Google

Dec 2024

Tackling scams and fraud together

Google white paper

Executive overview

Our mission at Google is to organize the world's information and make it universally accessible and useful. Core to this mission is the relevance and quality of the information we present to users. That's why we take our responsibility to tackle online fraud and scams and provide access to trustworthy information and content very seriously.

Financial fraud is harmful to consumers and legitimate businesses, and undermines individual users' trust in digital platforms. Online scams in particular are a growing global problem that can have a devastating impact on individuals and businesses, affecting people of all ages and backgrounds. Additionally, scammers are constantly evolving their tactics and taking advantage of new technologies and social trends. The COVID-19 pandemic saw an acceleration. People became more reliant on digital services, experiencing a significant increase in online scams and fraud. Accordingly, we are determined and committed to continue to innovate and invest heavily in identifying and combating fraud.

This policy white paper focuses on financial fraud, notably cases perpetrated by organized, often transnational, criminals. It details how Google is responding to this challenge, and offers recommendations for how we can better partner across the ecosystem to maximize the impact of our efforts to tackle this criminal threat to society and the economy.

Google's approach to tackling scams and fraud

Google fights scams and fraud by taking proactive measures to protect users from harm, deliver reliable information, and partner to create a safer internet. We do this through policies and built-in technological protections that help us to prevent, detect, and respond to harmful and illegal content, and we scale our industry-leading practices to keep users safe online through proactive partnerships and communication with experts and organizations such as national anti-scam agencies.

These protection mechanisms are tailored across each of our products. For example, Android incorporates multiple layers of protections across its mobile operating system to detect and prevent scams, such as *Phone by Google*, which screens for potential spam and scam calls, including thanks to AI technology. Chrome protects users through a variety of features, including *Google Safe Browsing, Chrome Password Alert, Advanced Protection Program (APP)*, and *Enhanced Safe Browsing*. Google Safe Browsing, for example, warns users if a website is dangerous and is attempting to phish their credentials. And Google Ads has developed - and regularly updates - a range of policies and safety features tailored to the ads ecosystem, such as its policy against public figure impersonation or Limited Ads Serving. The latter requires advertisers to go through a "get to know you" period before earning the privilege of advertising in abuse-prone areas. These dedicated policies have been key to dramatically reducing the volume of scammy ads.

Thanks to our years of experience in this space, and our policy and product protections, we can demonstrate significant impact. For example:

- Gmail blocks 99.9% of spam, malware, and dangerous links from reaching users' Gmail inboxes.
- Google Safe Browsing automatically protects more than 5 billion devices, keeping users secure from bad websites.
- In Ads, we blocked or removed 206 million ads in 2023 for violating our misrepresentation policy, and 273 million advertisements for violating our financial services policy.
- Google Play Protect scans 200 billion Android apps daily, helping keep more than 3 billion users safe from malware.
- We've expanded our financial services verification program to seventeen countries and regions globally, which requires financial services advertisers to demonstrate that they are authorized by relevant government authorities in order to promote certain financial products and services on Google's platforms through ads.
- On Android, our newly launched Google Play Protect anti-fraud feature has already blocked nearly 900,000 high-risk sideloaded app installation attempts on over 200,000 devices during the first six months of the pilot in Singapore and we are expanding that initiative.

Principles for responsible cooperation and regulation against fraud

While Google has made significant progress in combating fraud and scams, this is not an online problem for online platforms to address alone. It is a whole-of-society threat that necessitates a cross-industry response and collaboration with other stakeholder groups, from government and law enforcement to sectors such as financial institutions and telecommunication service providers. We cannot operate in silos, as we will miss synergies and leave gaps in our overall approach to tackle the transnational organized criminals who are behind much of the scams problem.

We need a global, comprehensive public policy and societal approach to tackle the problem. By combining information sharing, practical enforcement actions, innovation, education and individual awareness and responsibility, and supported by a spirit of collaboration among stakeholders, we can make the online environment safer for everyone.

In this white paper, we propose three key policy principles as a call to action across the ecosystem to more effectively, and more collaboratively, fight against online scams and fraud. Each of these principles is accompanied by a series of policy recommendations to guide these efforts.

Principles	Policy recommendations
Enable cooperation and sharing *Develop new information sharing frameworks across and between governments and industry, and facilitate increased cooperation through international forums.*	• Adopt laws to facilitate cooperation and information sharing between industry and governments, including sensitive data and enabling action on suspicious activity with opaque evidence. • Form cross-border working groups and access arrangements to coordinate and strengthen enforcement and promote interoperability. • Empower international forums like the Global Anti Scams Alliance to increase cross-industry dialogue and cooperation.
Incentivize action across the community *Provide the necessary legal framework to allow the community to take pre-emptive action against fraudulent content and actors.*	• Clearly define what is illegal activity and strengthen law enforcement capacity to act. • Incentivize preventive action by stakeholders through "Good Samaritan" liability protections. • Invest in Responsible AI and policies that encourage technological innovations that protect people from scams and fraud and help identify fraudulent content.
Invest in education & protection for users *Invest in education and awareness raising, particularly with other organizations and across sectors.*	• Launch public education campaigns about online scams and how to protect oneself. • Develop policies and product features that effectively protect users against harm. • Provide clear and accessible reporting channels for victims of frauds and scams. • Pursue secure digital transformation.

Introduction: The problem of online fraud and scams around the world

Fraud and scams are not a new phenomenon: they have existed since the early days of human interactions and have developed over the centuries, from everyday low level fraud to major financial bubbles that caused nation-wide turmoil, increasing in sophistication as new financial mechanisms and technologies appeared.

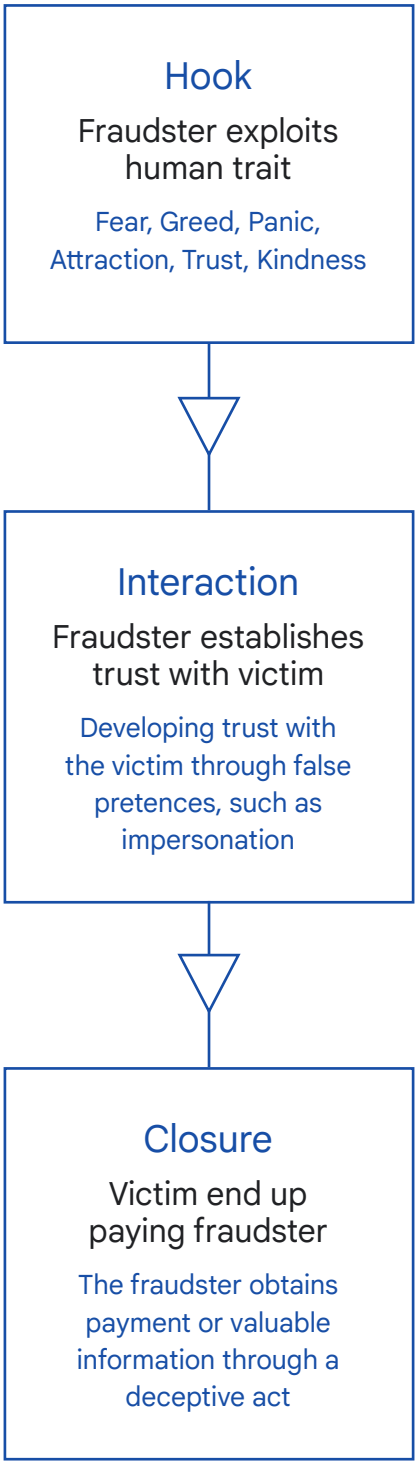
As the Internet and smartphones have integrated into our daily lives, they have become pivotal gateways to our personal data — whether it is our banking information, shopping accounts or health statistics. This has enabled fraudsters to increase the scale, scope, and speed of their illicit activities; they are taking advantage of social media and messaging platforms to operate across borders, and they are quick to exploit vulnerabilities that emerge through new digital financial institutions, e-commerce, and streaming platforms. Increasingly, online fraud is being driven by transnational criminal organizations that have access to greater resources and cross-border mobility, as well as access to more sophisticated technologies.

All of these factors have led to a dramatic increase in scams and fraud in the past several years, with estimates pointing to as high as US$1.026 trillion lost to fraudsters in 2022-23 alone. According to the Global Anti-Scams Alliance, nearly 1 in 4 individuals worldwide have been affected by online scams and fraud, with the Better Business Bureau reporting an 87% rise in online scams between 2015 and 2022.

Fraud types and trends

Given evolving tactics technologies, there are many variations of online scams and fraud, the most common of which include:

- Impersonation fraud: This includes scenarios where victims are contacted via mobile or social media applications by criminals pretending to be, for example, government officials, relatives or friends. They prey on the victims' emotions to induce payment, hand over control of payments accounts or to carry out financial activities such as a loan application or an account opening to receive criminal proceeds. A variant involves the misuse of AI technology to produce 'deep fake' images or videos of celebrities and trusted figures or celebrities to reinforce the appeal of fraudulent schemes, also known as 'public figure impersonation'.
- Investment scam (Online trading / trading platform fraud): Victims are lured by fake advertisements and online influencers to non-existent or fake platforms for trading or investment related to both fiat and virtual assets.
- E-commerce scam: With the increased popularity of online purchase and digital transactions, the rise of scams on digital marketplaces has continued to evolve. A common method is that users purchase goods on fake shopping websites that use online payment methods.
- Online romance fraud: Victims are duped into sending money either directly or indirectly through too good to be true investment schemes to criminals after being convinced they are in a romantic relationship.



