騷擾和詐騙的來電。Chrome 則透過多種功能保護使用者，包括 Google 安全瀏覽、Chrome 密碼防護警示、進階保護計畫 (APP)，以及安全瀏覽強化防護功能，其中 Google 安全瀏覽功能會在網站存在危險，並試圖竊取使用者憑證時發出警示。Google Ads 這項服務則已經針對廣告生態系統，制定一系列政策和安全防護功能並定期更新，例如假冒公眾人物身分的政策，或廣告放送量受限政策。這項措施要求廣告主必須經過「認識期」，才有權放送易遭濫用類型的廣告。這些特別制定的政策，就是大幅減少詐騙廣告數量的關鍵。這項措施要求廣告主必須經過「認識期」，才有權放送易遭濫用類型的廣告。這些特別制定的政策，就是大幅減少詐騙廣告數量的關鍵。

憑藉在該領域多年的經驗和努力，我們不僅達成了預期目標，更展現出顯著的影響力。舉例而言：

- Gmail 阻擋 99.9% 的垃圾郵件、惡意軟體和危險連結，保障使用者的收件匣安全。
- Google 安全瀏覽功能協助使用者避免惡意網站的侵害，為超過 50 億部裝置提供自動保護。
- 在廣告，我們在 2023 年封鎖/移除了 2 億 650 萬則違反「不實陳述或行為」政策的廣告 (其中包括許多詐騙和詐欺手法)，以及 2 億 7,340 萬則違反「金融服務」政策的廣告。
- Google Play 安全防護功能每天掃描 2,000 億個 Android 應用程式，保護超過 30 億位使用者免於惡意軟體威脅。
- 我們已將金融服務驗證計畫拓展到全球十七個國家和地區，該計畫規定金融服務廣告主必須證明已取得相關政府機關授權，才能在 Google 平台上透過廣告宣傳特定金融商品和服務。
- 在 Android 平台上，我們 新推出的 Google Play 安全防護反詐欺功能 在新加坡前測計畫的前 6 個月內，封鎖了超過 20 萬部裝置上近 90 萬次高風險側載應用程式安裝作業，而且這項計畫還將擴大適用範圍。

負責任的反詐欺合作與監管原則

儘管 Google 在打擊詐欺和詐騙方面已取得重大進展，但網路犯罪問題會威脅整個社會，只有網路平台致力防範並無法解決問題，而是需要跨產業攜手奮戰，並與政府部門、執法機關、金融機構和電信服務供應商等其他利害關係人合作。我們不能各自行動，這樣不僅無法發揮團結的力量，還會讓我們整體策略出現漏洞，留下可乘之機給犯下大部分詐騙案的跨國犯罪組織。

要解決這個問題，我們需要制定全面的全球公共政策，並集結社會的力量。結合資訊共享框架、採取可行的違規處置、發展創新技術、多方宣導，並提升個人警覺與責任承擔等做法，再加上利害關係人與我們攜手合作，就能打造更安全的網路環境，造福所有使用者。

在這份白皮書中，我們提出了三大政策原則，希望能推動整個生態系統強化合作，更有效打擊網路詐騙和詐欺。每項原則都搭配了一系列政策建議，為具體做法提供大方向。

原則	政策建議
攜手合作、資訊共享 建立新的政府與產業界資訊共享框架，透過國際論壇推動更多合作。	• 制定相關法律，促進政府和產業界之間的合作和資訊共享(包括機密資料的共享)，並在發現可疑活動但證據不充分時，也能採取行動。 • 組成跨境工作小組並商定取得資源的措施，以協調和加強執法，促進互通性。以協調和加強執法，促進互通性。 • 支援全球反詐騙聯盟 (Global Anti Scams Alliance) 等國際論壇，加強跨產業對話與合作。
鼓勵整個社群採取行動 提供必要的法律框架，讓社群成員能在詐欺內容和不肖分子得逞之前搶先行動。	• 明確界定非法活動，加強執法機關的能力。 • 倡導「互助友愛」的責任保護原則，激勵利害關係人採取預防措施。 • 投入資源發展負責任的 AI 技術並制定相關政策，鼓勵發展創新技術，以辨識詐欺性內容，防範詐騙和詐欺。
投入資源向使用者宣導並制定防範措施 主要藉由與其他機構和跨產業合作方式，投入資源宣導教育並提升大眾警覺。	• 推行大眾宣導活動， 解釋網路詐騙手法及自保方法。 • 制定政策和開發產品功能，t有效保護使用者。 • 為詐欺和詐騙受害者提供清楚易用的檢舉管道。 • 追求安全的數位轉型。

前言: 全球的網路詐欺和詐騙問題

詐欺和詐騙問題由來已久：事實上，詐欺和詐騙行為從人類互動的早期就已經存在，詐騙行為經過幾個世紀的發展，從日常的低級詐欺到引發 全國性動盪的重大金融泡沫，複雜度日漸提高。

隨著網際網路和智慧型手機融入我們的日常生活，它們也成為取得個人資料 (如銀行資訊、購物帳戶和健康統計資料) 的重要管道。詐騙者運用偽造訊息、誘人的優惠、冒用他人身分等欺騙性手段，透過不斷擴大的通訊管道展開行動。隨著數位化的推進，犯罪分子在技術的協助下持續發展並擴大其非法活動的規模、範圍和速度。他們利用各種工具和技術欺騙受害者，或利用受害者的心理狀態和情緒盡可能榨取金錢。社群媒體和訊息平台成為他們大規模跨國作案的工具，新型數位金融機構、產品以及電子商務和串流平台等非傳統產業的安全漏洞，都被他們趁機攻擊。

線上詐騙是一種日益猖獗的跨國組織犯罪。根據全球反詐騙聯盟估計，光在 2022 至 2023 年，詐騙者就造成了高達 $1.026 兆美元左右的損失，全球每 4 人中就有 1 人受到影響。根據商業改進局 (Better Business Bureau) 的報告，在 2015 年到 2022 年間，線上詐騙報告數量增長了 87%。

詐欺類型與趨勢

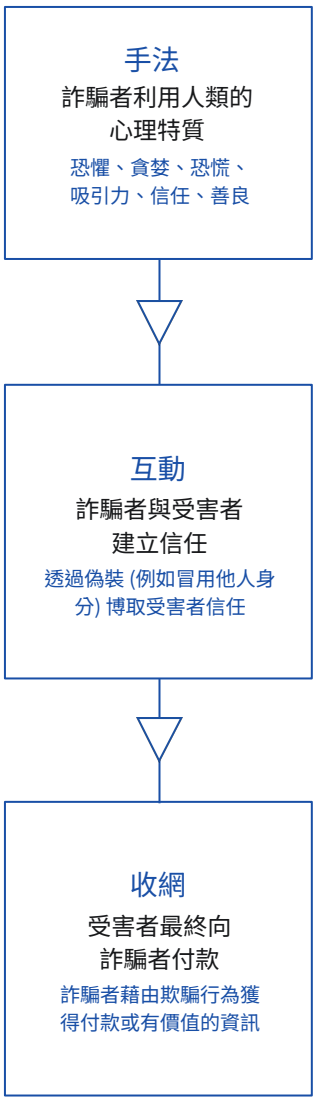
雖然基本手法保持不變，但線上詐騙和詐欺的形式卻變化多端，其中最常見的包括以下幾種：

- 冒用他人身分詐欺：這包括犯罪分子冒充政府官員、親友等，透過行動裝置或社群媒體應用程式與受害者聯絡。他們利用受害者的情緒誘導其付款、交出付款帳戶的控制權，或進行金融活動，例如申請貸款或開設帳戶以接收犯罪所得。其中一種手法是濫用 AI 技術，藉此製作名人和具公信力人物的深度偽造圖片或影片，以提升詐騙騙局的可信度，吸引受害者上當，這也稱為「公眾人物假冒騙局」。

- 投資詐騙 (線上交易/交易平台詐欺)：受害者遭到虛假廣告或線上顧問的矇騙，前往不存在或虛假的交易或投資平台，進行與法定資產和虛擬資產相關的交易或投資。

- 電子商務詐騙：線上購物和數位交易日益普及，隨之而來的是更多起數位購物平台詐騙案，手法也推陳出新，其中一種常見的手法是，誘騙使用者透過線上付款方式，在假冒的購物網站上購買商品。

- 網路愛情詐欺：受害者遭到欺騙，相信自己與對方陷入熱戀，直接或間接匯款給犯罪分子，將財產投入到子虛烏有的投資計畫



- 求職詐騙：隨著線上求職的增加，詐騙者會在看似合法的網站發布虛假的徵人啟事。線上的虛假工作職位以各種藉口誘騙受害者向詐騙者付款，包括預付商品費用以提高交易平台的銷售額，或支付訂金以保障工作機會。
- 付款劫持：詐騙者利用社交工程技巧從使用者處取得私密資訊，進而劫持使用者的付款方式。其中一種特殊手法對部分國家/地區的遊戲社群造成了影響，這是由於當地的支付系統要求玩家掃描 QR code 進行付款。
- 惡意軟體詐騙：使用者下載看似無害的實用應用程式，但實際上其中卻包含可用於竊取個人識別資訊等內容的惡意軟體。
- 網路釣魚詐欺：犯罪分子冒充銀行、公用事業服務或網際網路服務供應商 (ISP) 等可信單位的代表，誘騙受害者提供私密資訊，如個人資料、銀行詳細資料或帳戶登入憑證。犯罪分子接著會利用這些資訊從受害者的付款帳戶中榨取金錢、開立新的付款帳戶或進行詐欺性交易。這些行為可能會透過實體手段來完成，例如在信用卡詐騙案中，詐騙者會透過各種方式取得信用卡資訊，包括在實體銷售點使用卡號竊取軟體、使用社交工程手段或建立偽造網站。

第一部分

Google 防範網路詐騙和詐欺的措施

Google 的使命是彙整全球資訊，供大眾使用，使人人受惠。維護數十億使用者的線上安全是這項使命的核心。

這就是我們二十多年來持續積極對抗詐騙和詐欺威脅的理由。Google 將詐騙定義為「以謀取金錢和/或個人資訊為目的，針對個人或機構組織的欺騙性手段」。

Google 採用三管齊下的做法打擊詐騙與詐欺，分別是保護使用者免受傷害、提供可靠資訊，以及與各界聯手打造安全的網路環境。我們在所有產品中皆設有強大的技術與功能、政策以及程序，以保護使用者並提供可靠的資訊。這是一項需要持之以恆的挑戰，因為這些不肖分子不斷調整手段。在 2024年，所有線上平台的詐騙數量都顯著增加。面對日新月異的威脅，我們迅速更新政策、編組快速回應處置團隊，並且持續精進偵測技術。

1. 我們保護使用者免於威脅

Google 產品和服務中內建進階保護措施和政策，可防止發布詐欺性和詐騙內容、偵測和評估疑似違規內容，並對不肖分子和不當內容做出適當處置：

防範：產品的規範政策和內建安全功能是我們設立的第一道防線，可以遏阻或防止不肖分子發布詐欺性內容。

例如，不肖分子經常利用廣告來詐騙，因此我們制定了嚴謹的政策，盡可能避免使用者落入有害誤導性廣告的陷阱。其中一項政策是 Google Ads 不實陳述或行為政策，當中禁止廣告或到達頁面刻意遺漏相關產品資訊，或提供誤導性的產品、服務或業務資訊，意圖欺騙使用者。

偵測：雖然不肖分子濫用新技術來強化攻擊手段，但是我們的信任與安全團隊和安全專家則會運用 AI 和先進技術，來主動偵測有害內容和濫用行為，這讓我們可以與時俱進調整系統，早一步防範不斷進化的威脅。

應對：當有人標記了某項內容時，我們仰賴專人審查和 AI 技術來判斷內容是否違反我們的政策，並採取適當的因應措施。我們會為該服務採取適當行動，可能包括予以限制、移除內容、停用營利功能，或採取帳戶層級的行動，藉此減少日後的濫用情形。我們持續學習並改善做法：在對特定內容採取行動後，該內容會用於訓練我們的模型，以偵測類似的政策違規情形，並以此為參考制定新政策或開發新的產品功能，或者用於改進政策和功能。

2. 我們提供可靠資訊

我們提供可靠資訊，並讓使用者能夠透過一流工具評估資訊真偽，藉此提升使用者對我們產品和服務內容的信任。

3. 透過計畫與各方協作，並建立合作夥伴關係

打擊詐欺和詐騙行為是一項棘手的難題，需要跨產業攜手長期奮戰，並與政策制定者、專家、創作者、發布商、執法部門、企業主和一般民眾等利害關係人合作。

向大眾宣導知識也有助於打擊詐騙和詐欺行為。Google 積極與金融機構和監管單位合作，透過安全教育指南和工具包協助提升使用者素養。

大規模打擊不當行為

我們的使命是彙整全球資訊供大眾使用，使人人受惠。為此，確保資訊品質和審核內容是關鍵所在。不僅如此，我們還需要在資訊自由流通與社會責任之間取得絕妙平衡。Google 的產品、政策和違規處置決策都遵循我們的原則，體現出我們重視資訊開放和無障礙、尊重使用者選擇，並提供優秀的產品和服務，讓所有人都能受惠。

我們已制定以下政策和程序並開發相關工具，以保護使用者不受詐欺和詐騙侵擾：

- 使用限制政策：Google 的服務設有使用限制政策，其中明確禁止使用我們的服務從事任何詐騙行為。如果我們或使用者發現有人違反使用限制政策，使用者可以 (例如在 Gmail 上) 封鎖此類聯絡人，而我們可以採取行動處置侵權者、行為或內容。舉例來說，Google Ads 政策禁止欺騙行為。

- 檢舉活動：如果使用者懷疑或發現違反相關產品使用限制政策或 Google 一般服務條款的活動（例如分享垃圾或不當內容），可以向 Google 檢舉使用者違反政策。檢舉後，Google 會審查申訴，並可能會移除不當內容和停用違規帳戶。

- 封鎖功能：在我們的通訊工具中，使用者可以出於任何原因（包括傳送垃圾或不當內容）自行禁止其他使用者與他們聯絡。[1] 使用者可以在 Gmail、Chat 或訊息中封鎖即時訊息或使用者，禁止擾人的寄件者繼續與自己聯絡。廣告服務中也有類似機制，我的廣告中心提供控制選項，讓使用者只需按一下廣告，即可封鎖廣告並調整偏好設定，管理顯示的廣告類別。

- 自動偵測：Google 的系統會主動掃描服務，以偵測疑似違規內容。我們還會針對違反使用限制政策的潛在詐騙活動，採取違規處置。這包括主動將通訊內容歸類為垃圾內容，或在發現違反條款或政策的行為時，將違規者的服務或帳戶停權。Google 也會轉介犯罪案件至執法機關，協助將詐騙者繩之以法。

- 人工審查團隊：此外，我們設有人工審查團隊，負責審查使用者、優先檢舉人或機器學習系統檢舉的濫用情形。例如，如果使用者透過 Chat 使用者介面檢舉濫用行為，我們的人工審查團隊會審查所檢舉對話中最近傳送的 50 則訊息，並確認訊息傳送者的帳戶信號，以判斷是否有濫用行為。如果確定存在濫用行為，我們通常會針對濫用方停用此服務的帳戶。

- 驗證：我們使用保護隱私權的技術來驗證帳戶使用者是否為真人，並採用專門用於廣告主的驗證方法。廣告主身分驗證計畫的步驟包括驗證廣告主身分、提交關於業務營運的資訊，或提供必要的在地執照證明等等。