- Employment scams: With the increase in job searching online, fraudsters post fake job listings with websites that look legitimate. Fake job offers online dupe victims to pay fraudsters upon various pretenses including advanced payment for purchasing commodities to boost sales of a trading platform or a guarantee fee to secure employment.
- Payment hijacking: Fraudsters hijack the payment method for a user by fetching sensitive information from the user with social engineering techniques. An example affecting the gaming community is where local payment systems require the player to scan a QR code to pay money.
- Malware scams: Users are exposed to what appears like an innocuous download of a helpful app, which in fact contains malware which could be used, among other things, to steal personal identification information.
- Phishing fraud: Victims are deceived into revealing sensitive information such as personal data, banking details or account login credentials, for instance, by a person posing as representatives of a trusted service like a bank, utility service, or Internet Service Provider (ISP). The criminal will then use the information to drain the victims' money from their payments accounts, open new payment accounts or make fraudulent transactions. This can be supported by physical means, for instance in the case of credit card scams, whereby fraudsters use various approaches to obtain card information including use of credit card skimmers at physical points of sale, social engineering, or fake websites.

Part one

Google's approach to tackling online scams and fraud

Google's mission is to organize the world's information and make it universally accessible and useful. Keeping billions of users safer online is central to that mission.

That's why we've been actively working to counter the threat of scams and fraud for over two decades. Google defines scams as "deceptive schemes targeting individuals or organizations, with the intent of taking money and/or personal information".

Google fights scams and fraud by taking a three-pronged approach to protect users from harm, deliver reliable information, and partner to create a safer internet. We have robust technology and features, policies, and processes in place to do this across our products. However, bad actors are constantly evolving their tactics, and the volume of scams across all online platforms increased significantly in 2024. To counter these ever-shifting threats, we constantly update policies, deploy rapid-response enforcement teams, and sharpen our detection technology.

1. We protect users from harm

Across our products and services, Google has built-in advanced protections and policies that prevent the distribution of fraudulent and scam content, detect and evaluate potentially violative content, and respond to bad actors and abusive content appropriately:

Prevent: Our first line of defense are the policies that guide our products and the safety features we build into them to deter or prevent bad actors from publishing fraudulent content.

For example, bad actors will often attempt to use ads to run scams so we have developed robust policies that help protect users from harmful and misleading advertising. One of those policies is the Google Ads misrepresentation policy which prohibits ads or destinations that deceive users by excluding relevant product information or providing misleading information about products, services, or businesses.

Detect: While bad actors exploit new technologies to refine their attacks, our trust & safety teams and security experts leverage AI and cutting-edge technologies to proactively detect harmful content and abuse. This allows us to continuously adapt our systems and stay one step ahead of evolving threats.

Respond: When a piece of content is flagged, we rely on both humans and AI-driven technology to determine whether it has violated our policies and respond appropriately. We take appropriate action for that service, which could include restricting, removing, demonetizing, or taking account-level actions to reduce future abuse. We also learn and refine our practices on an ongoing basis: after taking action on certain content, that content is then used to train our models to detect similar policy violations, and to inform new or evolved policies or product features.

2. We deliver reliable information

We enable confidence in the content provided by our products and services by delivering information people can trust and best-in-class tools that put users in control of evaluating this information.

3. We collaborate through programs and partnerships

Tackling fraud and scams remains a complex and ongoing challenge that necessitates a cross-industry response and collaboration with stakeholder groups including policymakers, experts, creators, publishers, enforcement agencies, business owners and the general public.

Empowering individuals with knowledge is also an effective tool to address scams and fraud. Google has been actively working with financial institutions and regulators to help raise user literacy with educational safety guides & toolkits.

Tackling abuse at scale

Information quality and content moderation are integral to our mission of making information universally accessible and useful. This mission requires us to strike a careful balance between the free flow of information and social responsibility. The product, policy, and enforcement decisions we make are guided by our principles of valuing openness and accessibility, respecting user choice, and building products and services for everyone.

We have developed the following policies and tools to keep our users safe from fraud and scams:

- Acceptable use policies: Google's services have policies outlining acceptable use, which make clear that use of our services to engage in behavior in an attempt to scam others is not permitted. Where we or users identify that someone is not acting in accordance with the acceptable use policy, users can block them as contacts on Gmail for example and we can take action against the infringing actor, behavior or content. For example, Google Ads Policy prohibits deceptive practices.

- Report activity: If users suspect or identify activity which violates the relevant product's acceptable use policy or Google's general terms of service (such as sharing spam or abusive content), they can report that user to Google for these policy violations. Once reported, Google reviews the complaint and may remove the abusive content and disable the account.

- Block function: In our communications tools, users have the autonomy to block other users from contacting them for any reason, including for sending spam / abusive content.[1] A user may block direct messages or people in Gmail, Chat or Messages, which prevents unwanted senders from contacting that user further. Similarly for ads, My Ads Center gives users control to adjust preferences around what types of ad categories they see or to turn off personalized ads.

- Automated detection: Google's systems actively scan its services for potentially violative content and we engage in enforcement efforts for potential scam activity in breach of acceptable use policies. This includes proactively classifying communications as spam, or issuing service-level or account-level suspensions for violations of our terms or policies. Google also sends criminal referrals to law enforcement to help facilitate real world consequences against scammers.

- Human review: We also have manual reviewers that review abuse reported by users, priority flaggers or by machine learning. For example, if a user files an abuse report from the Chat user interface, our team of manual reviewers will review the last 50 messages sent in a reported conversation and review the account signals of the users sending the messages to determine if a user is abusive. If we find that abuse has occurred, we will typically disable the account for this service.

- Verification: We use privacy-safe techniques to verify that there are real users behind accounts, and specific verification methods for advertisers in particular. Advertiser verification programs can include steps such as verifying the advertiser's identity, submitting information about business operations, or providing evidence of necessarily local licenses.

- Controls: Wherever possible, we deploy technical solutions and features in our products to prevent fraud and other online harms. As one example, Gmail's AI-powered defenses stop more than 99.9% of spam, phishing and malware from reaching inboxes and block nearly 15 billion unwanted emails every day. Multi-factor authentication is another example of protective control that we have introduced.

1 See for example, for Google Chat, "Block and report someone", https://support.google.com/chat/answer/9277792?hl=en&co=GENIE.Platform%3DAndroid

Google's Trust and Safety approach

In order to organize our internal efforts strategically, we have developed a three pillar approach to tackling fraud and scams across our Trust and Safety teams:

The actor & behavior pillar focuses on increasing actor-level and behavioral enforcement, as well as strengthening our controls that help minimize the risk of successful scamming attempts. This is done by (i) enhancing the processes for identity and business operations verification; (ii) ensuring that new accounts are limited in capability and access until they have earned trust; and (iii) leveraging leads and signals from a range of sources in order to take action against bad actors and scams before harm occurs that impacts users.

The taxonomy, policy & measurement pillar is intended to strengthen our internal organization and strategy to combat scams and fraud. This includes having common definitions across our teams, unified procedures across product areas, developing and refining appropriate use policies, and measuring our effectiveness in enforcing our policies against frauds and scams.

The outreach and engagement pillar is about interacting with internal and external experts and partners to improve user safeguards and information sharing. This includes efforts to improve user experience safety features, and activities such as lead and signals sharing via industry coalitions with other digital platforms and relevant industry actors, and operational partnerships with law enforcement.

There is a crucial feedback loop between these three pillars of our Trust & Safety efforts, and our ongoing work to develop new and improved technology and product safety features. What we learn from detecting and taking action against incidents, from our exchanges with other stakeholders and experts, and our associated research, feeds into our consideration for the product roadmap of which 'safety by design' features we can develop or refine, contributing to ongoing improvements in supporting users across their journey online.

Google's product protections from scams and fraud

In addition to our policy actions to fight frauds and scams, we focus on designing each of our products with built-in security measures.

We work to ensure that whenever possible, all of our products automatically protect users online, allow them to easily manage their online security and choose the right level of protection. Our approach is tailored to the specifics of each product to ensure it is most relevant and effective:

- Android incorporates multiple layers of protections across its mobile operating system to detect and prevent scams, such as *Phone by Google*, which screens for potential spam and scam calls.
- Google Messages also introduces new levels of protection thanks to Rich Communication Services (RCS).
- Chrome protects users through a variety of features, including *Google Safe Browsing, Chrome Password Alert, Advanced Protection Program (APP)*, and *Enhanced Safe Browsing*. Google Safe Browsing, for example, warns users if a website is dangerous and is attempting to phish their credentials.
- Google Ads has developed - and regularly updates - a range of policies and safety features tailored to the ads ecosystem, such as its policy against public figure impersonation or Limited Ads Serving. The latter requires advertisers to go through a "get to know you" period before earning the privilege of advertising in abuse-prone areas. These dedicated policies have been key to dramatically reducing the volume of scammy ads.
- YouTube benefits from Google's work on ads safety, but YouTube's protections against scams extend to content shared on the platform via content policies, called the Community Guidelines. Under the deceptive practices and scams policies, YouTube prohibits scam content such as offering cash gifts, "get rich quick" schemes, or pyramid schemes (sending money without a tangible product in a pyramid structure) in videos, comments, and metadata posted by users.
- Gmail has long been known for its effectiveness in protecting users from spam and phishing attempts thanks to AI-enhanced filtering. We have adapted to more sophisticated phishing campaigns over the years, while also prioritizing protections against phishing attempts that are most immediately threatening to users' data and credentials.
- Google Shopping protects shoppers by removing fake reviews brought to our attention and utilizing numerous signals to proactively minimize abuse and spam, thanks to automated vetting, store badges and other visual cues which help point out quality businesses.
- Google Search protections focus on ranking and automated detection, using a variety of signals and training our ranking algorithms to recognize low quality or suspicious web pages.
- Google Cloud has dedicated counter-abuse teams which work to ensure our products are used in the intended manner and that our platform isn't misused or abused. To strengthen our customers' confidence in their ability to quickly detect and stop cryptomining attacks, we are introducing a new Cryptomining Protection Program, which offers financial protection up to $1 million to cover unauthorized Google Cloud compute expenses associated with undetected cryptomining attacks for Security Command Center Premium customers.
- Similar efforts extend to our entire range of products such as Payments and Maps.