- 控制：我們盡可能在產品中部署技術解決方案和功能防範詐欺和其他網路危害。例如，Gmail 的 AI 輔助防護機制可以攔截超過 99.9% 的垃圾郵件、網路釣魚郵件和惡意軟體，防止這些內容進入收件匣，每天封鎖近 150 億封垃圾電子郵件。多重驗證選項是另一個我們推出的保護及控制的例子。

1 例如，請參閱 Google Chat「封鎖並檢舉使用者」指南，https://support.google.com/chat/answer/9277792?hl=en&co=GENIE.Platform%3DAndroid

Google 信任與安全團隊的應對

為了有策略地安排我們的內部工作，我們為整個信任與安全團隊制定了打擊詐欺和詐騙的三大重點領域方針：

行動者與行為重點領域著重於強化行動者層級與行為方面的違規處置，以及加強上游控制，以盡量降低詐騙得逞的風險。具體做法如下：(i) 改善身分與商家營運驗證程序，以防止詐騙者造成下游傷害；(ii) 確保新帳戶的功能與存取權受到限制，直到建立信任為止；(iii) 運用來自各種來源的線索和信號，針對不肖分子採取行動，以及/或盡可能在傷害發生並影響使用者之前便偵測出詐騙。

分類、政策與評估重點領域的目的在於加強我們的內部組織和策略，以打擊詐騙和詐欺行為。這包括在不同團隊之間採用一致定義、統合各種產品領域的處理程序、開發恰當的使用政策與作出修正，並且評估我們執行詐欺和詐騙處理政策的成效。

交流與合作重點領域的目的是與內外部專家及合作夥伴交流，以改善使用者保護措施並增進資訊分享，這包括改善使用者體驗中的安全功能，透過產業聯盟與其他數位平台和業界相關人士分享線索與信號，以及與執法機關展開營運方面的合作。

在我們「信任與安全」工作的這三大重點領域，以及我們持續開發全新改良技術及產品安全功能的工作之間，存在著重要的回饋機制。我們透過事件偵測及處置、與其他利害關係人和專家的交流，以及相關研究學到的內容，都會納入產品藍圖中供我們參考，讓我們可以開發或改良以安全為設計宗旨的功能，以持續改善使用者的線上體驗。

Google 打擊詐騙和詐欺行為上的產品保護措施

除了在政策方面對抗詐欺和詐騙的行動，我們還著重為每款產品設計內建防護措施。

我們盡可能努力確保所有產品都能自動保護使用者的上網安全，並讓使用者能夠輕鬆管理上網安全，選擇合適的防護等級。我們根據每項產品的特性量身打造安全防護措施，以確保這些措施最合適有效。

- Android 的行動裝置作業系統採用多層式保護措施以偵測和防範威脅，例如 Google 電話可以過濾掉疑似騷擾和詐騙號碼的來電。
- Google 訊息還透過 RCS 進階通訊解決方案技術為傳訊功能引入全新的保護機制。
- Chrome 透過多種功能保護使用者，包括 Google 安全瀏覽、Chrome 密碼防護警示、進階保護計畫 (APP)、以及安全瀏覽強化防護功能。例如， Google 安全瀏覽功能會在網站存在危險，並試圖以網路釣魚竊取使用者憑證時，向使用者發出警告。
- Google Ads 這項服務則已經針對廣告生態系統，制定一系列政策和安全防護功能並定期更新，例如假冒公眾人物身分的政策，或廣告放送量受限政策。這項措施要求廣告主必須經過「認識期」，才有權放送易遭濫用類型的廣告。這些特別制定的政策，就是大幅減少詐騙廣告數量的關鍵。
- YouTube 也受惠於 Google 的廣告安全措施，但更進一步地擴展針對詐騙的保護措施，藉由《社群規範》內容政策約束平台上分享的內容。根據欺騙行為及詐騙政策，YouTube 禁止使用者以影片、留言和中繼資料形式發布任何詐騙內容，例如提供現金贈禮、以快速致富為噱頭，或透過多層次傳銷模式，要求金字塔結構的下層會員付錢給上層會員，但未提供實質產品的內容。
- Gmail 採用 AI 技術強化篩選功能，一直以有效防範 垃圾郵件和網路釣魚攻擊而著稱。多年來，我們已因應手法更高明的網路釣魚活動做出調整，同時亦優先防範即時威脅使用者資料和憑證的網路釣魚攻擊。
- Google 購物會移除受到檢舉的不實評論，並利用眾多信號主動減少濫用行為和垃圾內容，例如透過自動審核、商家徽章和其他視覺提示，協助使用者識別優質商家，進而保護購物者。
- Google 搜尋的保護機制側重於排名和自動偵測，運用各種信號，並訓練我們的排名演算法來辨識品質不佳或可疑的網頁。
- Google Cloud 擁有專門的濫用防制團隊，致力於維護產品的正當使用，並防範我們的平台遭到濫用。為幫助客戶更有信心快速偵測並遏止加密貨幣挖礦攻擊，我們推出全新的加密貨幣挖礦防護計畫，提供最高 $100 萬美元，來補助由於未偵測到的加密貨幣挖礦攻擊，所帶來的未授權 Google Cloud 運算費用。這項計畫適用於 Security Command Center 進階方案客戶。
- 此外，我們也將這些措施擴展到全系列產品，例如付款服務和 Google 地圖。

憑藉這些產品保護措施，我們已收得顯著成果。
舉例來說：

- Gmail 會阻擋 99.9% 的垃圾郵件、惡意軟體和危險連結，保障使用者的收件匣安全。
- Google 訊息和手機應用程式內建來電顯示、垃圾內容防護和來電過濾功能，可封鎖不安全的來電，並警示可疑的來電者，進而協助防範語音釣魚和詐騙。
- Google 安全瀏覽功能協助使用者避免惡意網站的侵害，為超過 50 億部裝置提供自動保護。
- 廣告：我們在 2023 年封鎖或移除了 2 億 650 萬則違反「不實陳述或行為」政策的廣告 (其中包括許多詐騙和詐欺手法)，以及 2 億 7,340 萬則違反「金融服務」政策的廣告。我們封鎖或移除了超過 10 億則濫用廣告聯播網的違規廣告 (包括宣傳惡意軟體的廣告)，施以停權處置的廣告主帳戶則高達 1,270 萬個 (相較於前一年增加將近一倍)。人工智能在擴大這些措施並增強其有效性方面發揮了重要作用。到 2024 年第 4 季度為止，我們與多國政府合作推出多項計畫，例如將金融服務驗證計畫推廣至全球 17 個國家和地區。這項政策要求金融服務廣告商必須證明他們已獲得相關政府部門的授權，才能透過廣告宣傳其產品和服務。
- Android: 在 2024 年 2 月，我們與新加坡網路安全局合作，在 Google Play 安全防護中為所有當地 Android 裝置推出新的加強版詐騙防護功能。此功能透過封鎖可能具有風險的側載應用程式，為行動裝置使用者防範惡意軟體詐騙。在六個月內，該功能封鎖了超過 20 萬部裝置上近 90 萬次高風險的應用程式安裝作業。這些應用程式大多假冒熱門訊息應用程式、遊戲應用程式或電子商務應用程式，可能用於詐欺行為。

附錄 1 詳細說明整個 Google 主要產品領域範圍內實施的保護措施。

第二部分

共同打擊詐騙和詐欺：為社群提供政策建議

詐欺是整個社會的問題，需要各方長期密切合作。我們深知，如此大規模的問題無法單憑一己之力解決。

Google 致力於持續改進，以保護使用者並協助他們找到可信任的資訊。不過，我們深知，如此大規模的問題通常由跨國犯罪組織主導，無法單憑任何一家機構或公司解決。為了讓這些努力發揮最大功效，我們必須與其他相關方攜手合作，包括政策制定者、監管機構、民間團體和私部門。

我們提出以下三大領域措施，包括提供大方向的具體政策建議，協力推動打擊詐欺和詐騙的社群。

1. 攜手合作、資訊共享

利害關係人之間的多方合作是有效打擊線上詐欺和詐騙的關鍵。打擊詐騙是眾人之事，我們應在業界、政府、技術社群、學術和民間團體的利害關係人之間，促進新的合作與資訊共享模式。首先，我們需要加強對話和實務協調，包括讓大家共享情報訊號以協助偵測詐騙行為及背後犯罪者，以及促成執法機關合作。這些措施都需要 Google 等私人實體的參與，更牽涉跨政府合作。我們還需要在法律和法規方面，就安全、資料交換和國際反犯罪協作等方面，更加密切地合作。政策制定者應確保為這類資訊分享，預先規劃出必要的法律框架。

制定相關法律，促進政府和產業界之間的合作和資訊共享

線上平台和金融機構需要獲得法律授權，以便在發現可疑活動但證據不充分時，也能合作並針對詐騙者採取行動。這包括與政府機關相互分享私密資料等相關資訊，還有帳戶詳細資料等重要資訊，以便追蹤金錢流向，從而追蹤犯罪活動。

政策制定者應確保採取清楚明瞭且面面俱到的有效法律框架，實現這類資訊共享。大規模的詐騙防範屬於重要公眾利益，因此在當地管轄區已採取措施保障資料當事人基本權利和利益的前提下，若各平台能基於偵測和防範詐欺的限定目的來處理私密資料，這對防範詐騙相當重要。目前，許多管轄區仍不允許為支援犯罪調查，而與執法機關分享可做為行動依據的濫用行為訊號，阻礙了公私部門的合作。在某些管轄區，法律允許金融機構與當地執法機關分享這類私密資料，但這些法律保障範圍並未擴展至線上服務。

在其他司法管轄區，可能還需要澄清消費者安全和反詐欺法與資料保護和競爭規則等其他立法之間的交集。

政策制定者也應簡化提交和處理詐欺調查相關要求的法律程序，確保私人公司能夠按照正當程序，有效地配合主管機關的合法要求。在這方面，有多項自願性框架可供政府採用，例如《雲端法案》協議、《布達佩斯公約》(Budapest Convention)，以及最近的《布達佩斯公約第二附加議定書》(Second Additional Protocol to the Budapest Convention)，這份議定書建立了以尊重法治的方式處理跨境存取的機制。我們尤其鼓勵政府考慮簽署並核准第二附加議定書。

組成跨境政府工作小組，以協調和加強執法，促進互通性

為了讓執法機關和私人機構 (例如線上平台和金融機構) 可以跨境合作，打擊跨國組織性犯罪網路，*推動和加強跨境調查必不可少*。詐欺交易所涉及的犯罪者、受害者、重要文件和第三方通常跨境廣泛分散各地，如果單憑一個國家/地區的執法機關和其他相關政府機關，實在難以集齊所有必要資訊來偵測詐騙和詐欺，無法有效調查案件。

這類跨境合作的範例之一，就是經濟合作暨發展組織 (OECD) 在 2023 年通過的《跨境保護消費者免受詐欺和欺騙性商業行為的指導原則》，當中提出了境內法律和執法框架的具體增強措施，並提出了多種建議機制，主要用以促進跨境作業，以及公私部門之間的資訊通知、資訊共享、協助調查和機密性等。對於線上平台和金融機構等跨國企業而言，這種國際間的協調以及法律義務的互通性，將大幅提升法規遵循作業和私部門訴訟的效率和成效。

我們鼓勵政策制定者透過區域性合作或 OECD 等組織，嘗試實行類似的框架和協作模式，更有效偵測組織犯罪的活動網路，更全面將其繩之以法，並在未來持續提升，拓展至跨境偵測及執法行動。

支援全球反詐騙聯盟 (Global Anti-Scams Alliance) 等國際論壇

在反詐騙社群之間必須加強對話，包括偵測和分享關於新威脅、趨勢和模式的相關資訊，以及實際調查的合作方式。政策制定者應鼓勵線上平台和私部門 (尤其金融機構等高風險機構組織) 定期主持多方利害關係人對話。同樣地，政府應參與這些合作論壇或為其背書，以確保在整個生態系統內維持密切且持續的合作。這樣有助於主動防範、快速且有效地回應事件，還能幫助其他機構做好更充分的準備。

為了有效分享詐騙資訊，相關的利害關係人之間可透過不同通路合作，而全球反詐騙聯盟 (GASA) 正是範例之一。Google 已於 2024 年 3 月加入該聯盟，成為創始成員。GASA 是一個全球非營利聯盟，擁有來自各領域超過 100 位成員，並與 FBI、歐洲刑警組織等執法機關以及銀行業、大型科技公司建立了密切關係。該聯盟的工作小組和創新計畫，不僅包括詐騙訊號資料共享的工作流程，還提供有關惡意軟體和網路釣魚報告等指標。

Google 於 2024 年 10 月與全球反詐騙聯盟 (GASA) 以及 DNS Research Federation (DNS RF) 建立合作關係，推出全球訊號交換中心 (GSE)。GSE 是一個全球資訊交流所，用於交換關於線上詐騙和詐欺不肖分子的訊號，而 Google 是該中心的首位創始成員。這項合作聚集了 GASA 無與倫比的全球利害關係人網路、DNS Research Federation 已儲存超過 4,000 萬訊號的資料平台，以及 Google 在防範詐騙和詐欺方面的深厚專業知識。GSE 集結了我們的努力成果，建立了一個中央平台，並在整個網際網路內，以高效又易用的方式交換濫用行為訊號，填補了過往的空缺，能更快地橫跨各領域、平台、服務來辨識及打擊詐欺活動。

GASA GSE 平台專門為分享詐騙相關的洞察和情報而設，並為全球的利害關係人提供獨特的共同環境，讓大家攜手解決問題。因此，*包括政府機關和相關私人機構 (例如金融機構) 在內的其他各方，都不妨考慮加入此平台。*

我們也需要在整個生態系統內建立合作關係，來解決安全性事件。這個領域中有一項潛力十足的技術，就是風險與事件分享協調 (RISC)。RISC 最初是做為 Google 框架推出。每當 Google 偵測到使用者帳戶狀態有重大變更時，就會使用該框架與企業合作夥伴分享安全性事件。舉例來說，使用者的 Google 帳戶遭不肖分子入侵時，Google 可以向連結的應用程式和平台傳送訊號，以便其採取恰當的行動，或實施額外保障措施。我們鼓勵相關機構組織加入 RISC 框架。

2. 鼓勵整個社群採取行動

政府需要制定必要的公共政策和法律框架，讓社群成員能夠採取行動對抗詐欺性內容和不肖分子，並投入資源發展負責任的 AI 等各種新技術，以提升集體打擊詐騙和詐欺的能力。因此，政府的角色舉足輕重。

明確界定非法活動，加強執法機關的能力

為了讓執法機關和私人機構 (例如線上平台和金融機構) 能夠有效採取行動對抗詐欺和詐騙者，政策制定者必須明確闡明所在管轄區非法活動的定義。若缺乏明確的法律規範，公私部門實體可能會躊躇不定，難以做出果斷的決策或採取預防措施。

*因此，政府應審視現有的法律框架，評估是否有助於打擊詐騙和詐欺。*如果發現漏洞，政策制定者應採用全新或更全面的框架，或者發布輔助性指導方針，闡明應如何執行現有法律規範。這些法律框架應包括合理的執法機制，以及此情境下非法活動的明確定義，這點特別重要，因為數位平台無法掌握整個詐騙局勢，包括其他平台的情況或線下活動，因此往往難以判定遇到的是否為詐騙。

此外，讓執法機關接受充足訓練並培養相關能力，就能更有效地調查和起訴網路犯罪分子，這有助於加強執法力度，是政府應優先考量的事項。在多數情況下，私部門可以幫助執法機關培養相關能力。

採用「善心助人」的責任保護原則，激勵利害關係人採取預防措施

政府可採取「善心助人」責任保護原則，鼓勵各公司以預防措施防堵詐欺和詐騙。此原則能夠保護線上平台及其他中介商，免於因為積極防範網路危害而需承擔責任。

為了理解採用這項原則的重要性，可以想一想，如果缺乏這項保護原則，公司可能會面臨什麼後果。舉例來說，平台可能會因為決定移除某些內容而遭到起訴，例如 Google 移除仇恨言論、兒童不宜內容或金字塔騙局相關影片 (如上所述，大部分國家/地區並未將金字塔騙局視為非法詐騙活動)。公司一旦面臨這類壓力，就不會積極開發強大的內容審核系統。

務必確定責任制度清楚明確，不會使服務供應商卻步，阻礙他們主動自願地採取措施，防堵詐騙和其他有害內容。不該僅因為服務供應商秉持善意，自願採取自動化或非自動化處置措施，就認定他們應為代管的所有資訊承擔責任。「善心助人」責任保護原則能夠解決此問題。這項原則為平台提供保護，鼓勵他們找出和移除有害內容，即便過程中偶有失誤，也不用擔心遭追究責任。

例如，歐盟《數位服務法》(DSA) 和美國《通訊端正法》(CDA) 第 230 節等法律條文鼓勵採取預防措施，保護主動防堵網路危害的線上服務供應商，讓他們免於為此承擔責任。其他管轄區也採用了類似的責任保護原則，以免公司不願主動揪出詐欺性內容和不肖分子，或者分享威脅資訊。

投入資源發展負責任的 AI 技術，以及制定鼓勵技術創新的政策

*政策制定者應扮演好角色，確保立法框架允許開發反詐欺創新技術，以及打擊濫用合成媒體亂象所需的工具。*最新一波的 AI 創新能夠翻轉現狀，改變政府、企業和線上平台辨識詐騙者和處置詐欺性內容的方式。例如，我們可以探究如何運用 AI，在個人資料未經授權從網路上流出時，更有效地識別問題並遏止。Google 團隊正在積極採用這項革新技術，以更有效地保障使用者的上網安全。藉由採用 AI，我們可以加快偵測詐欺和詐騙，並擴大行動範圍，有效打擊平台上的濫用與有害行為。

為了充分發揮這些先進技術的潛力，需要政策制定者、公司和民間團體齊心協力，投入資源發展負責任的 AI 技術，包括發展必要的基礎架構、培養所需的人才以及制定合乎需要的法律框架，如我們在《AI Opportunity Agenda》(AI 發展契機大綱) 中所述。某些監管措施可能會阻礙世界各地的創新者和各國政府運用 AI 造福社會，例如經濟的發展、醫學的進步以及網路安全的強化，因此政府應鼓勵利害關係人在 AI 領域追求創新，並充分運用這項技術來打擊詐欺。

投入資源發展負責任的 AI 技術時，應開發識別機制，幫助使用者辨識意圖欺騙消費者的 AI 生成詐欺性內容。Google 為產品和服務制定了政策，明確規定我們如何處置經過操弄的詐欺性媒體內容，以及其他形式的有害 AI 生成內容。Google 不僅在內部展開合作，也與業界夥伴攜手，共同開發 SynthID 等工具，確保使用者能夠識別圖片或影片是否由 AI 生成。僅僅詢問「這是由 AI 生成的內容嗎？」並不足以評估內容的可信度，因此 Google 與全球頂尖資訊素養專家緊密合作，運用最新研究成果，協助確保 Google 產品提供足夠的資訊和工具，讓使用者可以交叉比對在網路上找到的內容。

在制定標準與最佳做法、整合技術工具和共享資訊等層面上，政府可以投入資源持續研究發展，並與機構組織合作，扮演解決合成媒體風險問題的重要角色。此外，政府做為技術的使用者也發揮關鍵作用，可以運用獨有優勢協助開發、採用中繼資料及浮水印等來源追溯技術，並進行相關壓力測試，最終鼓勵普及這些技術，同時保護使用者的隱私權和言論自由。例如，政府可鼓勵更多使用者採用內容出處暨真實性聯盟 (C2PA) 制定的「內容憑證」技術標準等來源追溯技術。C2PA 匯集了技術、廣告和廣播領域的機構組織，為發布商、創作者和消費者提供多種可靈活選用的方式，幫助瞭解各種媒體內容的真實性和出處。

3. 投入資源向使用者宣導並制定防範措施

安排大眾宣導活動，介紹網路詐騙並說明如何自保

向大眾宣導知識是有效應對詐欺的重要工具。若能與其他機構組織攜手、跨產業合作，最能發揮效用。提高大眾警覺一向是重要的公共政策工具，能夠有效對抗垃圾內容和網路釣魚等威脅，這項做法也依然是打擊詐騙和詐欺的關鍵，因為只要掌握詐騙相關知識，就不容易落入陷阱。

政府機關、消費者保護團體和企業都可以採取行動，幫助消費者瞭解如何在使用網路時保護自己，例如舉辦宣導活動，協助大眾瞭解網路詐騙以及自保的方法，例如 Google 參與推出的 scamspotter.org。該網站教育民眾認識常見詐騙手法 (例如網路釣魚、感情詐騙、投資詐騙) 和學會如何辨別警訊，同時鼓勵查核資訊，提醒民眾在網路上分享個人資訊時應保持謹慎。Google 與當地機構組織合作，攜手推動許多相關領域的計畫，例如在印度促成 DigiKavach，以及在新加坡推行 Project PRAISE。我們還在產品和網誌中定期發布建議，幫助使用者辨別詐騙，並瞭解遇到時該如何應對。

制定政策和開發產品功能，有效保護使用者

我們深知這份工作永無止境，因為詐騙者不斷試圖利用新的趨勢和技術來犯罪，同時擴大犯罪活動的影響範圍，並提升手段的複雜程度。正因如此，Google 主動追蹤這些發展動向，不斷更新使用限制政策，制定新的政策和開發新的產品功能，面對日新月異的威脅先一步做好準備 (如前面章節所詳述)。例如，採用了進階通訊解決方案 (RCS) 後，Google 訊息的安全保護更升級，如果其他通訊服務也能採用 RCS，就有助於大幅降低詐騙文字對話的出現。

我們鼓勵整個生態系統的服務提供方採用類似標準，並持續改進政策和加強產品內保護措施，以打擊詐騙。我們期待能與各方利害關係人合作，分享詐騙活動的洞察報告、攜手提高大眾意識、部署最新反詐騙技術等等。

為消費者提供清楚易用的檢舉管道

政府應為詐欺和詐騙受害者提供明確方便的官方檢舉管道。在某些情況下，這包括簡化現有或重複的檢舉管道，以及消除受害者遇到的障礙。明確的檢舉管道不僅方便受害者使用，還讓政府和執法機關可以掌握資訊，有效處理詐騙案件，必定可以提升打擊詐騙的效果。

我們也鼓勵相關專家加入 Google 優先檢舉人計畫，以提升向 Google 檢舉詐騙的成效。透過這項計畫，我們為參與機構提供專屬檢舉管道，當 Google 產品和服務出現疑似違反政策和社群規範的有害內容 (包括詐騙和詐欺) 時，這些機構就能通知我們。只要 NGO 和政府機關等機構組織具備足夠的專業能力，能夠識別並打擊網路危害行為，便適合加入這項計畫。優先檢舉人所回報的問題內容不會自動移除，在政策方面也不會受到任何特別處置，而且會比照使用者檢舉的內容，適用相同的審查標準和正當程序。不過請注意，由於優先檢舉人可信度及專業素養較高，Google 團隊會優先審查他們提交的檢舉內容。

追求安全的數位轉型

我們應鼓勵社會各界機構組織開發和部署安全的技術。Google 持續加強預防、偵測和因應詐騙的保護措施和程序。同樣地，線上平台、金融服務和電信業者也應在產品設計中融入安全性考量，並根據特定業務需求制定、定期更新和改進反詐欺政策。

在這個過程中，機構組織應運用最新的網路安全防禦技術並落實最新良好做法，例如新型 AI 工具，這些工具有望從根本上提高針對詐騙等網路威脅的防禦能力。為了因應日新月異的威脅，並跟上資安技術發展的腳步，最好的做法便是推動安全的數位轉型，特別是採用雲端式平台、新型作業系統和硬體以及零信任安全機制 (一種用於保護機構組織的安全性模型，原理是預設不信任所有使用者和裝置，即使這些使用者和裝置原先便在內部網站中也一樣)。

適用於所有機構組織的良好做法還包括：

- 保護帳戶，例如使用不重複的高強度密碼；
- 為 IT 管理採用良好機制，例如使用兩步驟驗證以及加入帳戶救援資訊；
- 為應用程式和網路瀏覽器部署自動更新；
- 清楚傳達企業的官方管道 (網站、客戶服務電話專線)，以防範詐騙者假冒身分；採用安全的付款系統；
- 使用密碼金鑰等方式，提升登入的方便性和安全性；
- 如要進一步查看企業資訊保護相關建議，請參閱我們的專用資訊資源。

結論

應對詐欺和詐騙需要整個社會各個層面的合作。在 Google，我們主動採取大規模行動，每天保護數十億人免受各種威脅的侵擾，檢查 50 億部裝置、2,000 億應用程式和 10 億組密碼，同時阻擋 1 億次網路釣魚攻擊和 150 億封垃圾郵件。

我們的核心使命是協助使用者找到可靠且具公信力的資訊。對於 Google 而言，履行這項使命完全符合我們自身的商業利益。我們的業務極度依賴網路生態系統的正常運作，以及使用者對此生態系統的持續信任。如果消費者因糟糕的線上體驗而放棄使用網路，Google 核心業務的長期生存狀態將會岌岌可危。這點適用於我們所有的產品，無論是協助使用者透過 Google 搜尋在開放式網路中探索，還是讓使用者在 Google 地圖中查詢當地商家。

資訊安全是一場片刻不能鬆懈的挑戰，因為不肖分子不僅會以大規模運作，還會不斷調整手段，並結合線上和離線活動以規避偵測機制。因此，即使部署最嚴格的防禦機制，系統也難免有不盡完善之處。

我們致力於持續改進，以保護使用者並協助他們找到可信任的資訊；此外持續採行靈活策略很重要，不斷追蹤不肖分子的行為並從中汲取經驗。唯有如此，才能預做準備，妥善因應未來可能出現的詐欺和詐騙。

唯有攜手產官學界，共同合作，方能強化防禦，使影響更加深遠。

附件1

人工智慧與詐騙的關係

二十多年來，Google 一直致力以負責任的方式開發 AI。長久以來，我們在履行內容責任方面也一直居於領導地位，推出採用生成式 AI 技術的新產品時，也維持相同的信念與方法。AI 可以成為有力的打擊詐欺和詐騙幫手：採用這項技術，我們可以加速偵測濫用行為、採取行動，並擴大行動範圍，有效打擊平台上的濫用與傷害行為：

a) 防範有害內容

我們以負責任的態度建構生成式 AI，並運用 AI 技術大規模偵測和移除有害與非法內容，採用著重於防範 - 偵測 - 處置三大核心的方法：

(i) 防範：我們已制定生成式 AI 使用限制政策，當中概述了我們禁止發布的有害、不當、具有誤導性或非法內容。 這包括產生用於欺騙或詐欺活動、詐騙、網路釣魚、惡意軟體的內容，或旨在誤導的內容。這些政策得益於我們遍布全球多個國家/地區、精通多種語言的審查團隊，這些團隊 24 小時全年無休，謹慎評估遭標記的內容。

我們採用廣泛的分類器系統來偵測和防範違反生成式 AI 產品政策的內容。系統篩選訓練數據，排除高風險內容，確保訓練品質。此外，分類器也會分析使用者輸入的提示，以及模型產生的潛在結果。模型產生多個候選回應後，分類器會依據安全性等參數進行評分。所有程序皆在背景快速流暢地完成，使用者不會察覺。

如果發現違規提示或輸出內容，我們的產品將不會提供回覆。我們也可能會引導使用者參閱其他資源 (例如求助專線)，以尋求有關敏感主題 (例如與危險行為或自傷有關的主題) 的協助。

(ii) 偵測：近期大型語言模型 (LLMs) 的發展，大幅提升了我們在平台及服務中偵測違規內容的速度與規模。透過大型語言模型，不用耗費數週甚至數月，我們在幾天內就能快速建構並訓練模型，以找出產品上特定類型的不當內容。面對一定規模但需謹慎處理的領域 (例如偵測線上仿冒商品)，或是新興濫用領域，這特別有幫助。我們可以建立模型原型，精確識別這類新型態濫用行為，並自動轉交給我們的團隊進行違規處置。

例如，Google 信任與安全團隊持續運用大型語言模型，以有效加強偵測將快速致富當成噱頭的廣告，這些廣告違反了我們針對不可靠的財務聲明的政策。投放這類廣告的不肖分子相當狡詐多端，他們會調整手法，依據新型態的金融服務或產品 (例如投資建議或數位貨幣) 量身打造廣告，以欺騙使用者上當。金融趨勢瞬息萬變，增加了區分合法與詐騙服務的難度，進而影響我們擴展機器學習技術和自動化違規處置系統打擊詐騙的能力。大型語言模型更能快速找出金融服務的新趨勢、識別濫用這些趨勢的不肖分子的模式，以及區分正當業務與快速致富騙局。這項技術提升了 Google 團隊的靈活度，讓我們面對各種新興威脅時更能輕鬆應對。

我們在資料生命週期的各階段使用 AI，透過偵測資訊提升資料豐富度，從非結構化資料擷取重要細節，將資料標準化管理並分類，優先處理期間產生的結果，並提供主動防禦新興威脅所需的情報。

結合 Mandiant 和 Google Cloud 等 Google 系列服務的高精度分析和掌握全球情報的能力，我們能夠識別攻擊，同時深入瞭解不肖分子的手段、技術和程序 (TTP)。收集到的資料不僅成為我們威脅情報產品的基礎，還讓我們能夠使用優質範例來訓練 AI 模型，進而提高偵測準確度並改善結果，帶動良性循環。

我們仍在打擊濫用行為領域測試這些新技術，到目前為止，這些技術已交出亮眼成績，將大力推動內容審核的發展及大規模主動保護使用者的工作，尤其是針對新興風險的保護。

(iii) 處置：雖然 AI 是大規模識別潛在違規內容的重要工具，但專人審查仍然是整個過程中重要的一環。內容經標記並送交審查後，就會交由人工審查員評估教育、紀實、科學、藝術或新聞性質的背景資訊和細微差異。

建立或分享內容的背景資訊是評估內容品質或用途的重要因素。Google 擁有超過 20,000 名來自世界各地經過培訓的審查人員，負責偵測、審查和移除數十種語言的違反政策內容。這些審查人員具備必要的當地背景知識和專業技能。

結合自動化系統快速、大規模運作的能力，問題的內容通常會在廣泛流傳或讓使用者看到之前就移除，進而減少人工審查員接觸到的有害內容數量。

b) 幫助用戶識別人工智慧生成的內容

我們致力為使用者提供優質資訊，在整個社會創作者和使用者之間維繫信任。我們肩負的其中一項責任，就是為使用者提供更先進的工具來辨識 AI 生成內容。越來越多人運用人工智慧創作內容，因此我們也不斷努力開發新工具和制定新政策，來幫助目標對象辨識 AI 生成內容：