Thanks to these product protections, we are already having a significant impact. For example:

- Gmail blocks 99.9% of spam, malware, and dangerous links from reaching users' Gmail inboxes.
- Google's messages and phone apps help protect against voice phishing and scams with built-in caller ID, spam protection and Call Screen by blocking dangerous calls and warning you about suspicious callers.
- Google Safe Browsing helps keep users secure from bad websites, automatically protecting more than 5 billion devices.
- Ads: In 2023 we blocked or removed 206.5 million advertisements for violating our misrepresentation policy and 273.4 million advertisements for violating our financial services policy. We also blocked or removed over 1 billion advertisements for violating our policy against abusing the ad network, which includes promoting malware. and suspended 12.7 million advertiser accounts (nearly double from the previous year). AI played a major part in scaling these efforts and reinforcing their effectiveness. As of Q4 2024, we've partnered with governments on expanding our financial services verification program to seventeen countries and regions globally. Under this policy, financial services advertisers are required to demonstrate that they are authorized by relevant government authorities.
- Android: In Singapore, we partnered with the Cyber Security Agency in February 2024 to launch a new enhanced fraud protection feature within Google Play Protect for all local Android devices. The feature protects mobile users against malware scams by blocking potentially risky sideloaded apps. Within six months, the feature had already blocked nearly 900,000 high-risk app installation attempts on over 200,000 devices. Many of these apps are impersonations of popular messaging apps, gaming apps or e-commerce apps that were potentially used for fraud.

Annex 1 provides details on the protections in place across the range of Google's main product areas.

Part two

Tackling scams and fraud together: Policy recommendations for the community

Fraud is a whole-of-society issue that requires strong and sustained cooperation from a range of actors.

Google is committed to constant improvement to protect our users and help them find information they can trust. But we know that a problem of this scale, often driven by transnational criminal organizations, cannot be tackled by any organization or company alone. In order for these efforts to be most effective, we must work together with other interested parties, including policymakers, regulators, civil society, and the private sector.

We present below three key areas for action by the fraud and scams community, including specific policy recommendations to guide these efforts.

1. Enable cooperation and sharing

Multi-stakeholder cooperation is key to countering online fraud and scams effectively. As a concerned community, we should foster new cooperation and sharing between stakeholders across industry, government, the technical community, academia and civil society in order to interact and share information and good practices. First, we need greater dialogue and practical operational coordination. This includes being able to share intelligence signals that help detect scams and their perpetrators, and enabling law enforcement cooperation - both with private entities like Google, but also across government borders. We also need ever-closer legal and regulatory cooperation around safety, data exchange, and international anti-crime collaboration. Policymakers should ensure that the legal framework necessary for this type of information sharing is put in place.

Adopt laws to facilitate cooperation and information sharing between industry and governments

Online platforms and financial institutions need to be legally authorized to cooperate and act against scammers based on suspicious activity with opaque evidence. This includes sharing relevant information such as sensitive data with Government agencies, and vice versa. This also encompasses critical information like account details to be able to track the money, and therefore criminal activities.

Policymakers should ensure that there is a clear, balanced, and effective legal framework in place that enables this type of information sharing. Fraud prevention at scale represents a significant public interest and, therefore, it is important for the processing of sensitive data to be allowed for the narrow purpose of detection and prevention of fraud, provided platforms local jurisdictions have in place measures to safeguard the fundamental rights and interests of data subjects. Currently, sharing actionable abuse signals with law enforcement in order to support criminal investigations is not allowed in many jurisdictions, which hampers the ability of public and private sectors to work together. In some jurisdictions, the law enables financial institutions to share this type of sensitive data with local law enforcement; however, these legal protections do not extend to online services.

In other jurisdictions, it may also be necessary to clarify the intersection between consumer safety and anti-fraud laws with other legislations such as data protection and competition rules.

Policymakers should also streamline the legal process for submitting and processing requests related to fraud investigations, ensuring that private companies can efficiently comply with lawful requests from authorities, while respecting due process. In this respect, there are various possible voluntary frameworks that governments could put in place such as CLOUD Act Agreements, the Budapest Convention, and more recently the Second Additional Protocol to the Budapest Convention, which establishes mechanisms for dealing with cross-border access in a manner that respects the rule of law. *Specifically we would encourage Governments to consider signing and ratifying the Second Additional Protocol.*

Form cross-border government working groups to coordinate and strengthen enforcement and promote interoperability

Facilitating and strengthening cross-border investigations is essential so that law enforcement agencies, and private organizations - like online platforms and financial institutions - can work together across borders to tackle transnational organized crime networks. The perpetrators, victims, key documents, and third parties involved in the fraudulent transaction are often widely dispersed across borders, which makes it challenging for enforcement agencies and other relevant government entities in a single country to gather all the information necessary to detect scams and fraud and investigate them effectively.

As one example of this type of cross-border cooperation, in 2023 the Council of the OECD adopted Guidelines for Protecting Consumers from Fraudulent and Deceptive Commercial Practices across Borders, The Guidelines recommend specific enhancements to both domestic legal and enforcement frameworks, as well as mechanisms to facilitate notification, information sharing, assistance with investigations, and confidentiality, notably across borders and in cooperation with private sector entities. For global businesses like online platforms and financial institutions, this international alignment and legal interoperability would greatly improve the effectiveness and efficiency of compliance efforts, as well as private litigations.

We would encourage Policymakers to explore implementing similar frameworks and collaborations, either regionally or through bodies like the OECD, in order to achieve better detection and enforcement against organized criminal networks, and strengthen these actions in the future, especially across borders.

Empower international forums like the Global Anti-Scams Alliance

Greater dialogue among the anti-scam community is essential, from detecting and sharing information about new threats, trends and patterns, to collaboration on actual investigations. Policymakers should encourage online platforms and the private sector - particularly high-risk organizations like financial institutions - to hold regular, multi-stakeholder dialogues. Similarly, Governments should join or endorse these cooperation forums to ensure close and consistent coordination across the ecosystem. This can support proactive prevention, enable quick and effective responses, and help others be better prepared.

An example of the channels that can be used for relevant stakeholders to partner for effective information sharing against scams is the Global Anti Scam Alliance (GASA), which Google joined in March 2024 as a Foundation Member. GASA is a global non-profit coalition with a broad network of over 100 members and close relationships with Law Enforcement Agencies, including the FBI and Europol, and industry sectors such as banks and major tech companies. GASA working groups and innovation include workstreams dedicated to data sharing for scams signals and metrics such as malware and phishing reports.

In October 2024 Google entered into a partnership with the Global Anti-Scams Alliance (GASA) and DNS Research Federation (DNS RF) to launch the Global Signal Exchange (GSE). The GSE is a global clearing house for online scams and fraud bad actor signals, with Google becoming its first Founding Member. The collaboration brings together GASA's unparalleled global network of stakeholders, DNS Research Federation's data platform already storing over 40M signals, and Google's deep expertise in fighting scams and fraud. By combining our efforts, and creating a central platform, GSE will fill a gap of how abuse signals are exchanged to identify and disrupt fraudulent activity faster across different sectors, platforms and services in a way that is easy to use, efficient and works at the scale of the Internet.

We would recommend that other parties, including government agencies and relevant private organizations such as financial institutions, consider joining the GASA GSE Platform, as it has been developed specifically for the purpose of sharing scams-focused insights and intelligence, and provides a unique common setting for stakeholders around the world to join forces to tackle the problem.

We also need partnerships across the ecosystem to address security events. A promising development in this area is Risk and Incident Sharing and Coordination (RISC). RISC was launched initially as a Google framework to share security events with business partners any time that Google detects a major change in a user account's status. For example, if a user's Google Account were hijacked by a bad actor, Google can send a signal to connected apps and platforms, allowing them to take appropriate actions or put additional protections in place. We encourage relevant organizations to join the RISC framework.

2. Incentivize action across the community

Governments have a key role to play in providing the public policy and legal frameworks necessary for the community to take action against fraudulent content and actors and invest in new technologies that will improve our collective ability to fight against scams and fraud, including the development of Responsible AI.

Clearly define illegal activity and strengthen law enforcement capacity to act

In order for law enforcement agencies and private entities, such as online platforms and financial institutions, to take effective action against fraud and scammers, policymakers must clearly codify what constitutes illegal activity in their jurisdiction. Without this legal clarity, public and private entities may be reluctant to take decisive, or preventive, action.

Governments should therefore review their existing legal frameworks to assess whether they are fit for purpose for fighting scams and frauds. If gaps are found, policymakers should look to introduce new or enhanced frameworks, or issue supporting guidance to clarify how existing frameworks should be enforced. These legal frameworks should include reasonable enforcement mechanisms, and clear definitions of what constitutes illegal activity in this context. This is particularly important as oftentimes digital platforms are not in a position to determine what is a scam because it does not have visibility on the overall scheme including behavior occurring on other platforms or offline.

Additionally, providing sufficient training and capacity for law enforcement agencies to better investigate and prosecute cybercriminals can help strengthen enforcement, and should be prioritized by governments. In many situations, the private sector can help with these capacity building efforts.