- 運用 SynthID 加上浮水印

  我們的 SynthID 工具包可識別 AI 生成內容並加上浮水印。這些工具將數位浮水印直接嵌入 AI 生成的圖像、音訊、文字或影片中。人眼難以察覺 SynthID 浮水印，但在各模態中系統都可以偵測並識別 AI 生成內容。此工具包目前為 Beta 版，未來仍會持續改善，並且會整合入越來越多的 Google 產品，協助使用者和機構以負責任的方式使用 AI 生成內容。儘管 SynthID 無法直接阻止網路攻擊者或駭客等不肖分子的惡行，但是能增加難度，讓 AI 生成內容更難被濫用於惡意用途。

- YouTube 上的內容標籤

  生成式 AI 正在顛覆 YouTube 上的創作方式，而我們的做法需要與時俱進，繼續為創作者和觀眾維護透明健康的生態系統。因此我們推出了一款工具，要求創作者揭露經實質變造或合成的逼真內容。現有的上傳流程中已加入這款工具，方便創作者輕鬆加入揭露聲明。我們會套上資訊公開標籤，來提示使用者他們正在觀看這類內容。大部分影片標籤會顯示於展開後的說明。但如果影片涉及較敏感的主題，例如：健康、時事、選舉或金融，我們還會在影片本身加上更醒目的標籤。

- 搜尋功能中的「關於這張圖片」

  我們在 2023 年推出了「關於這張圖片」新功能，方便使用者檢查網路上圖片的可信度和掌握背景資訊，包括圖片是否有 SynthID 浮水印。現在使用者可以在 Android 手機上，透過「畫圈搜尋」更直覺地使用這項功能。如有中繼資料，使用者還可以查看圖片創作者和發布商在圖片中加入的資料，包括可能反映圖片是由 AI 生成或修飾的欄位。在 Google 所有 AI 生成圖像的原始檔案中，我們都加入了這類標記。

- 選舉廣告揭露聲明

  Google 於 2023 年領先所有科技公司，率先要求若放送廣告經數位變造或含有由 AI 等工具生成的仿真合成內容，選舉廣告主應一律醒目揭露這類資訊。

c) 聯手打造安全網路環境

Google 明白，要負責任地發展 AI 技術，需要研究人員、社會科學家、業界專家、政府、創作者、發布商，以及在日常生活中使用 AI 的大眾共同努力。我們已建立合作機制，並運用我們的專業知識開發和分享工具，幫助其他組織在平台上偵測和移除不當和有害內容。這包括深偽內容資料集，可協助其他機構偵測以 AI 操弄的內容，並以負責任的方式打造 AI 技術。

單憑一間公司的力量，無法推動負責任的 AI 做法。我們必須攜手合作才能取得進展。因此，我們致力聯手業界、社會和學術機構，以達成目標。

- C2PA：我們最近加入了內容出處與真實性聯盟 (C2PA) 的指導委員會。C2PA 是跨產業合作的成果，目的在於幫助大眾更清楚瞭解 AI 生成內容及掌握相關背景資訊。
- 前沿模型論壇：Google 與 Anthropic、Microsoft 和 OpenAI 共同成立了前沿模型論壇，旨在確保以安全又負責任的方式開發前沿 AI 模型。
- Partnership on AI：我們加入了 PAI，這是一個專家社群，致力於推動以負責任的做法開發 AI 技術、建立和分享使用生成式 AI 製作的媒體內容。
- MLCommons：我們是 MLCommons 的成員，這個聯盟的宗旨是加速機器學習創新，並擴大對社會的正面影響。

其他計畫包括在 2024 年 2 月推出的 AI 網路防禦計畫，目的在於利用人工智慧 (AI) 提升網路安全，並扭轉「防禦者困境」。這項計畫將包括舉辦一系列網路安全研討會，以及推動各項實際解決方案研究專案，探討威脅偵測、惡意軟體分析、安全漏洞偵測與修復及事件應變等議題。除此之外，我們還開放了 Magika 的原始碼。這是一個全新 AI 工具，可協助防禦者識別檔案類型，這是偵測惡意軟體不可或缺的一環，因為惡意軟體是詐騙和詐欺的主要工具。這些計畫潛力十足。舉例來說，Gmail 已使用 RETVec，這是一種以神經網路為基礎的多語言文字處理模型，將垃圾郵件偵測率提高了 40%。

除了自行研究，我們也委託或協助其他機構進行研究，以更有效地瞭解和對抗詐騙。例如，在 2024 年，Google 和雅加達的策略與國際研究中心 (Centre for Strategic and International Studies, CSIS) 為了進行錯誤資訊相關研究而合作設立的「更安全的網際網路研究室」(Safer Internet Lab, SAIL)，推出了全新工作專案，研究以東南亞地區為目標的詐騙者如何操弄合成媒體內容、這對大眾的影響，以及可能的解決方案。

濫用 AI 功能

對於深偽技術用於詐欺的疑慮反映出實際的問題，隨著 AI 工具日益普及，這些工具也可能會遭到濫用，以推動各種惡意活動。新型生成式 AI 工具能夠以空前的速度和規模製作內容，包括文字、圖片和語音。儘管這些工具可以激發創造力和促進理解，但也可能用來製作對個人和社會有害的內容。此類技術未來可能會大幅加劇惡意攻擊行動，即使不肖分子資源和能力有限，也能輕易帶來威脅。

不過，到目前為止，根據我們的觀察和開放原始碼帳戶，不肖分子採用 AI 和有效運用的情況仍然有限，主要以社交工程為主，但在未來，我們可以預見，資源與能力有限的不肖分子可能會濫用 AI 來大規模製作並散布更讓人難辨真假的內容。他們可能會使用各種模型來撰寫文章或製作政治卡通等內容，以傳達特定觀點，或生成看似正當的填充式內容，讓虛假的人物角色更有可信度。

比起過去沒有使用 AI 技術偽造的內容，超逼真的 AI 生成內容更具說服力，更容易讓目標對象上當。我們還已經看到證據顯示確有其事，例如使用合成媒體 (深偽技術) 重現可信人物的圖像，以此宣傳詐騙內容。

濫用 AI 技術還可提高詐欺性內容的品質，進而提升「可信度」。藉由使用大型語言模型，威脅發動者無需瞭解目標對象使用的語言，就能量身打造更具說服力的內容。不僅如此，還有利於不肖分子取得貼近人類平常說話模式的文字內容，進而掌握更有用的材料，進一步發動網路釣魚，並得以初步入侵。

要負責任地採用這項技術，非常重要的一點就是預測和測試各種安全風險，包括這些新型 AI 生成媒體的興起。儘管這項技術可以應用在許多領域，有效發揮作用，例如，為語言機能障礙或閱讀障礙者開創新的可能性，或啟發全球藝術家和電影工作室構思出新點子，發揮無限創意，但我們也意識到，它也可能會遭到濫用，淪為詐欺或詐騙等惡意用途的工具。

我們與多個相關單位合作，致力於三大重點領域，以降低「深偽」等技術的風險：

- *偵測*：運用機器學習技術來開發系統，以大規模偵測合成和/或經過操弄的媒體內容。
- *出處*：為了達成提供可靠資訊的目標，讓使用者能夠瞭解線上內容的記錄，我們使用驗證鏈 (例如可保證內容從相機擷取後不曾遭到竄改的數位簽章) 或提供有關內容線上記錄的背景資訊 (例如協助使用者瞭解圖片或影片過去在網路上哪些其他情境出現過)。
- *媒體素養*：協助讓社會大眾在面對圖片和影片時更謹慎，不會直接當成某件事確實有發生的有力證明，而且更清楚可以根據哪些情境元素判斷內容具有可信度。

善用 AI 技術並解決相關濫用問題的後續步驟

Google 多年來一直致力於對抗 AI 濫用問題。舉例來說，我們一直在研究和開發工具以對抗「合成媒體」(深偽技術) 的濫用，而且越來越常與網際網路相關利害關係人合作，以分享應對這些威脅的最佳實務與資訊。 政策制定者在這個方面可以發揮作用，確保立法框架允許反詐欺創新和開發必要工具，以打擊濫用合成媒體的亂象。

我們也樂觀的看著 AI 能真正展現能力，成為「防禦者」重要的助力，幫助對抗詐騙與詐欺等惡意行為。藉由防禦措施的改進，包括採用生成式 AI，安全社群有機會比威脅發動者更早採取行動，讓從業人員和使用者受惠。

我們深信 AI 可以在打擊詐欺與詐騙方面發揮顯著的正面影響，正如大型語言模型在 2023 年提升廣告安全方面所展現的成效。我們已經在應對偽造內容方面取得進展：例如，多虧了我們不斷演進的機器學習演算法，我們今年在 Google 地圖上封鎖或移除了超過 1.7 億則違反政策的評論，這個數字比 2022 年增加了 45% 以上。我們還移除或封鎖了超過 1,200 萬筆偽造的商家資料。

附件2

我們如何支援使用者

附件 2.1 使用者如何檢舉詐欺與詐騙行為

消費者可使用 Google 產品檢舉不當內容和行為。舉例來說，使用者現在可以更輕鬆地檢舉他們認為違反我們政策的廣告，只需點選直接在廣告介面中顯示的「關於這則廣告」選單，即可提出檢舉。廣告遭檢舉時，我們會進行審查，確認是否符合我們的政策，並視需要撤下廣告。

使用 Google 購物時，如果使用者看到不對勁或不符合我們政策的內容，例如超高價格或暴力武器，可以透過產品頁面右下方的「檢舉產品資訊」連結或我們的說明中心向我們檢舉。

在產品內提供此類體驗至關重要，以便讓使用者能盡可能輕鬆操作。

使用者受到威脅時如何檢舉 ：

以下是 Gmail 遇到網路釣魚攻擊的範例

附件 2.2 賦能使用者

賦能使用者是應對詐騙和詐欺威脅的主要方法。我們透過內建的產品保護功能以及產品內資訊來達成此目的，這包括產品內的揭露聲明以及其他對使用者有用的資訊，協助他們做出明智且理性的決策，例如廣告資訊公開中心、我的廣告中心、廣告主身分驗證、金融服務廣告主認證等。

此外，我們認為要賦能使用者，特別重要的一點就是提高使用者的意識，讓他們瞭解可能面臨的風險以及學習如何自保。分析證實提高使用者意識非常重要，因為使用者通常難以辨別詐騙和真實互動：三分之一的使用者表示曾受到誘人的優惠吸引，另外三分之一的使用者則表示，無法辨別欺騙行為或缺乏識別詐騙的知識 (根據 GASA 全球與亞太地區問卷調查，2023年)。

我們致力於透過多種管道，例如平面媒體、全國性廣告宣傳以及產品內警示，提升使用者防範詐欺的意識，這些工作需要涵蓋所有年齡層和社群，因為所有人都可能成為詐騙者的目標。

我們也意識到，單憑傳統宣導教育，可能不足以提升大眾對詐騙和詐欺的抵抗力：單純提供更多資訊不一定能改變使用者的行為。為此，我們一直在試行所謂的「預防接種」方法，協助使用者建立對詐騙的「心理抗體」。事實上，最近一項針對美國受眾的研究顯示，透過「預防接種」形式的干預措施，可以大幅提高使用者對冒用他人身分詐騙的辨別能力，而且不會損害他們對正常通訊的信任 (資料來源：Robb & Wendel, 2023)。

我們自己也舉辦了許多提高意識的宣導活動，但發現與其他機構合作特別有效，不但有助於擴大防範宣導範圍，還能更貼近當地社群的需要。我們的行銷團隊已開發出一套使用者宣導工具包。我們希望透過自家管道，以及與金融機構和業界機構等對象的合作，讓這套工具包可以幫助更多人。

以下列舉我們在宣導教育和提高使用者意識方面的幾項合作：

在泰國，我們與泰國銀行及泰國銀行業電腦緊急應變小組 (Bank of Thailand and Thailand Banking Sector Computer Emergency Response Team, TB-CERT) 合作推出 #31days31tips 線上安全宣傳活動。在活動期間，每天都會分享一則安全建議，協助泰國民眾更有效地保護自己的線上帳戶、識別金融詐騙並保障私人資訊的安全。

在澳洲，我們在過去一年與澳洲消費者網路 ACCAN 合作，推動禮物卡詐騙防範活動 (透過廣告和 YouTube 內容宣傳)。我們還與澳洲競爭與消費者委員會 (ACCC) Scamwatch 團隊合作，每年舉辦「反詐騙意識週」，具體行動包括推廣安全檢查 (Scamwatch 是 Google Workspace 優先檢舉人計畫的成員，可透過專屬管道向我們檢舉違反政策的行為)。除了「反詐騙意識週」外，我們每年還會在 2 到 3 次活動中推廣安全檢查。共有約 1,100 萬名澳洲人觀看這些宣導內容，而且觀眾參與相當踴躍。

在新加坡，我們承諾支持 Project PRAISE，這是一項與 RSVP Singapore The Organisation of Senior Volunteers 及新加坡警方合作推出的計畫，旨在訓練一群志工，透過一系列詐騙防範工作坊，針對特別容易成為網路攻擊目標的銀髮族，協助提高他們對詐騙的意識。

在香港，我們支援香港社會服務聯會推動「做個智慧網樂人」計畫，幫助提高學生的數位素養和網路安全意識。我們分享了保護線上帳戶安全的實用訣竅，包括啟用兩步驟驗證、在搜尋時使用「安全搜尋」和「安全瀏覽」提升安全保障，以及在 Gmail 和 Chrome 中啟用由 AI 技術支援的反網路釣魚功能。

印度：與此同時，Google 的慈善計畫 Google.org 發放新的 $400 萬美元援助金給 CyberPeace Foundation，用於在印度推動一項為期四年的全國意識提升計畫，並建立一個多語數位資源中心，協助近 4,000 萬名弱勢者提升抵禦錯誤資訊的能力。這筆資金是繼「Digi Kavach」重大計畫之後的額外援助。

亞太地區：過去五年來，Google.org 資助亞太地區 26 間具有社會影響力的機構，捐助的金額超過 $3,500 萬美元。這些金援有助於教導弱勢群體保障網路安全、打擊錯誤資訊，並提高機構的網路韌性。

在美國，Google 與 Cybercrime Support Network 合作，聯手找出詐騙者最常見的犯案模式，並在 scamspotter.org 網站上提供實用建議，幫助使用者阻止詐騙者的惡行。

使用者援助資源：

- Google 網誌：如何辨別詐騙以及遇到時該如何應對
- 防範並檢舉詐騙行為 - Google 說明
- 行動裝置安全機制與隱私權 - Android 安全中心
- Google Play 安全防護 - Android
- 運用 Google Play 安全防護保障應用程式安全和個人資料隱私
- 瞭解 Pixel 的 Android 安全性認證
- Android 生態系統安全性資訊公開報告
- 安全中心 - 緊急救援 - Android

附件3

Google 各產品打擊詐騙的對策

附件3.1 Android

數十年或甚至數百年以來，詐騙者的手段不斷演進，從信件詐騙轉向利用電話和簡訊來犯罪，並逐漸加入線上通訊工具。因此，我們致力於開發一系列產品功能，希望降低使用者在使用 Google 支援的 Android 行動裝置時遭受詐欺和詐騙的風險。

Android 是一款全球數十億部裝置仰賴的作業系統 (OS)。它就像是讓手機可以正常運作的軟體。這個行動作業系統以修改版 Linux 核心系統及其他開放原始碼軟體為基礎，主要用於配備觸控螢幕的行動裝置，例如智慧型手機和平板電腦。

Android 的首要之務是確保使用者的安全，我們非常重視這項職責。我們採用業界領先的安全做法，並與整個生態系統的開發人員和裝置製造商密切合作，以確保從開啟裝置的那一刻起，使用者就能受到保護。