Incentivize preventive action by stakeholders through "Good Samaritan" liability protections

Governments can incentivise companies to take preventive action against frauds and scams through the adoption of 'Good Samaritan' liability protections, which shield online platforms and other intermediaries from liability for their proactive efforts against online harm.

A good way to understand why this is important is to think about what might happen to a company if the protections are not in place. For instance, platforms could be sued for decisions around removal of content from their platforms, such as our removal of hate speech, mature content, or videos relating to pyramid schemes (as mentioned above, the latter is not deemed an illegal scam except in a minority of countries around the world). The result of these pressures would be to disincentivize companies from developing robust content moderation systems.

It is important that the liability regime provides clarity, and does not disincentivize services from taking positive, voluntary measures to tackle scams and other content challenges. A service should not become liable for any of the information that it hosts simply by virtue of the fact that it has taken voluntary action in good faith, whether of an automated or a non-automated nature. 'Good Samaritan' protections would address that concern by giving protection for platforms to seek out and remove harmful content, without risking the loss of liability protections for occasional failures in that process.

Examples of legal provisions to incentivise preventive action by shielding online services from liability for their proactive prevention efforts against online harms are included in the EU Digital Services Act (DSA) and section 230 of the US Communications Decency Act (CDA). In other jurisdictions, similar protections could avoid disincentivizing companies to proactively search for fraudulent content and actors or to share threat information.

Invest in responsible AI and policies that encourage technological innovation

There is a role for policymakers to play in ensuring that legislative frameworks allow for anti-fraud innovation and for tools necessary to counter the misuse of synthetic media. The latest wave of AI innovation has the potential to revolutionize the way that governments, businesses, and online platforms identify scammers and take action against fraudulent content. For example, opportunities to explore how AI could help better identify and stop instances of personal data leaving networks in an unauthorized way. Our teams at Google are embracing this transformative technology so that we can better keep people safe online, and AI has enabled us to accelerate detection of frauds and scams, scaling our approaches towards fighting abuse and harm on our platforms.

Unlocking the potential of these advanced technologies requires that policymakers, companies, and civil society work together to invest in Responsible AI - including the infrastructure, the workforce, and the legal framework needed - which we outline in Google's AI Opportunity Agenda. There is a risk that some regulatory approaches could block innovators and governments around the world from leveraging AI to achieve societal benefits such as economic opportunity, medical progress, and improved online safety. So Governments should also encourage stakeholders to innovate in - and utilize - AI in combating fraud.

Investments in Responsible AI should include mechanisms to help people identify fraudulent content that has been generated by AI for the purpose of scamming consumers. At Google, we have policies, across our products and services, that inform how we approach deceptive manipulated media and other forms of harmful AI-generated content; we are working both within Google and across industry to develop tools such as SynthID to ensure people know when an image or video is AI generated; and since simply asking "Is this generated by AI?" does not suffice in assessing content trustworthiness, we work closely with leading information literacy experts around the world to incorporate the latest research and help ensure that our products are empowering users with information and tools to cross-check what they find online.

Governments can play an important role in addressing risks of synthetic media by continuing investment in research and partnerships among organizations in areas such as development of standards and best practices, coordination on technical tools, and information sharing. Governments also play an important role as end users of technology, and therefore are uniquely positioned to help develop, adopt, pressure test, and ultimately encourage wider adoption of provenance technology like metadata and watermarking, while preserving the privacy and freedom of expression of users. For example, governments can encourage wider adoption of provenance technologies like the Coalition for Content Provenance and Authenticity (C2PA)'s technical standard, Content Credentials. C2PA, which brings together a range of organizations across the technology, advertising and broadcasting sectors, aims to provide publishers, creators, and consumers with opt-in, flexible ways to understand the authenticity and provenance of different types of media.

3. Invest in education & protection for users

Launch public education campaigns about online scams and how to protect oneself

Empowering individuals with knowledge is a key and effective tool in addressing fraud, and is most effective if done as a joint effort with other organizations, across sectors. Just as raising awareness has been seen as a key and effective public policy tool in addressing threats such as spam and phishing, it remains an essential pillar in the fight against scams and fraud: informed people are simply less likely to fall prey to fraudsters.

Government agencies, consumer protection groups, and businesses can all play a role in educating customers about safe online practices. An example would be campaigns to educate the public about online scams and how to protect themselves, like the Google co-created scamspotter.org : teaching people about common scam tactics (like phishing, romance scams, investment scams) and how to spot warning signs; and encouraging people to verify information and be cautious about sharing personal details online. At Google, we have initiatives in this space in partnership with local actors, for instance, DigiKavach in India and Project PRAISE in Singapore, and we regularly publish advice for users in our products and blogs on how to spot scams, and what to do if you encounter one.

Develop policies and product features that effectively protect users against harm

We know that our work is never done: scammers are constantly seeking to take advantage of new trends and technologies to perpetrate and extend the reach and sophistication of their criminal activities. That is why at Google, we are actively monitoring these evolutions, and constantly evolving our acceptable use policies, and creating new policies and product features to get ahead of these developments (as was detailed in the previous sections of this paper). An example is the new levels of protection provided in Google Messages thanks to the adoption of the Rich Communication Services (RCS), and others also adopting the RCS standard would help greatly reduce scam text conversations.

We encourage actors across the ecosystem to similarly introduce and constantly enhance their own policies and in-product protections to fight scams. We look forward to collaborating with a broad range of stakeholders in this perspective, from sharing insights into scams to jointly raising awareness and deploying the latest anti-scamming technologies.

Provide clear and accessible reporting channels for consumers

Governments should make reporting to official channels clear and accessible for victims and witnesses of fraud and scams. This includes, in some cases, streamlining existing or duplicative reporting channels, and removing barriers to access by victims. A clear reporting channel, that is both usable for victims and provides actionable information for government and law enforcement, only makes efforts in this space more meaningful.

Expert stakeholders are also encouraged to join Google's Priority Flaggers Programme in order to enhance reporting of scams to Google. Through this programme, we provide dedicated channels for participating organizations to notify us of potentially harmful content on our products and services that may violate our policies and Community Guidelines, including scams and fraud. This program is most suitable for organizations, such as NGOs and government agencies, with an identified expertise in recognizing and fighting harm online. Issues reported by Priority Flaggers are not automatically removed or subject to any special policy treatment. They receive the same standards of review and due process applied to use flags. However, because of their high degree of trust and expertise, our teams prioritize flags from Priority Flaggers for review.

Pursue secure digital transformation

Organizations across all sectors of our society should be encouraged to develop and deploy secure technology. Google constantly reinforces its protections and processes to prevent, detect and respond to scams. Similarly, online platforms, financial services, and telecom operators should build safety-by-design into their products, as well as create and regularly update and reinforce anti-fraud policies applying to their particular business.

In doing so, organizations should leverage the latest cybersecurity defenses and good practices, such as newer AI-based tools, which hold the promise of fundamentally enhancing the defense against cyber threats including scams. The best way to keep up with the threat landscape and advances in security technology is through secure digital transformation, particularly the adoption of cloud-based platforms, modern operating systems and hardware, and zero-trust principles (a security model used to secure an organization based on the idea that no person or device should be trusted by default, even if they are already inside an organization's network).

Additional good practices for all organizations include:

- protecting your accounts, such as by using unique and strong passwords;
- deploying good practices for IT admins, such as 2-Step Verification and adding recovery information to accounts;
- deploying auto update for apps and internet browsers;
- clear communication about businesses' official channels (website, customer support phone line) so as to counter misrepresentation by fraudsters; implementing secure payment systems;
- using passkeys to make login easier and more secure; etc.
- More detailed suggestions on how businesses can protect their information, for example you can visit our dedicated information resources.

Conclusion

Addressing fraud and scams requires cooperation among a range of actors across society. At Google, we have not waited to act. Every day we protect billions of people from a range of threats, checking 5 billion devices, 200 billion apps and 1 billion passwords while blocking 100 million phishing attacks and 15 billion spam messages.

The core of our mission is helping users find reliable and authoritative information. It is very much in Google's business interest to do the right thing. Our business is heavily dependent on the proper functioning of the online ecosystem, and the continued trust of users in that ecosystem. If consumers abandon the web due to bad online experiences, the long-term viability of Google's core business is at stake. This applies across our products, whether it's helping users navigate the open web through Search, or find local businesses in Maps.

Information integrity is an ongoing challenge, where bad actors operate at scale, constantly adapt their methods, and combine online and offline activity to avoid detection. No system will ever be perfect even with the best defenses in place.

But we are committed to constant improvement to protect our users and help them find information they can trust. It is important to be nimble, tracking bad actors' behavior and learning from it. In doing so, we are able to better prepare for future fraud and scams that may arise.

These efforts will only be stronger, and more effective and long-lasting, if we join forces with other stakeholders - from governments to other industry actors and civil society - to address these threats together.

Annex one

AI and scams

Google has been responsibly developing AI for more than two decades. We have also long been a leader in content responsibility, and we are applying the same ethos and approach as we launch new products powered by Generative AI technology. AI can play a very positive role in fighting fraud and scams: it enables us to accelerate abuse detection and action, scaling our approaches towards fighting abuse and harm on our platforms:

a) Protect from harmful content

We build generative AI responsibly and use AI to power detection and removal of harmful and illegal content at scale, with the 3 pillars of Prevent - Detect - Respond:

(i) Prevent: We've created generative AI prohibited use policies outlining the harmful, inappropriate, misleading or illegal content we do not allow. This includes generating content for deceptive or fraudulent activities, scams, phishing, malware, or content intended to misinform, misrepresent or mislead. In developing these policies, we leveraged our investment in review teams that are located in countries around the world, are fluent in multiple languages, and carefully evaluate flagged content 24 hours a day, seven days a week.