讓最多使用者安全使用裝置，其他平台無法比擬：Android 已應用於超過 24,000 多種不同類型的行動裝置，而且有 30 億部使用者裝置已啟用 Google Play 安全防護功能。這項功能不僅針對從 Play 商店下載的應用程式提供惡意軟體保護，還針對第三方商店/網站提供相同的安全防護。

Android 符合全球最嚴格的安全標準：我們已獲得最高標準的行動產業認證，包括美國國防部和 31 個國家/地區認可的共通準則。

Android 的安全防護方針著重於三大核心要素：

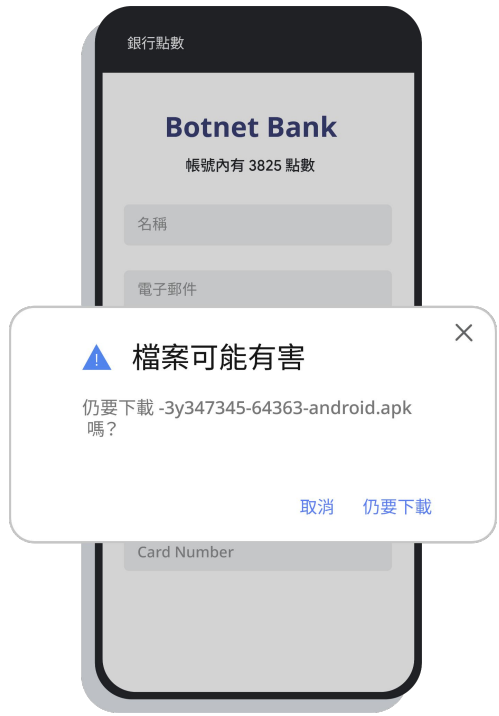
- 多層式：Android 生態系統的每個部分搭配與協作，共同打造出運作流暢又有效的強大防護機制。
- 資訊公開：我們與安全研究社群合作，共同發掘、修復安全問題並確認解決方案是否有效。問題解決後，我們便會與全世界分享，以確保資訊公開並協助他人。
- 跨 Google 服務的技術：我們運用 Google 的資安專業知識，將領先的安全防護功能整合至 Android OS、Play 商店以及裝置上的應用程式。

我們的團隊致力於打擊詐欺，尤其專注於打擊透過電子郵件、電話和訊息應用程式等管道對受害者遠端詐欺的行為。犯罪分子利用各種攻擊向量實施犯罪計畫，包括散布惡意軟體、濫用權限、分享螢幕畫面以及包括網路釣魚在內的各種社交工程策略。我們在核心作業系統中內建防護功能來防禦這些攻擊向量，並附加額外的安全服務，持續掃描裝置以偵測惡意軟體和其他有害行為。

附件 3.1（一）Android 手機和訊息服務

針對手機和訊息服務中的主要通訊功能，Android 採用多層保護措施，具體包括：

- Google 電話內建來電顯示、垃圾內容防護和來電過濾功能，可封鎖不安全的來電，並警示可疑的來電者，進而協助防範語音釣魚和詐騙。
- Google 訊息內建詐騙與網路釣魚防護功能，可警示使用者並自動過濾可疑的垃圾內容與不安全的網站。這個功能運用 AI 技術評估寄件者的信譽、確認是否出現已知模式及危險連結，藉此找出可疑訊息。
- Chrome 下載警示可在使用者即將下載 Android (APK) 檔案時發出警告，確保使用者知曉連結即將觸發應用程式下載。
- 每部 Pixel 裝置都具有來電顯示和垃圾內容防護功能。 此外，我們還透過「訊息」應用程式的「已驗證簡訊」功能，在訊息串中顯示商家名稱和標誌以及驗證徽章，協助使用者識別潛在詐騙。
- 我們也正著手運用 AI 工具，進一步強化 Android 的詐騙電話偵測功能。



進階通訊套件 (RCS)：針對簡訊面臨的問題進行改善，特別是強化防範垃圾內容與詐騙的能力

傳統的手機簡訊 (又稱為 SMS，即 Short Messaging Service 的簡稱) 不僅是一種古老過時的通訊方式，也是一個有缺陷的系統，尤其是在防範網路釣魚和詐騙偵測方面。由於簡訊網路呈分散式，因此沒有人可以辨別哪些網路可信，哪些不可信。這樣一來，從訊息建立到使用者接收的過程中，都沒有任何信任機制，導致簡訊詐欺解決方案成效有限，這也是為什麼使用者會收到大量企圖詐騙和進行網路釣魚攻擊的簡訊。

打擊這些濫用行為媒介、防範垃圾內容和濫用行為的方法有很多，包括結合防火牆封鎖來源可疑的訊息、要求電信業者/集結網站簡訊平台掃描訊息內容，以及制定商業條款以規定連接的集結網站和互連的電信營運方合法營運。然而，由於網路拓撲結構分散，這些方法無法以足夠一致的方式嚴格實施，因而無法大規模用於防範詐欺，亦無法靈活快速地應對新威脅。

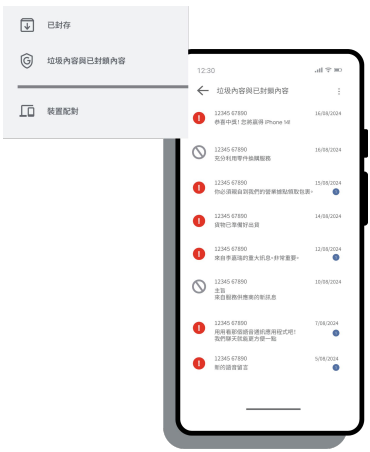
進階通訊解決方案 (RCS) 的新訊息系統可提供更有效的詐欺偵測：在電信平台的推動下，RCS 已發展成為一個集中式系統，讓我們能夠以一致的方式實施這些方法，迅速部署新的反制措施來應對新威脅。這種融合式系統讓單一平台也能驗證使用者、監控流量和辨識可疑流量模式。舉例來說，如有新使用者傳送多個國際訊息，這些訊息很可能是詐騙內容，系統便會標示這些訊息，以助使用者辨別。

很重要的一點是，RCS 促成了一個更安全的訊息架構。在這個架構中，企業傳送給消費者的訊息經過驗證，而人際訊息傳送則在平台端和用戶端設有廣泛的惡意訊息偵測和封鎖功能。

RCS 結合經過驗證的商家訊息和人際訊息保護機制，透過以下方式大幅降低了目前普遍存在於簡訊中的詐騙和網路釣魚攻擊數量：

*1. 人際訊息詐騙偵測：*RCS 可偵測異常流量，並在流量傳送到目標位置前予以封鎖

- 集中式的平台可驗證新用戶及其信譽。
- 集中式的平台可封鎖傳輸中的訊息，並暫停有傳送垃圾內容行為的違規電話號碼。
- 集中式的平台可根據使用者的信譽和行為，向連結的用戶端提供垃圾內容通知。
- 平台通知和裝置端防護機制可能會觸發訊息自動封鎖，或顯示警告橫幅。
- RCS 還讓可信賴的實體可以將垃圾簡訊歸類至垃圾內容資料夾，防範這些內容進入使用者的收件匣。
- 在 Google 訊息應用程式中，詐騙偵測功能可用於簡訊和 RCS 訊息。



每月，我們在各簡訊、多媒體訊息和 RCS 等管道中，針對超過 20 億則疑似有害的訊息，顯示警告橫幅並自動封鎖。

2. Verified Business Messaging - RCS 可驗證商家身分，讓使用者可以信任通訊對象

- 每年商家向使用者傳送超過 2 兆則訊息，佔全球簡訊流量的 30% 以上。
- 詐騙者會冒充商家傳送訊息，使用者經常受騙上當，因此能夠確認商家的真實性非常重要。
- RCS Business Messaging (RBM) 要求商家必須通過電信服務供應商驗證。
- 完成驗證後，商家將獲得「勾號」驗證標誌，這增強了使用者對訊息傳送者的信心和信任。
- 使用者可以認明這個標誌，判斷商家資料真假 (區分出未經驗證的短碼簡訊訊息、長串號碼的簡訊或 RCS 訊息，以及合法商家傳送的訊息)。
- RCS/RBM 功能提供品牌驗證、上游企業 (代理) 驗證、內容核准、主動流量管理工具和更豐富的互動體驗，有望改變商家與顧客之間的訊息溝通方式，建立更穩固、更信賴的關係。

簡訊

無可見驗證機制

無商家資訊

平台政策執行分散

RBM

商家驗證

商家資訊

平台政策執行具有一致性

簡單來說，RCS 可為使用者提供更安全有效的訊息體驗。簡訊技術以過時的網路為基礎，導致不肖分子得以濫用系統。相較之下，RCS 以安全的 IP 數據連線為基礎，作為融合式系統，在詐騙和網路釣魚偵測方面皆大有改進。RCS 還是主要使用者安全功能的基礎，例如垃圾內容篩選器和商家驗證。未來，採用 RCS 標準將大幅減少網路釣魚攻擊和詐騙訊息。

附件 3.1（二）Android 和 Google Play

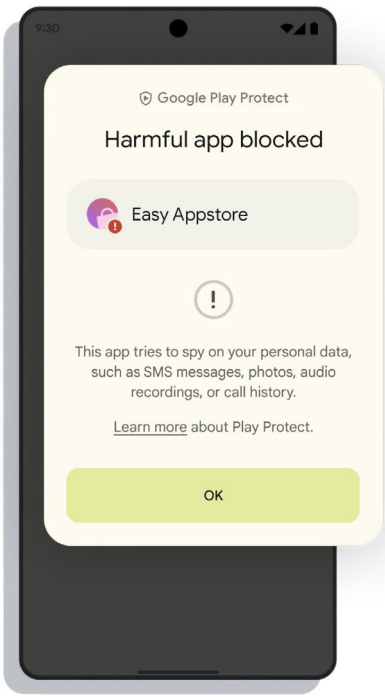
Google Play 網路商店可供使用者發掘喜愛的應用程式、遊戲、電影、電視節目和書籍等精彩內容，為全球數十億 Android 裝置使用者提供超過兩百萬款應用程式和遊戲。

Google Play 是全球領先的行動應用程式商店，因此成為詐騙者利用的對象，例如他們會提供來自第三方網站的偽造應用程式，並聲稱出自合法來源，以引誘使用者下載，或特意建構應用程式，企圖感染使用者的裝置及竊取資料。

為了規避偵測機制，網路犯罪分子現在會利用精密的技術，例如可改變識別特徵的多型態惡意軟體。他們轉而利用社交工程手段，誘騙使用者做出有危險性的行為，例如提供機密資訊或從臨時來源下載惡意應用程式 (最常見的情況是透過連結下載)。為了應對不斷演變的作案手法，我們持續投入心力強化 Android 生態系統的安全與防護。

防範：技術與產品保護措施

我們持續努力提升 Android 的安全與防護，同時為各大機構提供工具與支援，協助強化官方應用程式的安全性，使這些應用程式更能有效應對惡意軟體攻擊，例如側載應用程式的攻擊。以下是我們已部署的幾項重要保護措施。



Play Integrity API - 應用程式存取風險：對金融機構而言，提升官方應用程式的安全性特別重要。例如在巴西、印度、新加坡和泰國，Google 目前協助銀行和金融科技公司加強官方應用程式的安全性，提升應對惡意應用程式攻擊的能力。我們邀請各大銀行與金融科技開發人員加入搶先體驗計畫，將 Play Integrity API 應用程式存取風險功能整合至自家應用程式。開發人員加入這項搶先體驗計畫後，可將應用程式存取風險信號新增至 API 回應，該回應中還已包含裝置完整性、應用程式完整性和帳戶授權判定結果資訊。根據應用程式存取風險信號，他們可以偵測裝置上的其他應用程式能否用於存取或控制自家應用程式。

Google Play 安全防護：Google Play 商店中的所有應用程式都必須先經過嚴謹的安全性測試，才能獲准上架。我們的機器學習系統每天都會掃描多達 2,000 億個應用程式，持續在幕後確保裝置、資料和應用程式安全無虞，不受惡意軟體和垃圾軟體侵擾，它是全球使用範圍最廣的行動威脅防護服務。

我們持續改進機器學習系統和審查流程。在 2023 年，我們阻止 228 萬個違反政策規定的應用程式在 Google Play 上發布，其中一部分歸功於我們不斷推出新的改良安全防護功能、政策更新，以及先進的機器學習和應用程式審查流程。我們還持續努力避免開發人員建立惡意和詐騙應用程式。在 2023 年，我們出於違規行為 (應用程式確認為惡意軟體和多次嚴重違反政策等) 原因禁止 333,000 個惡意帳戶在 Play 上架應用程式 (詳情請參閱我們的網誌，瞭解我們如何打擊不肖分子以及如何保護 Android 使用者免受詐欺)。

2023 年 10 月，我們宣布進一步強化 Google Play 安全防護的安全功能，新增程式碼層級的即時掃描，以應對新型態的惡意應用程式。截至 2024 年 2 月，在這項即時掃描強化功能的協助下，Google Play 安全防護已識別出 515,000 個新型態的惡意應用程式，而且針對這類應用程式的警示或封鎖次數超過 310 萬。

高風險權限是 Google 發現經常遭到詐騙者利用的裝置設定權限，例如閱讀或傳送簡訊，以及存取控管。系統會標記向使用者請求這些權限的應用程式，以提醒使用者注意，甚至可能會在高風險情況下直接封鎖應用程式。

打擊側載式金融詐騙 (在新加坡進行前測計畫)：根據我們對主要詐欺性惡意軟體系列的分析，我們發現超過 95% 私密資訊權限 (RECEIVE_SMS、READ_SMS、BIND_Notifications 和無障礙權限) 漏洞攻擊行為透過 Android 裝置上的側載應用程式進行。側載是指使用者透過獨立網站下載應用程式，而非預先載入的應用程式商店。由於 Google 已投注大量心力強化 Play 商店的安全與防護，因此詐騙者越來越難透過 Play 商店發布的應用程式實施詐欺行為。

因此，為了進一步保護消費者免受金融詐欺和詐騙侵擾，Google 於 2024 年開始與新加坡政府 (MCI & CSA) 合作推出全球第一個前測計畫，以打擊 Android 上的網際網路側載金融詐欺和詐騙應用程式。這項合作讓我們能更即時地識別並移除潛在的詐欺性應用程式，並攜手增加使用者宣導，協助大眾更有效自保，免於受到不肖分子欺騙。這項前測計畫是 Google 現有 Play 安全防護功能的延伸，可協助保護消費者免受詐欺性和惡意應用程式的侵害。每次識別出應用程式，Play 安全防護的偵測能力都會不斷提升，讓我們得以加強對整個 Android 生態系統的保護。

Android 撥號應用程式針對訊息和撥號功能提供額外保護，可在使用者可能受騙時提供通知。此外，我們將推出一項針對 Play 商店的掃描計畫，專門找出針對老年人的詐騙應用程式。

Play 商店政府徽章：我們近期致力保障 Google Play 的安全，Google Play 於 2024 年推出的全新徽章就是其中一項措施，旨在協助使用者識別全球官方聯邦及州政府應用程式。由於政府應用程式通常要求使用者提供高度機密的資料，因此經常成為被冒用身分的目標，讓不肖分子能夠乘機竊取身分資料並進行財務詐欺。為經過驗證的政府應用程式加上徽章十分重要，有助為使用者提供安全、高品質、實用且相關的體驗，而我們也一直在探索各種方法不斷改進。