We use an extensive system of classifiers to detect and prevent content that violates our policies for Generative AI products. Training data is filtered for high-risk content and to ensure all training data is sufficiently high quality. We also apply classifiers to both the user's prompts and potential outputs from the model. The model proposes multiple response candidates, and classifiers are applied to those candidates, rating them on certain parameters, including safety. This is all done quickly and seamlessly under the hood, without the user realizing it.

If we identify a violative prompt or output, our products will not provide a response. We may also direct the user to additional resources (like a helpline) for help on sensitive topics such as those related to dangerous acts or self harm.

(ii) Detect: Recent advances in large language models (LLMs) expand our ability to find abusive content that violates our policies even more quickly and at scale, across our platforms and services. Using LLMs, we can rapidly build and train a model in a matter of days - vs weeks or months - to find specific kinds of abuse on our products. This is especially valuable for new and emerging abuse areas, or for nuanced scaled challenges, like detecting counterfeit goods online. We can prototype a model that expertly identifies this new kind of abuse and automatically routes it to our teams for enforcement.

For example, Google's Trust & Safety teams have been leveraging LLMs to efficiently increase our detection of ads promoting get-rich-quick schemes, which violate our policy against Unreliable Financial Claims. The bad actors behind these types of ads are sophisticated; they adjust their tactics and tailor ads around new financial services or products, such as investment advice or digital currencies, to scam users. The fast-paced and ever-changing nature of financial trends make it, at times, harder to differentiate between legitimate and fake services, which impacts our ability to quickly scale our machine learning and automated enforcement systems to combat scams. LLMs are more capable of quickly recognizing new trends in financial services, identifying the patterns of bad actors who are abusing those trends and distinguishing a legitimate business from a get-rich-quick scam. This has helped our teams become even more nimble in confronting emerging threats of all kinds.

We also use AI at each stage in the lifecycle to enrich the data with detection information, extract critical details from unstructured data, normalize and categorize that data, prioritize the intermediate outcomes, and give the intelligence necessary to proactively defend against emerging threats.

Combining the capabilities of hyper-focused analysis and global visibility of Google's family of services like Mandiant and Google Cloud gives us the ability to identify attacks, while simultaneously gaining tremendous insight into malicious actors' tactics, techniques, and procedures (TTPs). Not only does the collected data form the foundation of our threat intelligence products, it also allows us to train our AI models with high-quality examples, which in turn leads to a virtuous cycle of better detections and improved outcomes.

Although we are still testing these new techniques in abuse fighting, they have demonstrated impressive results so far, and will be a major advance in content moderation and our effort to proactively protect our users at scale, especially from new and emerging risks.

(iii) Respond: Whilst AI is an important tool in identifying potentially violative content at scale, human review remains a critical part of the process. Once content has been flagged for review, it is sent to human reviewers who assess educational, documentary, scientific, artistic or journalistic context and nuance.

The context in which a piece of content is created or shared is an important factor in any assessment about its quality or its purpose. Across Google, around 20,000 trained reviewers from around the world work to detect, review and remove content that violates our policies in dozens of languages. These signals bring the necessary local context and expertise.

Combined with the ability of automated systems to work quickly and at scale, this process also means problematic content is often removed before it is widely viewed, or viewed at all, reducing the amount of harmful content human reviewers are exposed to.

b) Helping users identify AI-generated content

We're committed to connecting people with high-quality information and upholding trust between creators and users across society. Part of this responsibility involves giving users more advanced tools for identifying AI-generated content. With more people using artificial intelligence to create content, we are building on the ways in which we help our audiences identify AI-generated content through several new tools and policies:

- Watermarking with SynthID

  Our SynthID toolkit watermarks and identifies AI-generated content. These tools embed digital watermarks directly into AI-generated images, audio, text or video. In each modality, SynthID's watermarking technique is imperceptible to humans but detectable for identification. The toolkit is currently launched in beta and continues to evolve. It's now being integrated into a growing range of Google products, helping empower people and organizations to responsibly work with AI-generated content. And while SynthID isn't built to directly stop motivated adversaries like cyberattackers or hackers from causing harm, it can make it harder to use AI-generated content for malicious purposes.

- Content labels on YouTube

  Generative AI is transforming creativity on YouTube, which means that we need to transform the ways we are maintaining a transparent and healthy ecosystem for both creators and viewers. That's why we rolled out a tool that requires creators to share when the content they're uploading is meaningfully altered or synthetically generated and seems realistic. This tool lives in the existing upload flow, making it easy for creators to add this disclosure. We apply transparency labels to signal to users that they are watching this type of content. For most videos, a label will appear in the expanded description, but for videos that touch on more sensitive topics—like health, news, elections, or finance—we will also show a more prominent label on the video itself.

- About this Image on Search

  In 2023, we announced a new feature called About this image, which gives people an easy way to check the credibility and context of images they see online including whether they carry a SynthID watermark. This feature can now be accessed more intuitively via the "circle to search" feature on Android phones. Users can see metadata - when available - that image creators and publishers have added to an image, including fields that may indicate that it has been generated or enhanced by AI. All Google AI-generated images have this markup in the original file.

- Elections ads disclosures

  In 2023, Google was the first tech company to require election advertisers to prominently disclose when their ads include realistic synthetic content that's been digitally altered or generated, including by AI tools.

c) Partner to create a safer web

We know that building AI responsibly must be a collective effort involving researchers, social scientists, industry experts, governments, creators, publishers and people using AI in their daily lives. We use our technical expertise to develop and share tools to help other organizations detect and remove abusive and harmful content from their platforms. This includes datasets of deep fakes, to help others detect AI-manipulated content and build AI responsibly.

No one company can progress this approach alone. We must work together for progress and we are committed to working in partnership with industry, civil society, and academia to get it right.

- C2PA: We recently joined the Coalition for Content Provenance and Authenticity (C2PA) coalition as a steering committee member. C2PA is a cross-industry effort to help provide more transparency and context for people on AI-generated content.

- Frontier Model Forum: Together with Anthropic, Microsoft and OpenAI, Google launched the Frontier Model Forum to ensure the safe and responsible development of frontier AI models.

- Partnership on AI: We joined the PAI, as part of a community of experts dedicated to fostering responsible practices in the development, creation, and sharing of media created with generative AI.

- MLCommons: We are part of MLCommons, a collective that aims to accelerate machine learning innovation and increase its positive impact on society.

Other initiatives include launching the AI Cyber Defense Initiative in February 2024 to leverage artificial intelligence (AI) to boost cybersecurity and to reverse the "Defender's Dilemma". This will include a series of cybersecurity seminars, and applied research projects into threat detection; malware analysis; vulnerability detection and fixing; and incident response. Among other efforts, we are also open-sourcing Magika, a new, AI-powered tool to aid defenders through file type identification, an essential part of detecting malware - a major enabler of scams and fraud. The potential is enormous. For instance, already Gmail uses RETVec, a multilingual neuro-based text processing model which improved spam detection by 40%.

We also undertake ourselves and also commission or support others doing research to understand and counter scams more effectively. For instance, in 2024 the Safer Internet Lab (SAIL), a Google-supported collaboration with the Jakarta-based Centre for Strategic and International Studies (CSIS) to conduct research around misinformation, launched a new workstream to examine how fraudsters targeting the South East Asia region can manipulate synthetic media, how that impacts the public, and what possible solutions may look like.

Misuse of AI capabilities

As concerns like the use of deepfakes in fraud make clear, the increasingly wide availability of AI tools means that they can also be misused to facilitate a variety of malicious activities. New generative AI tools enable the creation of content – text, images, speech – at unprecedented speed and scale. While these tools can fuel creativity and human understanding, they can also be used to create materials that are harmful to individuals and societies. Such technologies have the potential to significantly augment malicious operations in the future, enabling threat actors with limited resources and capabilities.

However, so far based on our own observations and open source accounts, adoption of AI and effective operational use by malicious actors remains limited and primarily related to social engineering. In the future, we can see that malicious actors with limited resources and capabilities could misuse AI to produce, and disseminate, higher quality content at scale. Models could be used to create content like articles or political cartoons in line with a specific narrative or to produce benign filler content to backstop inauthentic personas.

Hyper-realistic AI-generated content may have a stronger persuasive effect on target audiences than content previously fabricated without the benefit of AI technology, and we are already seeing evidence of this activity with the use of synthetic media ('deepfakes') for example reproducing the image of trusted individuals to promote scammy content.

The quality and therefore 'believability' of fraudulent content can also be enhanced by the misuse of AI techniques. Threat actors can use LLMs to generate more compelling material tailored for a target audience, regardless of the threat actor's ability to understand the target's language. LLMs can help malicious operators create text output that reflect natural human speech patterns, making more effective material for phishing campaigns and successful initial compromises.

An important part of introducing this technology responsibly is anticipating and testing for a wide range of safety and security risks, including the rise of these new forms of AI-generated media. While this technology has useful applications - for instance, by opening new possibilities to those affected by speech or reading impairments, or new creative grounds for artists and movie studios around the world - we are conscious of how it can be misused for malicious purposes such as fraud and scams.

In collaboration with a range of stakeholders, we are working on three key pillars to mitigate the risks of such technologies as 'deep fakes':

- *Detection* – using machine learning to develop systems that are able to detect synthetic and/or manipulated media at scale.
- *Provenance* – in line with our effort to deliver reliable information, making it possible for users to understand the history of a piece of online content, either by means of a chain of authentication (e.g., digital signatures guaranteeing that a piece of content hasn't been tampered with since its capture on a camera) or by providing context on the online history of a piece of content (e.g., helping users understand other contexts in which an image or video appeared online in the past).
- *Media literacy* – helping society become more mindful of image and video as authoritative proof that something happened, and more attuned to elements of context that may indicate that a piece of content may be trustworthy.

Next steps in using AI for good and addressing its misuse

Google has been working to counter the misuse of AI for several years. For example, we have been researching and developing tools to counter the misuse of 'synthetic media' (deepfakes), and we are increasingly working with the wider Internet ecosystem to share good practices and information on these threats. There is a role for policymakers to play here in ensuring that legislative frameworks allow for anti-fraud innovation and tools necessary to counter the misuse of synthetic media.

We are also optimistic that AI will prove a major help for 'defenders' against malicious acts including scams and fraud. The security community has an opportunity to outpace threat actors with defensive advancements, including generative AI adoption, to benefit practitioners and users.

We strongly believe that AI could have a decidedly positive effect in fighting fraud and scams, as already demonstrated in how LLMs have boosted our Ads Safety efforts in 2023. We are making strides in dealing with fake content: for example thanks to our evolving machine learning algorithms, we blocked or removed over 170 million policy-violating reviews on Google Maps from this year — over 45% more than in 2022. More than 12 million fake business profiles were removed or blocked too.

Annex two

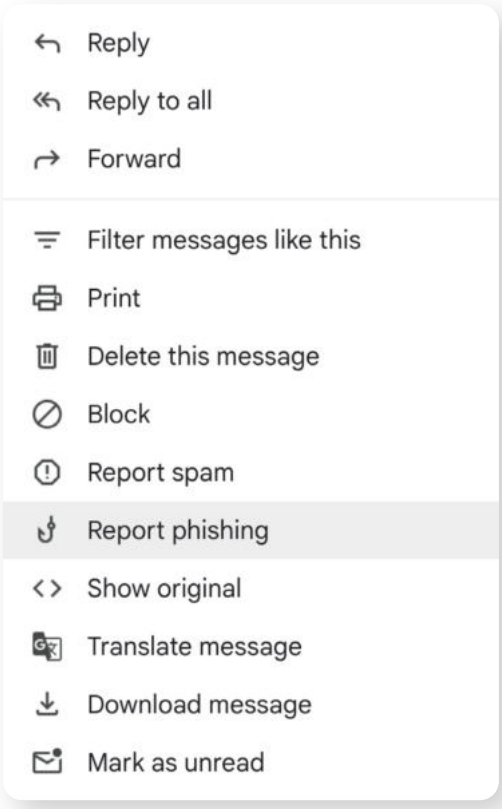
How we support users

Annex 2.1 How users can report fraud and scams

Google's products enable consumers to report inappropriate content and behaviors. For example, users can now more easily report an ad if they believe it violates one of our policies by clicking on our "about this ad" menu, which is available directly in the ad surface. When an ad is reported, it will be reviewed for compliance with our policies and will be taken down if appropriate.

On Google Shopping, if a user sees something that doesn't look right or seems out of sync with our policies, like a very high price or a violent weapon, they can report it to us, using the "report a listing" link on the bottom right of a product page, or they can report content through our Help Center.

It's important to provide this "in-product" experience to make this as easy as possible for the user.



How users can report when they encounter threats:

here the example of Gmail concerning a phishing attempt

Annex 2.2 Empowering users

A major way to address the threat of scams and fraud is to empower users. We do this through in-built product protections, as well as in-product information. This includes in-product disclosures and other pieces of information that users can benefit from to make an informed and educated decision, such as the Ads Transparency Center, My Ads Center, Advertiser Verification, Financial Services Advertiser Certification, etc.

Additionally, we feel it particularly important to empower users by raising their awareness of the risks they may be exposed to, and how they can look after themselves. Analysis confirms the usefulness of users being more aware, as they find it difficult to discern scams from genuine interactions: one third of users report getting lured by attractive offers, while another third report being unable to identify deceit or lack of knowledge to identify scams (according to the GASA Global & APAC Surveys, 2023).

We endeavor to create user awareness against frauds through multiple channels like print media and national level ad campaigns, as well as in-product warnings. These efforts need to target all age groups and communities, as all are susceptible to being targeted by fraudsters.

We also realize that traditional education on its own may not be sufficient in uplifting resilience to scams and fraud: people who are simply given more information may not always change their behavior. This is why we have been piloting so-called 'inoculation' methods that help users build 'mental antibodies' to scams. Indeed, a recent study with a US audience has shown inoculation styled interventions to significantly improve user's scam discernment to impersonation scams, without harming their trust in legitimate communications (Source: Robb & Wendel, 2023).

We have run awareness raising campaigns ourselves, but find it particularly useful to do it also in partnership with others, notably in order to extend the reach of our preventive messages, and their relevance to local communities. Our marketing team has developed a user education toolkit which we are keen to amplify not just through our own channels but also through cooperations with the likes of financial institutions and industry bodies.

Here are some of many examples of these partnerships for education and awareness:

In Thailand, we worked with the Bank of Thailand and Thailand Banking Sector Computer Emergency Response Team (TB-CERT) to launch the #31days31tips online safety campaign. This campaign shared daily safety pointers to help Thai people better secure their online accounts, spot financial scams and keep their private information safe.

In Australia, we have worked with Australian consumer network ACCAN over the past year on a gift card scam campaign (available through ads and YouTube content). We also partner with the government agency ACCC Scamwatch team every year for Scam Awareness Week, which in particular involves promoting a Security Checkup (Scamwatch is a member of the Google Workspace Priority Flagger program, which enables them to warn us of policy violations through a dedicated channel). The Security Checkup is also promoted approx 2-3 times a year (outside of Scam Awareness Week); these promotions are seen by approximately 11 million Australians and receive strong engagement from viewers.

In Singapore, we pledged to support Project PRAISE — an initiative with RSVP Singapore: The Organization of Senior Volunteers and the Singapore Police Force. This project trains volunteers to raise awareness of scams among seniors, a population particularly vulnerable to cyberattacks, through a series of workshops trained to help people spot scams.

In Hong Kong, we supported the Be a Smarter Digital Citizen program by the Hong Kong Council of Social Service to improve students' digital literacy and online safety awareness. We shared practical tips for securing online accounts, including enabling 2-step verification, safer search experiences with SafeSearch and Safe Browsing, and AI-enabled anti-phishing features across Gmail and Chrome.

India: Meanwhile, Google.org, Google's philanthropic arm, supported the CyberPeace Foundation with a new grant of US$4 million for a four-year nationwide awareness-building program in India and a multilingual digital resource hub to help nearly 40 million underserved individuals build resilience against misinformation. This is on top of the major 'DigiKavach' initiative.

APAC-wide: Over the past five years, Google.org has supported 26 social impact organizations in Asia Pacific with over US$35 million in grants. These grants have helped train vulnerable people to stay safer online, tackle misinformation, and enhance cyber resilience for organizations.

In the United States, Google partnered with the Cybercrime Support Network, spotting the most common patterns used by scammers and offering practical advice on the website scamspotter.org to help users stop them in their tracks.

Supporting user resources:

- Google blog: How to spot scams, and what to do if you encounter one
- Avoid and report scams - Google Help
- Mobile Security & Privacy - Android Safety Center
- Google Play Protect - Android
- Use Google Play Protect to help keep your apps safe and your data private
- Learn about Pixel security certifications on Android
- Android ecosystem security transparency report
- Safety Center - Emergency Help - Android

Annex three

Approach to tackling scams across specific Google products

Annex 3.1. Android

Building on decades, if not centuries, of letter-based fraud, fraudsters started using phone calls and text messages to commit their crimes, gradually adding online communications tools. This is why we have been developing a range of product features aimed at mitigating risks of fraud and scams for users when using Google-supported mobile devices running Android.

Android is an operating system (OS) that powers billions of devices worldwide. You can think of it like the software that runs your phone. This mobile operating system is based on a modified version of the Linux kernel and other open-source software, designed primarily for touchscreen mobile devices such as smartphones and tablets.

Android's top priority is the safety of its users. This responsibility is not taken lightly. We use industry-leading security practices and work closely with developers and device implementers across the entire ecosystem to ensure users are protected as soon as they power on their device.

No platform keeps more users safe on their devices – Android is incorporated on over 24,000 distinct Android devices, and Google Play Protect is active on 3 billion user devices. It stretches beyond the Play Store to protect users from malware in apps downloaded from the Play Store and third party stores/sites.

Android meets the toughest security standards in the world – we've attained the highest mobile industry certification standards, including the US Department of Defense and the Common Criteria recognized in 31 countries.