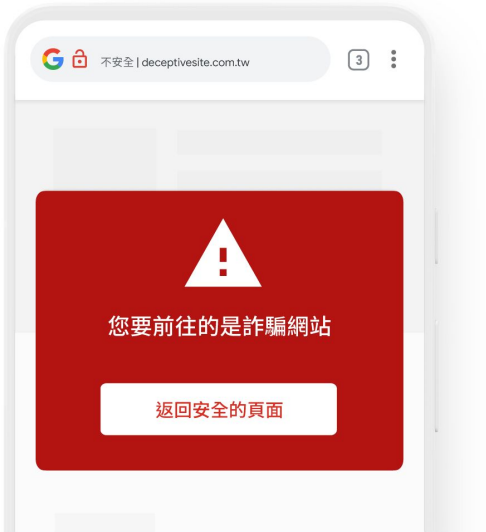
偵測及處置

針對 Play 應用程式，我們量身打造了一系列使用政策，其中一項目標是確保金融服務應用程式不會讓使用者接觸到欺騙性或有害的金融商品和服務。因此，我們已建立隱私權政策，完整揭露個人及敏感使用者資料的存取、收集、使用與分享方式，並遵守這項政策所列出的限制。

個人信貸權限：2023 年 4 月，我們更新了個人信貸政策，明確規定提供或促成個人信貸的應用程式不得存取使用者的私密資料，例如聯絡人或相片。

附件3.2 Chrome

Chrome 是 Google 的官方網路瀏覽器，不僅速度飛快、安全可靠，使用者也能視需求自訂設定。Chrome 預設採用安全設定：系統每四週自動更新一次，以確保安全防護功能和修正程式能夠保護使用者遠離危險和詐騙網站，並在憑證遭盜用時發出警示。我們還採用安全瀏覽技術，當使用者嘗試存取我們判定具有危險性的網站時，瀏覽器便會顯示警示。



防範

安全瀏覽：Google 安全瀏覽功能於 2005 年推出，目前已保護全球超過 50 億部裝置，面臨看似正當的惡意連結時，能為使用者提供額外防護。

如果某個網站看起來具危險性，並試圖騙取使用者的憑證，Google 安全瀏覽功能會向使用者發出警示。在收到警示後，使用者只需點選「返回安全的頁面」選項，即可避免前往惡意網站或下載惡意檔案。我們還更新了機器學習模型，以專門識別看似常見登入頁面的網頁，以及包含魚叉式網路釣魚信號的訊息。

Google 將這項技術免費提供給其他瀏覽器和網際網路公司使用。除 Chrome 外，這項技術也部署在 Firefox 和 Safari 等多款競爭對手的瀏覽器中，並在 iOS 和 Android 等不同平台上使用。

每當 Google Pixel 手機使用者嘗試前往危險網站或下載危險檔案時，安全瀏覽技術都會顯示停止訊號警示，避免 Pixel 裝置受到網路釣魚攻擊。如果網站遭到惡意人士入侵，安全瀏覽功能還會通知網站管理員，協助他們診斷及解決問題。

我們正在部署安全瀏覽的商用功能，以進一步推廣我們打擊詐騙的做法。Web Risk 讓使用者可偵測網址是否違反任何安全瀏覽政策、評估網址的風險程度，並將網址提交至 Google Cloud 進行掃描。為了保護超過 50 億部裝置，如果網址違反任何安全瀏覽政策，系統會在幾分鐘內將違規網址加入安全瀏覽封鎖清單中。這項服務提供一系列獨特功能，可保護使用者免受惡意攻擊侵擾，例如網路釣魚和惡意軟體。隨著我們與全球各地的政府機關合作，Web Risk 將整合至新加坡政府網路安全運營中心 (Government Cyber Security Operations Centre, GCSOC) 主導的網路安全生態系統。 GCSOC 使用自有的自動化工具 (例如 PhishMonSG)，每天從商業動態消息和 ScamShield 等群眾推動的計畫中，彙整數千個潛在惡意網路釣魚網站。經過簡化和標準化後，他們會使用 Google Web Risk Evaluate API 評估可疑網站的風險等級。只有判定為具有風險的網站才會提報給 Web Risk 的 Submission API 進行封鎖。

Chrome 的其他保護措施包括我們開發的各種工具，讓使用者可根據自身的威脅模型和感知風險採取更多措施，例如瀏覽器上的 Chrome 密碼防護警示擴充功能、透過安全金鑰提供 Google 最強大帳戶防護機制的進階保護計畫 (APP)，以及安全瀏覽強化防護功能。

- Chrome 密碼防護警示：我們在 Chrome 瀏覽器中提供擴充功能。只要使用者啟用，就會在他們於Google 以外網站中輸入 Google 憑證時發出警示 (使用者可能誤以為他們是在真正的 Google 網站中輸入憑證，實則為網路釣魚網站)。這項關於異常登入的警示有助於告知使用者，可能有不肖分子正試圖竊取他們的帳戶憑證，並即時提供如何減輕危害的建議。密碼防護警示還會檢查使用者造訪的每個頁面是否假冒 Google 登入頁面。若有此情形，便會發出警示。

- 進階保護：2017 年，我們推出了進階保護計畫 (APP)，這個計畫提供 Google 最強大的帳戶保護機制。APP 要求使用安全金鑰，這些金鑰可有效防範中間人攻擊。在這類攻擊中，使用雙重驗證碼可能會遭網路釣魚攔截。進階保護計畫在每次下載前都會執行更嚴密的檢查作業，以抵禦針對 Chrome 的線上攻擊，同時還為 Gmail 和雲端硬碟等其他 Google 產品提供額外防護機制。

- 安全瀏覽強化防護功能：2020 年，我們在 Chrome 中推出了安全瀏覽強化防護功能，為希望在瀏覽網路時獲得更高層級安全防護的使用者提供新選擇。使用者開啟此功能後，Chrome 會直接與安全瀏覽功能分享額外安全資料，提升威脅評估精確度。例如，Chrome 會即時檢查不常見的網址，以偵測使用者即將造訪的網站是否可能是網路釣魚網站。這項功能採用更進階的偵測技術，可快速因應惡意活動的演變調整。結果顯示，比起使用標準防護的使用者，使用強化防護的使用者遭到網路釣魚攻擊的可能性降低了 20-35%。

附件3.3 Google Ads

數十億使用者透過 Google 尋找正確資訊，我們仰賴使用者的信任，才能經營業務，但詐騙行為會破壞這種信任。廣告贊助的網際網路環境為廣告主、發布商和使用者建立了價值交換。各種規模的企業都能以更經濟實惠的價格投放廣告，並放送個人化廣告，透過成效評估資料調整行動，加速成長。發布商和內容創作者可透過隨內容放送的廣告賺取收益，並運用成效評估資料來證明廣告的成效，進而盡可能提高收益。所有這些都讓使用者能以極低或免費的方式存取大量內容和資訊，例如新聞、影片、社群媒體、地圖和路線。廣告的安全性和完整性是這個生態系統的基礎，讓這一切能夠成真。在 Google，我們不僅致力於維護健康的生態系統和開放式網路，這對我們的業務也至關重要。我們將保障使用者的安全視為首要之務，因此在政策的執行上投注大量心力。我們由數千人組成的團隊全天候分工，致力於大規模訂定並實行各項政策。

詐騙不斷進化。近幾年來，我們發現越來越多不肖分子運用精密的詐騙手段，例如利用偽裝技術來躲避偵測，或放送電話詐騙廣告，在現實生活中佈下陷阱。自 COVID-19 疫情爆發以來，我們發現投機型廣告和詐欺性行為增加，這些不肖分子試圖誤導使用者，或以不正當的方式透過廣告賺取收益。我們還觀察到一些不肖分子跨越國界從事詐騙。

生成式 AI 的出現同時帶來了機會和風險。好的廣告主會運用生成式 AI 改善廣告活動，製作出越來越多吸引人的廣告素材和到達網頁。然而，同樣的技術也可能為不肖分子所用，因此得以更快速製作出更精美的廣告，並擴大觸及範圍。我們也在運用生成式 AI 技術，提升偵測與違規處置能力。

為了解決這些難題，我們致力於創新和改善政策與產品。我們正投注大量心力改進政策、進階機器學習和大型語言模型 (LLM) 以及專人審查程序。我們的系統綜合考量網路信號、帳戶活動記錄、行為模式和使用者意見回饋。此外，我們還投資相關技術，在面對犯罪分子們籌畫的惡意行為時，可以強化偵測的能力，將不同帳戶中的蛛絲馬跡串聯起來，一次將多個不肖分子停權。最後，我們持續投入資源進行廣告主身分驗證，這也是我們落實廣告資訊公開及「融入安全性考量的設計」架構的重要基礎。

偵測及處置：制定並實施完善的廣告政策

我們透過結合 AI 與人工檢舉和評估，偵測詐欺和詐騙廣告。這個程序可確保 Google 平台上的廣告遵守我們制定的嚴格政策，包括禁止不實陳述或行為和助長欺騙行為的政策。此外，我們還為使用者提供簡便的詐騙廣告檢舉途徑。

我們會定期審查並更新政策，以確保 Google 聯播網中的使用者、廣告主和發布商都受到保護。舉例來說，在 COVID-19 疫情期間，我們執行敏感事件政策來防範惡意行為，例如針對手部消毒液、口罩和紙類產品等高需求產品哄抬價格，或宣傳不實療法的廣告。隨著大眾對病毒有進一步的瞭解，衛生機構也陸續發布新指南，我們的政策執行策略也有所調整，開始允許醫療服務提供者、衛生機構、各地政府機關和可信賴的企業提供重要新資訊和具公信力的內容，但仍謹慎防範投機分子從事不當行為。最近，在 2023 年末到 2024 年初這段期間，我們面臨了一系列廣告活動濫用行為，不肖人士常利用深偽技術，在廣告中放上公眾人物的假肖像，以此詐騙使用者。我們一偵測到這類威脅，便成立專責團隊迅速因應。我們先找出不肖人士的行為模式，然後訓練自動違規處置模型偵測類似內容，並開始大規模下架這些廣告。此外，我們也修訂了不實陳述或行為政策，以利迅速將不肖分子的帳戶停權。

我們的安全團隊已長期使用 AI 機器學習技術來大規模實施政策。多年來，在這套系統的輔助下，我們得以偵測並直接封鎖數十億惡質廣告，防止使用者看到這些廣告。儘管這套系統十分精密，但這些機器學習模型仍須不斷地進行大範圍訓練。我們必須提供數十萬甚至數百萬件違規內容案例，這套模型才能持續精進。

大型語言模型 (LLM) 就不一樣了。大型語言模型能迅速檢視並解讀大量內容，並從龐大的內容中擷取出重要的細微差異。運用這類先進的推理技術，不但能夠處理更複雜的政策，而且規模及精準度更勝以往，有助我們更快進行處置。以針對不可靠的財務聲明的政策為例，這項政策禁止在廣告中以快速致富為噱頭。投放這類廣告的不肖分子越來越狡詐多端，他們會調整手法，依據新型態的金融服務或產品 (例如投資建議或數位貨幣) 量身打造廣告，以欺騙使用者上當。

可以肯定的是，傳統機器學習模型經過訓練後能夠偵測這些違反政策的行為。然而，金融趨勢步調快速且不斷變化，有時讓我們難以區分正當與虛假服務，也不太能夠快速擴展我們的自動化違規處置系統，以打擊詐騙。大型語言模型更能快速找出金融服務的新趨勢、識別濫用這些趨勢的不肖分子的模式，以及區分正當業務與快速致富騙局。這項技術提升了 Google 團隊的靈活度，讓我們面對各種新興威脅時更能輕鬆應對。

為了協助大家瞭解我們在廣告安全方面的工作規模，在此舉例說明：2023 年，我們封鎖並移除了 55 億個惡質廣告，將違反政策的 1,270 多萬個廣告主帳戶停權 (這個數字遠高於 2022 年的 670 萬)。尤其在詐騙防範方面，我們封鎖/移除了 2 億 650 萬個違反不實陳述或行為政策的廣告 (涉及許多詐騙手段)，以及 2 億 7,340 萬個違反金融服務政策的廣告。我們還封鎖或移除了超過 10 億個濫用廣告聯播網的違規廣告 (包括宣傳惡意軟體的廣告)。[2]

2 廣告安全報告，2003, https://blog.google/products/ads-commerce/google-ads-safety-report-2023/#enforcement

廣告主身分驗證

為了替使用者打造安全又可靠的廣告生態系統，Google 要求廣告主完成一或多項驗證計畫。廣告主驗證計畫可能包括多個步驟：

- 商家資訊：在廣告主驗證計畫的第一個步驟，Google 會透過「商家檔案」部分，要求廣告主提供幾項有關 Google Ads 帳戶和商家的基本資料，這些資訊有助於 Google 進一步瞭解商家。例如，我們可能會詢問貴商家是否為廣告代理商、廣告費用的支付者、宣傳的產品或服務是由貴商家或其他商家提供，以及所在的特定產業。

- 驗證身分：填寫完「商家檔案」的資料後，商家可能需要透過廣告主身分驗證或商家營運驗證程序，進行法定名稱驗證。這項驗證程序必須由授權代表完成，即 Google Ads 帳戶和/或廣告費用付款資料的管理員。

- 驗證業務營運：根據商家在「商家檔案」提供的資料，我們可能會請商家提供佐證文件，以驗證商家營運詳情 (如適用)，例如商業模式、商業登記資訊、服務類型、商業行為，以及商家與所宣傳的品牌、產品或第三方之間的關係 (如適用)。

Google 驗證廣告主資訊後，會在我的廣告中心揭露廣告主的名稱和地理位置，使用者只要點選廣告，就能查看這些資訊。使用者也可以從我的廣告中心點選前往廣告資訊公開中心，即可搜尋並查看更多資訊，包括廣告主在 Google 聯播網放送的所有廣告。

我們目前在超過 240 個國家和地區主動驗證廣告主身分。如果廣告主在收到提示後沒有在期限內完成驗證計畫程序，帳戶會自動受到限制，無法放送廣告。在高風險地區，我們設置了更嚴格的驗證程序。舉例來說，如要在特定國家/地區宣傳金融服務，廣告主需要完成進一步的驗證程序，對大多數廣告主而言，他們需要證明已取得當地監管機構授權，可透過廣告宣傳自家產品和服務。截至 2024 年第 4 季，我們已在 17 個國家/地區推出金融服務廣告主驗證計畫。

廣告主身分驗證也提供了重要信號，有助於開發其他以安全為設計宗旨的產品功能。例如，在 2023 年 11 月，我們推出了廣告放送量受限政策，針對我們較不熟悉的廣告主，在高風險地區設下廣告觸及範圍限制，藉此保護使用者安全。廣告放送量受限政策針對投放高風險廣告的廣告主採用「瞭解期」機制，而且僅適用於特定廣告放送情境。在這些情境下，只有合格廣告主可以在沒有曝光限制的情況下放送廣告。

假設有人想預訂下一趟行程，搜尋喜愛的航空公司飛往舊金山的航班。採行這項新做法後，系統顯示與使用者搜尋相關的廣告中，絕大多數來自下方的廣告主：航空公司、航空公司的競爭對手、該地區的飯店，以及遵循政策且資訊公開的其他廣告主。根據這項政策規定，如果廣告主沒有良好行為記錄，廣告曝光次數可能會因而受限，直到在我們的平台上有良好記錄為止。這項政策讓使用者有機會與相關且實用的廣告互動，但如果廣告主未證實有良好記錄且投放的廣告具有誤導性或混淆不清，此政策會減少使用者看到此類廣告的可能性。