Android's approach to security focuses on three core pillars:

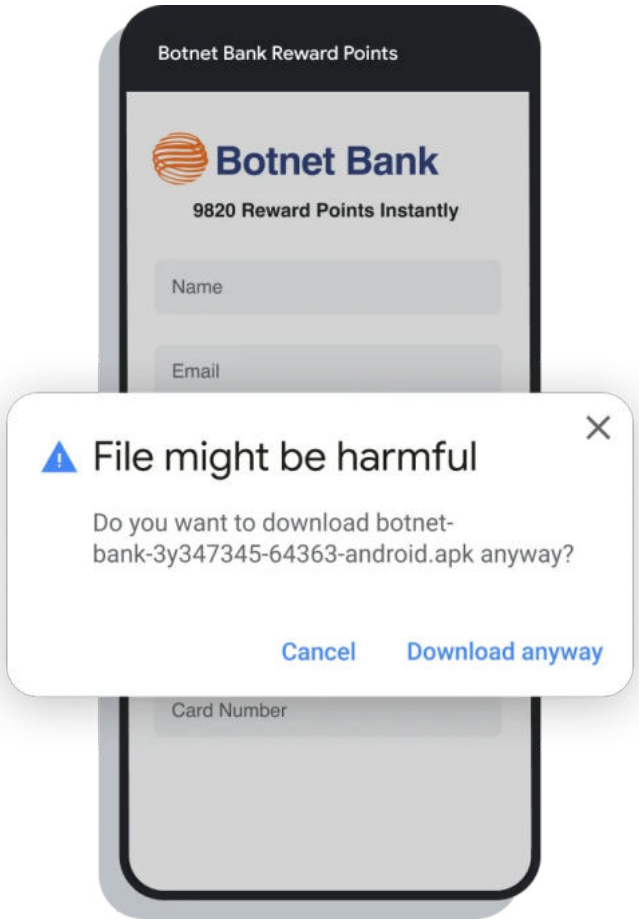
- Multi-layered: Each part of the Android ecosystem works together to build a strong defense that runs smoothly and effectively.
- Transparency: We work with the security research community to uncover, fix, and validate security issues. Once an issue is addressed we share that with the world to ensure transparency and help others
- Cross-Google technology: We leverage Google's security expertise and incorporate leading security features into Android's OS, the Play Store and apps on the device.

Our teams are dedicated to combating fraud, specifically focusing on cases where victims are targeted remotely through channels like email, phone calls, and messaging apps. Criminals exploit various attack vectors to carry out their schemes, including malware distribution, permission abuse, screen sharing, and social engineering tactics, such as phishing. We've built protections against these attack vectors into the core operating system, and we layer on additional security services that continually scan devices for malware and other harmful behavior.

Annex 3.1 (i) Android phone and messaging

For its main communications-enabling features of phone and messaging, Android incorporates multiple layers of protections, including:

- Phone by Google which helps protect against voice phishing and scams with built-in caller ID, spam protection and Call Screen by blocking dangerous calls and warning you about suspicious callers.
- Google Messages provides built-in spam and phishing protection that warns users and automatically filters suspected spam and unsafe websites, using AI to spot suspicious messages by assessing the reputation of the sender, looking for known patterns and dangerous links.
- Chrome download warnings that alert you if you're about to download an Android (APK) file, ensuring you're aware a link is about to trigger a download of an app.
- Every Pixel device comes with caller ID and spam protection and we help users to identify potential scams with verified SMS for messages which shows users the business name and logo as well as a verification badge in the message thread.
- We are also looking at leveraging AI tools to support enhanced scam call detection on Android.



RCS - Rich Communications Suite: improving on SMS especially to counter spam and scams

Traditional phone text messages, also known as 'SMS' (for 'short messaging service') are not only an outdated form of communication by now, but also a flawed system - especially when it comes to phishing and scam detection. Because the SMS network is decentralized, there is no way that anyone can identify which networks are trustworthy - and which aren't. This lack of trust along the travel of the message from creation to reception by the user limits the effectiveness of SMS fraud solutions, which is why users get so many phishing attacks and scam messages via SMS.

To combat these abuse vectors, spam and abuse prevention methodologies include a combination of firewalls that block messages from suspicious sources, message content scanning by carrier / aggregator SMS platforms, and business commercial terms that require connected aggregator and interconnected carrier parties to operate lawfully. However, owing to the decentralized topology, these methods cannot be applied with sufficient consistency and rigor to prevent fraud at scale, and they can also be inflexible and slow in addressing new threats.

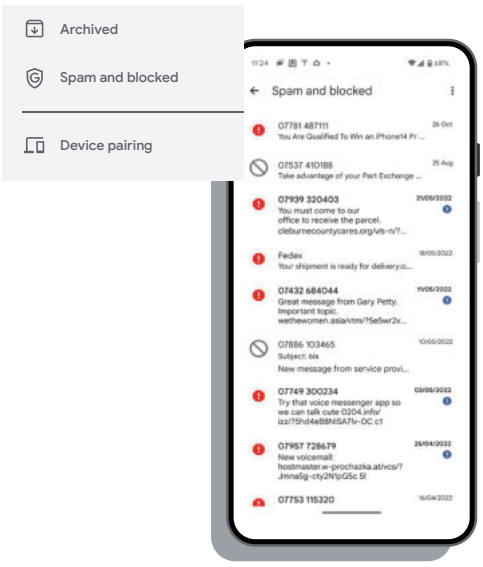
The new messaging system known as Rich Communications Suite or RCS offers much better fraud detection: RCS, through carrier platform choice, has evolved to the point where it is now a centralized system, so we can implement these methodologies consistently and deploy new countermeasures quickly to address new threats. This converged system allows a single platform to authenticate users, police traffic and identify suspicious traffic patterns. For example: when a new user sends several international messages these are likely scam texts, and will be labeled as such for the end use.

Crucially, RCS enables a safer messaging framework, comprising verified business to consumer messages, and interpersonal messaging with extensive platform and client malicious message detection and blocking features.

RCS, thanks to the combination of verified business messaging and interpersonal messaging protections greatly reduces the amount of scams and phishing attacks prevalent in SMS today through:

1. Interpersonal Messaging Scam Detection - RCS can detect unusual traffic and block it before it is delivered

- The common platform authenticates new users and their reputation.
- The common platform is able to block messages in-transit and temporarily suspend abusive phone numbers due to spam.
- The common platform will provide spam notices to connected clients based on user reputation and behavior.
- Platform notices, and on-device protections, may either result in messages being autoblocked or trigger a warning banner.
- RCS also allows for trusted parties to file spam texts into a spam folder, so they never reach the user's inbox.
- In Google Messages, the spam detection feature works for both SMS and RCS messages.



We display warning banners and autoblock more than 2 billion potentially harmful messages monthly across SMS, MMS and RCS channels.

2. Verified Business Messaging - RCS can verify businesses, giving users trust in who they're messaging

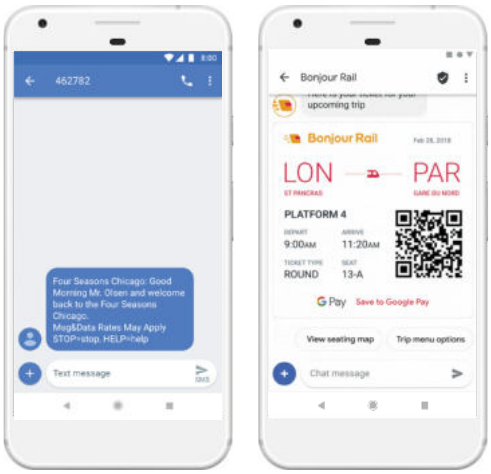
- Over 2 trillion messages are sent from businesses to users every year (over 30% of global SMS traffic).
- Users are often tricked into responding to scam messages when they pose as businesses. Being able to ascertain that a business is genuine is therefore very important.
- RCS Business Messaging (RBM) requires businesses to be verified by the carrier service provider.
- Once verified, businesses will get a verified "check mark", instilling user confidence and trust in the sender.
- This allows users to distinguish between authentic and fake business profiles (between messages sent from unverified short codes over SMS, or long numbers over SMS or RCS, and legitimate businesses).
- The RCS/RBM feature has the potential to transform how businesses communicate with their customers using messaging to build stronger, trusted relationships through brand verification, upstream business (agent) verification, content approval, active traffic management tools, and richer interactive experiences.

SMS

No visible verification

No business information

Distributed platform policy enforcement



RBM

Business verification

Business information

Consistent platform policy enforcement

In short, RCS offers a much safer and effective messaging experience for users. While SMS is built on an outdated network, which allows bad actors to abuse the system, by contrast RCS is built on secure IP data connections and, as a converged system, offers a vast improvement for scam and phishing detection. RCS also serves as the foundation for key user safety features, like spam filtering and business verification. Going forwards, adopting the RCS standard would greatly reduce phishing attacks and scam text conversations.

Annex 3.1 (ii) Android and Google Play

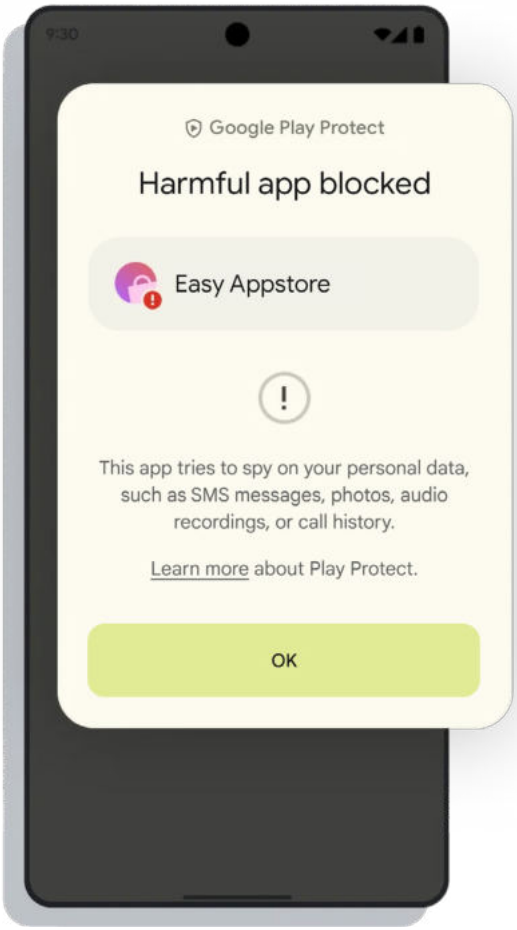
Google Play is an online store where people go to find their favorite apps, games, movies, TV shows, books and more. It provides over two million apps and games to billions of people around the world who use Android-powered devices.