資訊透明、使用者選擇與控管的重要性

維護使用者對 Google 產品及服務的信任，是我們的第一要務。針對在網路上看到的廣告，我們希望使用者能掌握充分資訊再做出決定。使用者信任 Google 平台廣告主，我們才有可能為每個人打造切合需求的網路體驗。因此，我們致力於維持資訊透明，揭示廣告主的身分、所在地和透過 Google 投放的廣告素材。我們長期以來致力於維持廣告資訊透明：

- 自 2011 年起，我們開始發布年度廣告安全報告，詳細說明我們為避免自家廣告平台遭人惡意使用所採取的行動。
- 2018 年，從美國開始，針對希望在我們平台上投放選舉廣告的廣告主，我們開始要求他們必須完成驗證程序，並在廣告內揭露聲明中明確指出廣告出資者身分。
- 2020 年，廣告主身分驗證的推出讓使用者可進一步掌握具體情況，協助他們深入瞭解特定廣告背後的公司，並能夠辨認生態系統中可信的廣告主，同時限制了不肖分子做出不實陳述或行為。
- 2022 年，我們推出了我的廣告中心，讓使用者可進一步控管在 Google 網站和應用程式上的廣告體驗。在我的廣告中心中，使用者可以封鎖敏感廣告，並進一步瞭解用於個人化廣告體驗的資訊。
- 2023 年，我們宣布推出廣告政策中心，匯總所有已驗證廣告主並開放搜尋，使用者可以在當中檢視廣告商的基本資訊，並查看廣告商在我們平台上放送的其他廣告。

對於試圖入侵我們防禦系統的攻擊者，Google 並不陌生，而且總會有更狡詐的不肖分子試圖操控或規避我們的系統。我們希望讓這些濫用 Google Ads 產品的不肖分子，不但需要面對重重阻礙，還需耗費高昂成本。透過政策變更和驗證要求來增加成本和阻力，就能幫助達到這個目的。

我們發現，在能夠部署可靠廣告主身分驗證機制的全球高風險地區，濫用情形已有所減少。另外值得注意的是，驗證只是多層防禦機制的一部分，並非所有情況皆適用的單一解決方案。我們還有多項執行程序和政策，可與驗證機制搭配。我們持續投入資源開發新程序，在詐騙者進駐我們的平台前就先發制人，同時審查和更新平台政策，以應對持續演進的詐騙手段。

附件3.4 YouTube

YouTube 是使用者上傳和分享影片的影音平台，而我們的首要目標是堅守職責。YouTube 的任何行動都會以此為準，包括對詐騙的處理方式。使用者每天都會在 YouTube 上尋找資訊和教育資源以豐富生活，包括學習金融知識。

Google 的廣告安全工作延伸至我們所有的產品：Google Ads 金融商品和服務政策適用於所有 Google廣告，包括 YouTube 上的廣告。但是，YouTube 針對詐騙的保護措施並不限於平台上的廣告：YouTube 的《社群規範》也禁止詐騙行為。例如，在使用者在平台上分享的任何內容中 (包括影片、影片說明和其他使用者發布的評論)，立下誇大不實的快速致富承諾，以及宣傳多層次傳銷計畫和現金贈禮計畫。

YouTube 向來會在《社群規範》明列禁止發布的內容類型。設計這些社群規範是為了在確保社群安全的同時，讓使用者能夠自由、開放地交流意見。

YouTube 的社群規範禁止使用者發布鼓吹危險或非法活動的內容。以下是幾項與應對詐騙相關的社群規範：

- 垃圾內容、欺騙行為與詐騙處理政策：YouTube 嚴禁發布垃圾內容，或以詐騙和其他欺騙行為誘騙 YouTube 使用者。我們禁止發布提供現金贈禮、以快速致富為噱頭或透過多層次傳銷模式，要求金字塔結構的下層會員付錢給上層會員，但未提供實質產品的內容。另外，如果內容的主要目的為誘導觀眾離開 YouTube 並前往其他網站(包括詐騙網站)，我們也一概禁止。相關政策包括外部連結處理政策 (規範安裝惡意軟體或促成網路釣魚的網站或應用程式連結)，以及我們的垃圾內容影片處理政策。
- 有害與危險內容處理政策：禁止內容宣傳仿冒商品、非法銷售 (例如銀行帳戶密碼或竊取而來的信用卡) 和數位安全 (網路釣魚、入侵行為等) 等產品和服務。使用者頻道可能會上述原因而遭到終止。
- 冒用他人身分處理政策：根據這項政策，我們禁止意圖冒用他人身分或頻道的內容。
- 不實資訊處理政策：YouTube 禁止可能造成重大危害的特定誤導性或欺詐內容，包括運用技術修改或竄改內容 (通常已經超出斷章取義的範圍) 來誤導使用者，因而可能造成重大危害。

我們會採取行動，盡快移除違反政策的內容，並結合人工作業與機器學習技術，大規模偵測和處置違規內容。對於屢次違規或情節重大的單一濫用事件，我們可能會終止使用者的頻道或帳戶。

機器學習技術讓我們能夠主動識別和標記有害內容，以交由人工審查員審核，某些情況下，也能自動移除與之前移除過的內容十分相似的特定內容，例如垃圾內容。因此，我們可以在違規內容廣泛流傳之前採取行動。

這種方法成效顯著。2024 年 4 月至 6 月，我們的違規影片收視率為 0.9% 至 0.11%。也就是說，在每 10,000 次的 YouTube 內容觀看次數中，只有 0 到 11 次是觀看違反《社群規範》的內容。在同一期間內，YouTube 移除了超過 840 萬部違反《社群規範》的影片，並終止了超過 320 萬個違反《社群規範》的頻道。絕大多數的頻道都是因為違反垃圾內容政策而遭到終止。如果想瞭解詳情，請參閱每季發布的 YouTube 社群規範違規處置報告。

營利: 創作者必須達到更高的門檻，才能藉由 YouTube 合作夥伴計畫，透過 YouTube 上的內容賺取收益。過去幾年，YouTube 持續加強規範營利資格，以防止發布垃圾內容、冒用他人身分和從事其他不當行為的不肖分子利用平台營利，並危害創作者生態系統。例如，要符合加入 YouTube 合作夥伴計畫以透過內容營利的資格，頻道的觀看時間和訂閱者人數都需要滿足資格標準。創作者申請加入計畫後，若屢次違反 YouTube 的內容和營利政策，我們也會暫停相關權益。YouTube 會審查頻道，以確定是否違反營利、內容和著作權政策。詳情請參閱 YouTube 合作夥伴計畫總覽與資格條件。

YouTube Shopping

YouTube Shopping 是我們推出的一套功能，目標是在 YouTube 建立更真實、可信且愉快的購物體驗3。創作者可運用 YouTube Shopping，標記和宣傳自家產品或其他品牌的產品，並輕鬆追蹤收益和成效。

YouTube 已訂立標準來規範哪些類型的影片和創作者適用 YouTube Shopping 功能 (包括推出了專門的 YouTube Shopping 聯盟計畫)，當中列有明確的資格條件，這有助於為創作者、品牌和觀眾建立一個更負責任的平台。根據我們的負責任做法方針，如果內容包含著作權聲明、主要觀眾為兒童、或創作者有未解除的《社群規範》警告 (未解除的《社群規範》警告也許會重新審查)，那麼觀眾將不會看到產品標記。

YouTube Shopping 設有一系列政策、條款及細則，以解決許多消費者保護相關問題，但我們將繼續與政策制定者合作，為消費者、創作者和品牌找出最佳政策解決方案。雖然 YouTube Shopping 致力於為觀眾提供順暢的體驗，但目前交易不是在 YouTube 上進行。除了我們訂立的資格條件外，許多與我們緊密合作的夥伴也各自有平台使用條款及細則。舉例來說，政策制定者可能會制定關於以下情事的規則：虛假評論和網紅揭露聲明、詐欺性交易和仿冒商品、退款和退貨、詐欺性設計模式以及隱藏性費用。

觀眾通常會信任在 YouTube 上追蹤的創作者分享的購物推薦。YouTube 創作者的優質資訊和真誠分享，讓消費者更有信心，可以更聰明購物。

3 YouTube Shopping 指南, https://support.google.com/youtube/answer/12257682?hl=zh-Hant&sjid=13923454110448904427-NC

附件3.5 Gmail and Workspace 應用程式套裝組合

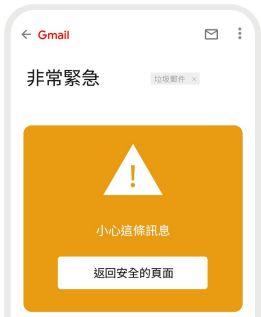
自 2004 年推出以來，Google 的電子郵件服務 Gmail 便以保護使用者免於垃圾郵件和網路釣魚攻擊而聞名，而這都歸功於採用 AI 技術的篩選功能。類似的安全措施已擴展至整個 Google Workspace 應用程式套裝組合，例如 Google 文件、Google 表單等。

垃圾郵件、網路釣魚和惡意軟體仍然是所有電郵使用者面臨的嚴重威脅。我們已因應手法更高明的網路釣魚活動調整，同時也優先防範對使用者資料和憑證最具即時威脅性的網路釣魚攻擊。

- Gmail 每天可防堵 99.9% 的危險電子郵件，讓使用者不受侵擾，這包括含有垃圾內容、網路釣魚連結或有害惡意軟體的電子郵件。
- 我們在 Gmail 防堵的惡意文件中，有 63% 是每天都會變化。
- 目前 Gmail 封鎖的網路釣魚電子郵件中，有 68% 是從未見過的新型態。

在 COVID-19 疫情期間，垃圾郵件與網路釣魚攻擊大幅增加。我們便加倍努力開發內建防護功能。例如，Gmail 惡意軟體掃描工具每週處理超過 3,000 億個附件，封鎖有害內容。

機器學習技術協助我們識別 Gmail 中超過 95% 的垃圾郵件和網路釣魚活動。在這個領域中，更多的資料可提升我們對網際網路使用者的保護能力。我們在這個領域不斷改進技術，讓使用者收件匣免受許多帳戶盜用活動 (包括網路釣魚活動) 侵擾。



2023 年，我們在產品內加入技術保護措施，開始要求傳送至 Gmail 地址的電子郵件必須經過特定形式的驗證。結果顯示，Gmail 使用者收到的未經驗證郵件數量大幅下降了 65%，這不僅讓收件匣免於雜亂，還更精確地封鎖數十億封惡意郵件。自 2024 年初，我們將要求傳送大量郵件的 Gmail 寄件者驗證電子郵件、提供使用者輕鬆取消訂閱，並將遭檢舉的垃圾郵件數量控制在門檻以下。

此外，Google 的威脅分析小組是專責的安全專家團隊，可進一步偵測、防範及緩解政府資助的威脅行動。這個團隊會定期在季度公告和特別報告中分享主要研究結果。

當我們認為使用者可能成為政府資助的網路釣魚攻擊目標時，Google 會持續向使用者發出警示。自 2012 年以來，我們一直在發出此類警示，包括提供建議，告訴使用者如何提升 Google 帳戶安全性。

我們已建立新系統來偵測可疑的電子郵件附件，並將這類附件提交至安全瀏覽服務以進行進一步檢查 (請參閱上方關於 Chrome 的部分)。這項措施可保護所有 Gmail 使用者 (包括 Enterprise Workspace 客戶)，避免遭到附件中可能隱藏的惡意軟體攻擊，不肖分子可能會將這些惡意軟體當做詐騙工具，例如用來竊取使用者的登入資訊和其他個人識別資料。

附件3.6 Google 搜尋

Google 重視使用者的每一筆搜尋。因此，無論使用者何時使用 Google 搜尋功能尋找實用的相關資訊，我們都會確保提供最優質的搜尋結果。然而，網路上仍不乏干擾行為，包括詐騙和其他威脅線上使用者安全的行為。

有一些線上詐騙和詐欺會影響搜尋結果，例如詐欺網站冒充商家或服務供應商，誘騙使用者向錯誤對象支付款項。

例如，許多詐騙者假冒提供熱門服務和產品的客戶服務專線，誤導撥打電話的使用者透過銀行轉帳或禮物卡付款，他們會使用所謂的「濫填關鍵字」方法、仿冒品牌的標誌，並提供希望使用者撥打的電話號碼，建立大量劣質網站。這些網站可能會誘騙使用者揭露敏感的個人資訊，導致使用者損失錢財，或使裝置感染惡意軟體。這種詐騙行為通常稱為「客戶服務詐騙」或「技術支援詐騙」，目前已有數十萬名使用者通報遇到這類詐騙，每起案件都可能造成數百美元落入詐騙者手中。

我們打擊詐欺和詐騙工作的基礎是：

- 透過研究發展出可規模化的自動化解決方案，以及
- 從過去數年來收集的具體案例汲取經驗。例如，根據使用者透過 Google 詐騙和網路釣魚檢舉工具提交的網路釣魚攻擊和詐騙行為，我們找出最常見的詐騙手法，然後開發解決方案，並採取行動來處置。

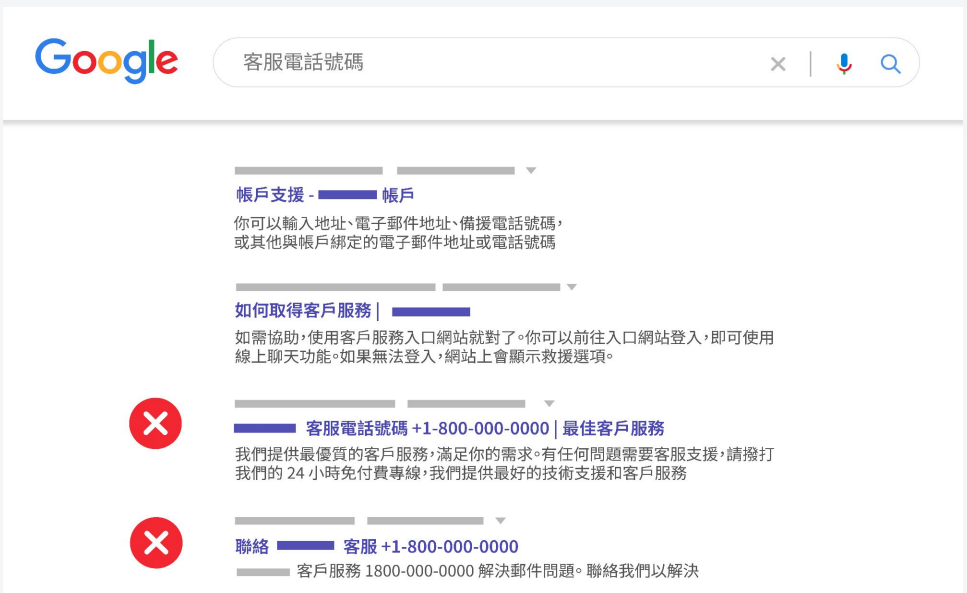
防範

我們一直密切監測自然搜尋結果，找出較有可能傳回詐騙網站的查詢，並致力於針對這些查詢主題做好詐騙手段防範措施，不讓使用者掉入陷阱。自 2018 年以來，我們的系統偵測疑似詐騙網站，並防止這些網站出現在 Google 搜尋結果中，每年能夠避免數億筆搜尋結果出現詐騙內容。

我們能夠在數十億個網頁中識別出干擾行為和惡意行為，因此能夠讓 99% 以上的搜尋結果不含垃圾內容。我們垃圾內容保護措施雖然廣泛，仍有可能遇到無法處理的疑似詐欺和詐騙，因此我們運用各種方法，並參考最佳做法和經驗教訓來補足。舉例來說，我們改良的演算法讓 2021 年與 2020 年相比，特定詐騙性質的搜尋結果減少了 40%，例如客服詐騙。

當然，還有許多其他形式的詐騙，我們要努力的方向包括確保防禦措施跟上最新趨勢。

我們的專用分類器運用演算法，協助偵測垃圾頁面並採取行動，讓我們每天能過濾超過 20 億筆搜尋結果，並攔截 96% 的隱蔽型垃圾內容。



排名保護措施：減少劣質、非原創的搜尋結果

Google 致力於在產品中提供最相關且具公信力的搜尋結果。我們運用排名演算法，顯示相關且優質的來源，確保滿足使用者的期望，並盡量避免讓劣質或有害內容出現在搜尋結果功能或搜尋結果的顯眼位置，因為這類內容不會是使用者想看到的。這些系統的設計是我們對抗劣質有害內容的最強防線。

舉例來說，當不肖分子意圖尋找非法內容時，我們會特意將維護使用者安全相關內容的排名提高。例如，如果使用者搜尋「尋找金融憑證暗網」，我們的搜尋結果會包含建議，告訴使用者如何防止帳戶憑證遭到竊取、如何確定憑證是否已外洩到暗網上，以及該採取哪些行動。

2022 年，我們開始調整排名系統，以減少 Google 搜尋中不實用、非原創的內容，將這類內容的數量控制在非常低的範圍內。我們從這項工作中汲取教訓，推出了隨後的更新。

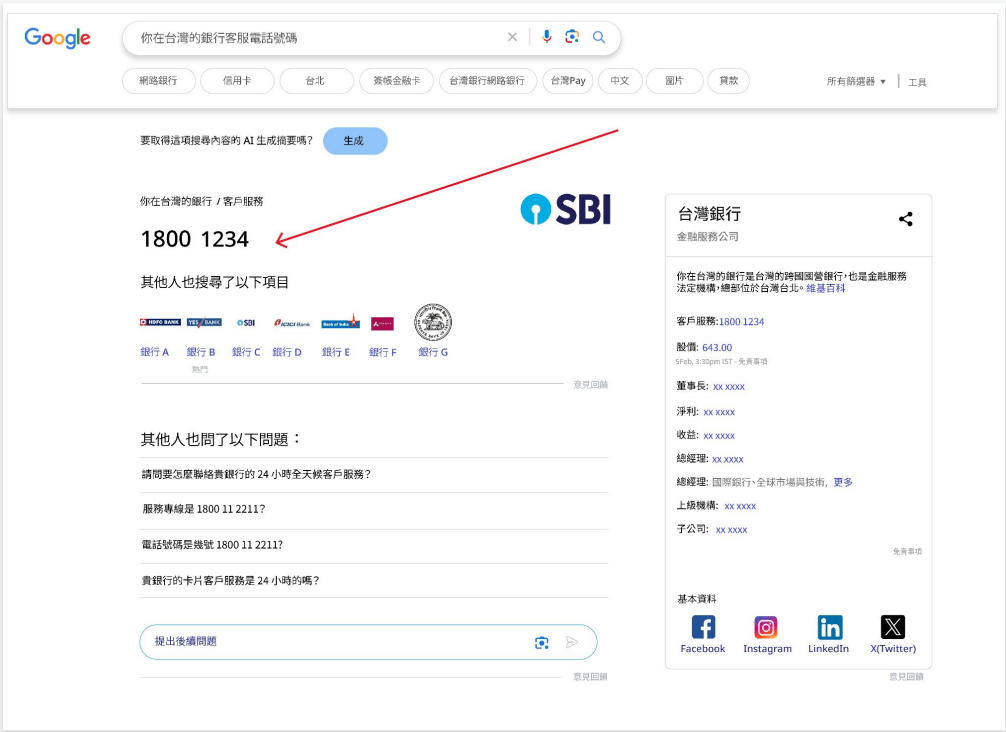
這項更新是關於調整我們的部分核心排名系統，以幫助我們進一步瞭解網頁是否沒有幫助、使用者體驗是否不佳，或者網頁的製作是否像是針對搜尋引擎，而非真人，包括主要針對特定搜尋查詢建立的網站。

我們相信這些更新將減少 Google 搜尋中劣質內容的數量，並將更多使用者導向至實用的優質網站。根據我們的評估，結合這項更新與我們先前的努力，預計搜尋結果中劣質的非原創內容將減少 40%。

使用者賦能

為了進一步保護使用者，我們也會協助為使用者呈現具公信力的內容。舉例來說，我們會針對內容發布的最佳做法向網站發布商提供建議，以協助確保 Google 搜尋結果顯示最準確的發布商業務或服務資訊，這有助於使用者直接與商家聯絡，而非聲稱具有授權關係但又無法確認是否屬實的第三方，達到打擊客服詐騙的效果。

針對大多數商家和服務供應商，我們還在 Google 搜尋結果附近顯示資訊面板 (也稱為觸發式精選摘要 (Trigger Featured Snippets) 或知識面板)，可協助使用者快速取得特定商家具公信力的資訊和客戶服務電話號碼。



偵測

Google 搜尋服務防護機制著重於排名和自動偵測：

- 首先，我們透過自動偵測技術，使用各種信號來識別詐騙。
- 其次，我們訓練排名演算法，以識別劣質網頁，這包括不具權威性或不可信的網址。

這樣一來，我們可以運用這些信號，降低可能品質不佳的網站的排名，以遵循我們的整體信念和原則，在 Google 搜尋中顯示具高度公信力的內容。

處置

(i) 依法移除內容

我們會限制違反產品和服務營運所在地國家或地區法律的內容。

我們提供各種工具，協助使用者根據適用的國內法律，檢舉他們認為應從 Google 服務中移除的內容。例如，根據英國法律，兜售竊取財務資料的網站屬於違法行為，因此受到我們的依法移除內容政策規範。

經此類移除要求所檢舉的內容，將於相關國家或地區移除；但若涉及著作權問題，則將在全球各地撤下。

(ii) 依政策移除內容

和其他 Google 產品一樣，我們也制定並維護各項政策 (稱為「內容政策」)，並在當中說明不允許在各項產品或服務中出現的內容與行為類型。我們的目標是為所有使用者提供易於取得的清楚政策。

針對 Google 搜尋，我們的搜尋政策說明了不允許出現的內容類型，以及從 Google 搜尋結果中移除內容的程序。

例如，有鑑於個人資訊可能涉及身分遭盜用、金融詐欺或其他特定危害等重大風險，我們透過個人識別資訊 (PII) 移除政策為使用者提供保護。根據這項政策，使用者或授權代表可提出要求，移除 Google 搜尋結果中包含特定 PII (例如機密的身分證字號、銀行帳號等) 內容的連結。

我們持續努力改進這些政策並加強落實。因此在 2024 年，Google 搜尋推出了多項防堵垃圾內容相關更新，並提升搜尋結果的整體實用性和品質。這些更新包括修改用於偵測垃圾內容和實施違規處置決策的規範，以及強化我們的排名系統，以幫助我們進一步瞭解網頁是否無益且使用者體驗不佳 (但這些網頁也可能不符合我們對垃圾內容的正式定義)。

除了強化排名系統之外，我們針對垃圾內容的更新還有以下三個關鍵元素：

- 大規模內容濫用：我們長期以來一直以政策禁止使用自動化手段大規模生成劣質或非原創內容，以達到操控搜尋排名的目的。隨著這些行為變得日益複雜，我們的政策現在將重點放在濫用行為上，即透過自動化或人工方式，或者兩者結合，大規模製作內容來提升搜尋排名。
- 網站信譽濫用：某些網站允許第三方發布劣質內容來營利 (例如，第三方可能會在受信賴的資訊網站上發布發薪日貸款評論)。現在只要屬於低價值第三方內容，主要以提升排名為目的，缺乏網站擁有者嚴格監管，我們會將其視為垃圾內容。這項政策將於 5 月 5 日開始實施，以確保網站管理員有時間瞭解政策內容並視需要調整做法。
- 過期網域濫用：有時候，有心人會購買過期的網域並重新利用，以提升劣質或非原創內容的搜尋排名，這可能會導致使用者誤以為新內容是舊網站的一部分。現在，如果發現為了操控劣質內容的搜尋排名而購買和重新利用的過期網域，我們也會視為垃圾內容。

為了更有效地呈現實用資訊給使用者，我們做了許多努力，都是希望可以將 Google 搜尋上的劣質內容數量控制到最低，而這些變更只是其中的一環。

如要進一步瞭解 Google 搜尋如何保護使用者，請造訪我們的專門網站「Google 搜尋的運作方式」，其中說明我們如何大規模地協助使用者存取實用的相關資訊、遏止垃圾內容，以及如何透過嚴謹的測試不斷改進 Google 搜尋。

附件3.7 購物

我們推出了多種工具和功能，確保使用者能夠安心購買商品並信賴商家。以下是我們協助使用者在 Google 上安全購物的三種主要方式：

(i) 利用自動化技術，協助我們快速準確地審核產品資訊

產品和商家在上架 Google 之前，都必須通過深入的安全審查。有了購物圖譜 (我們的全球產品和賣家資料集)，我們的系統可以快速審查商家是否合法、展示的產品是否準確，以及呈現的內容是否符合我們的政策。

和任何社群一樣，我們需要設立規則，才能維持彼此尊重的環境。我們的購物政策涵蓋產品資訊和購物廣告，列出了 Google 允許及禁止的內容，包括從網路上檢索並顯示在購物結果中的產品。這些政策讓使用者更安心，確定看到與收到的產品不會有差別，而且無需費心過濾不當的內容，例如暴力武器、商家的不實業務資訊或仇恨內容。

我們結合自動化系統和人工評量方式，確保產品資訊符合我們的政策規定。我們的違規處置技術採用演算法和機器學習技術，根據人工審查員做出的判斷進行模擬，有助於保護消費者並確保自家平台安全無虞。自動化技術協助我們更有效率、更精準地審查大量產品。若特定案例的情況更加複雜、情節重大或無法一概而論，通常會交由經過特訓的專家審查及評估。

光在 2023 年 1 月，我們就擋下了超過 1 億則產品優惠資訊，並拒絕核准將近 30 萬個有品質問題或未遵循政策規定的帳戶。

(ii) 運用商店徽章和提供其他視覺提示，幫助使用者認明優質商家

為了提高消費者對商家的信賴感，Google 會頒發商店徽章給購物體驗較為優良的商家，協助消費者辨別優良商家。商店需要滿足出貨快速、退貨便利、網站內容優質和使用者評分優良等標準，才能獲得此徽章。

我們還會顯示產品及不同賣家的評分，讓消費者瞭解其他購物者對這些產品和商家的體驗。此外，由於我們的產品資訊可將消費者直接導向至商家網站，因此消費者還可直接在商家網站上瞭解更多資訊。

(iii) 透過自動化系統與人工審查團隊，即時監控商家和產品資訊

產品資訊發布後，Google 仍然會繼續執行安全措施。我們的自動化系統會持續監控違規活動，人工審查員也會隨時待命，審查可能需要更深入調查的問題。

例如，我們的系統可能會發現某電子產品公司將價格降低了 70%，並從網站上移除了運送資訊；或者發現原本販售毛衣的商家現在開始販賣家用電器。這些都可能是反映需要我們團隊進一步檢查的信號，以確定是否有誤導性內容。

此外，在商家加入之後，我們會密切留意商家及相關產品資訊，以確保從他們加入 Google 以來，未出現任何可疑變動。例如，如果商家在加入時符合國家或地區適用的酒精飲料銷售規定，我們會定期檢查，以確保商家依舊符合這些資格條件。

發現可疑行為時，我們會採取多種行動，包括移除看似可疑或違反政策的產品資訊，以及禁止商家在 Google 發布產品資訊。如果使用者看到不對勁或不符合我們政策的內容，例如超高價格或暴力武器，可以透過產品頁面右下方的「檢舉產品資訊」連結或說明中心提出檢舉。

附件3.8 付款

對所有付款系統而言，保障隱私權與安全性應是必備的基本功能。無論消費者使用的是 Google 錢包還是 Google Pay，Google 的首要之務都是確保付款的安全性。

我們運用世界級機器學習和詐欺偵測演算法來達成以下目的：

- 確保無論是線上交易還是在店內感應支付，都快速、安全又輕鬆
- 保障財務安全。

我們也為生態系統的合作夥伴提供詐欺管理工具。在註冊權杖期間，GPay 會提供工具包，協助發卡機構驗證使用者的帳戶擁有權，而在交易期間，則會提供權杖和密文，供合作夥伴自行評估風險。採用裝置權杖化技術，整個付款流程中都無需擔心信用卡資訊外洩，而針對線上或高額交易的雙重認證機制，則可提供最高標準的安全保護，同時確保使用者享有順暢體驗。

我們將所有資料儲存在加密的安全伺服器上。

我們根據資訊透明、使用者選擇權與安全控管等嚴格的隱私權原則，設計並開發旗下的付款產品。例如，你可隨時在帳戶中儲存或移除付款方式，並輕鬆開啟/關閉付款卡的自動填入功能。此外，我們用於保護 Google 帳戶的所有安全措施也適用於付款卡。舉例來說，我們會使用與其他 Google 功能所用的相同技術來確認身分，然後才會授權透過裝置付款。如需保護 Google 帳戶安全的建議和指引，我們建議使用者定期進行安全檢查。

除了 Google 的安全基礎架構之外，使用者還可享有卡片服務供應商平常提供的所有保護措施。感應支付不僅方便，也比刷卡更安全。

以下是我們防止透過 Google 錢包未經授權付款的方式：

- Google 錢包不會在手機上儲存付款卡資訊
- 相關資料會儲存在安全加密的伺服器上
- 如果裝置遭竊，也不用擔心，因為付款方式安全無虞。要使用 Google 錢包進行感應支付，必須在近期使用安全螢幕鎖定機制 (例如 PIN、密碼、臉部辨識或生物辨識) 完成驗證
- 如果裝置遺失或遭竊，使用者可使用「Google 尋找我的裝置」應用程式遠端清除裝置。

此外，當使用者將付款卡加入 Google 錢包時，系統會建立裝置專屬虛擬帳號 (也稱為裝置權杖)。這樣一來，就不用將真實卡號儲存在裝置上或提供給商家。

附件3.9 Google 地圖

我們每天都會收到 Google 地圖使用者貢獻的約 2 千萬筆內容，這包括最新營業時間、電話號碼、相片和評論等。

和任何接受貢獻內容的平台一樣，我們必須保持警覺，努力打擊濫用行為，並確保這些資訊正確無誤。我們採取嚴密的措施防堵不實評論。考慮到收到的評論數量之多，我們投入心力發展機器學習技術和真人審查員團隊，致力於在消費者、商家和使用者之間建立信任。只要有人發布評論，我們就會立即將評論傳送至我們的審核系統，以確保評論並未違反我們的任何政策。

由於結合了機器學習技術和人工審查機制，我們得以持續減少 Google 地圖上出現的詐欺性或不當內容，事實上，在 Google 地圖顯示的所有內容之中，這類內容占不到百分之一。

機器擅長找出模式，因此是我們的第一道防線。這些模式通常有助於機器立即判斷內容是否正當合法，並在有人實際看到之前，就將絕大多數虛假和詐欺性內容移除。我們的人工審查員團隊則毫不懈怠地審查遭標記的內容。發現違反政策的評論時，我們會從 Google 移除這些評論，有時可能會將使用者帳戶停權，甚至提出訴訟。除了審查遭標記的內容外，我們的團隊還會主動識別潛在的濫用行為風險，以降低濫用行為和攻擊得逞的可能性。