Because it is the leading mobile application store globally, fraudsters have sought to exploit Play, for instance through fake applications available for download on third party websites, that purport to be from genuine sources, or purpose-built apps that are built so as to infect a user's device and steal their information.

In order to try and avoid detection, cybercriminals are now leveraging sophisticated techniques such as polymorphic malware, which can change its identifiable features. They're turning to social engineering to trick users into doing something dangerous, such as revealing confidential information or downloading a malicious app from ephemeral sources – most commonly via links to download malicious apps. In order to combat the evolving modus operandi, we are continuing to invest in strengthening the safety and security of the Android ecosystem.

Prevent: Technical and product protections

We continually strive to enhance the safety and security of Android, while also providing tools and support for organizations to harden the security of their official apps so they become more resilient to malware attacks, such as by sideloaded apps. We highlight below some of the key protections we have deployed.



Play Integrity API: App access risk: enhancing the security of official apps is particularly important when it comes to financial institutions. For example in Brazil, India, Singapore, and Thailand, Google is supporting banks and fintechs to strengthen the security of their official apps and make them more resistant to malware attacks by malicious apps. We have invited top banks and fintech developers to join the Early Access Program for integrating Play Integrity API App access risk feature into their apps. Developers in this early access program can add app access risk to their API response, which already contains the device integrity, application integrity, and account license verdicts. App access risk can help you detect whether another app on the device could be used to access or control your app.

Google Play Protect - All the apps in the Google Play Store undergo rigorous security testing before they're approved. Our machine learning system scans up to 200 billion apps each day, continuously working behind the scenes to keep your device, data and apps safe from malware and unwanted software. This is the most widely deployed mobile threat protection service in the world.

We continue to enhance our machine learning systems and review processes, and in 2023 we prevented 2.28 million policy-violating apps from being published on Google Play in part thanks to our investment in new and improved security features, policy updates, and advanced machine learning and app review processes.. We also continued in our efforts to combat malicious and spammy developers, banning 333,000 bad accounts from Play in 2023 for violations like confirmed malware and repeated severe policy violations (more details in our blogs on how we fought bad actors and ways to protect Android users from fraud).

In October 2023, we announced that Google Play Protect's security capabilities would be made even more powerful with real-time scanning at the code-level to combat novel malicious apps. As of Feb 2024, as a result of the real-time scanning enhancement, Google Play Protect has identified 515,000 new malicious apps and issued more than 3.1 million warnings or blocks of those apps.

Risky Permissions is an area whereby Google has identified device settings permissions that are frequently abused in scam attempts (e.g., reading or sending text messages, accessibility controls). Apps that request these permissions from users can be flagged for user attention or even blocked directly in risky situations.

Combating sideloaded financial fraud (pilot in Singapore): Based on our analysis of major fraud malware families, we found that over 95% of apps working to exploit sensitive permissions (RECEIVE_SMS, READ_SMS, BIND_Notifications, and Accessibility), are via apps sideloaded onto Android devices, meaning that they are not downloadable by users from preloaded app stores but through independent websites. As Google has invested in improving the safety and security of the Play Store, it is increasingly harder to conduct fraud via Play-distributed apps.

This is why, as a further example of our efforts to further safeguard consumers against financial fraud and scams, starting in 2024 Google teamed up with the Singapore Government (MCI & CSA) on a global-first initiative to combat Internet-sideloaded financial fraud and scam apps on Android. This partnership enables us to identify and remove potentially fraudulent apps in a more timely manner, and jointly expand user education to help people better protect themselves against bad actors. This pilot initiative is an extension of Google's existing Play Protect feature that helps protect consumers against fraudulent and malicious apps. Play Protect is constantly improving its detection capabilities with each identified app, allowing us to strengthen our protections for the entire Android ecosystem.

Android Dialer provides additional protections in relation to messages and dialer by notifying users when they may be getting scammed, and we will be launching a program in Play to do store sweeps looking for elder scams in apps.

Play Store Government Badge: As part of our latest efforts to help keep Google Play safe, in 2024 Google Play introduced a new badge to help users identify official federal and state government apps globally. Government apps are often targets of impersonation due to the highly sensitive nature of the data users provide, giving bad actors the ability to steal identities and commit financial fraud. Badging verified government apps is an important step in helping connect people with safe, high-quality, useful, and relevant experiences, and we're already exploring ways to build on this work.

Detect and Respond

As one example of a use policy tailored to Play apps, we're working on ensuring that financial services apps do not expose users to deceptive or harmful financial products and services. This is why we set up a privacy policy that comprehensively discloses the access, collection, use and sharing of personal and sensitive user data, subject to the restrictions outlined in this policy.

Personal Loans Permissions: In April 2023, we updated our Personal Loans policy to state that apps aiming to provide or facilitate personal loans may not access sensitive data, such as user contacts or photos.

Annex 3.2 Chrome

Chrome is the official web browser from Google, built to be fast, secure and customisable. It is secure by default: Its systems are automatically updated every four weeks to ensure that the security features and fixes can protect users from dangerous and deceptive sites and warn them if credentials have been compromised. We also use our Safe Browsing technology, meaning that any time a user tries to access a site we have identified as dangerous, they will be shown a warning.

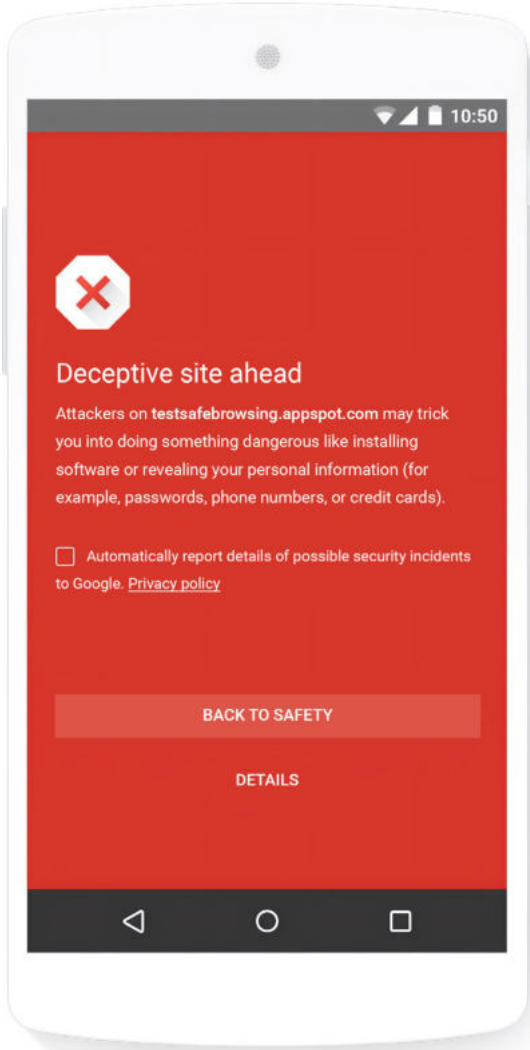
Prevent

Safe Browsing: Launched in 2005, today Google Safe Browsing protects more than five billion devices across the world, and provides more protection in cases where a link appears to be legitimate.

Google Safe Browsing warns users if it looks like a website is dangerous and is attempting to phish their credentials, upon receiving this warning. People can simply click on the "Go back to safety" option to avoid going to a malicious site or downloading a malicious file. We've also updated our machine learning models to specifically identify pages that look like common log-in pages and messages that contain spear-phishing signals.

Google makes this technology freely available to other browsers and Internet companies and it is deployed in multiple, competing browsers in addition to Chrome (e.g. Firefox, Safari) and across many different platforms, including iOS and Android.

On Google Pixel mobile phones, Safe Browsing technology helps protect the Pixel device from phishing attacks by showing the user Stop Sign warnings any time that they attempt to navigate to dangerous sites or download dangerous files. Safe Browsing also notifies webmasters when their websites are compromised by malicious actors and helps them diagnose and resolve the problem.



We are deploying the business features of Safe Browsing in order to extend the potency of our approach to fight fraud. Web Risk enables users to detect if a URL is violating any Safe Browsing policies, evaluate how risky a URL is, and submit URLs to Google Cloud for scanning. If a URL violates any Safe Browsing policies, it will be added to the Safe Browsing blocklist that protects more than 5 billion devices within minutes. This provides a unique set of capabilities to protect end users against malicious attacks such as phishing and malware. As part of the range of partnerships we have with government agencies around the world, Web Risk will be integrated into the cybersecurity ecosystem led by Singapore's Government Cyber Security Operations Centre (GCSOC). GCSOC uses their own automated tools, such as PhishMonSG, to consolidate thousands of potentially malicious phishing websites daily from commercial feeds and crowdsourced initiatives such as ScamShield. After deduplication and normalization, they use the Google Web Risk Evaluate API to evaluate the risk level of a suspected website. Only sites deemed as risky are escalated for blocking via Web Risk's Submission API.

Additional protections for Chrome include tools we have developed that empower users to take more steps based on their own threat modeling and perceived risk, such as the Chrome Password Alert extension on our browser; Advanced Protection Program (APP), which provides the strongest account protection that Google offers through security keys; and Enhanced Safe Browsing.

- Chrome Password Alert: In the Chrome browser, we offer an extension that, when turned on by the user, alerts them when they've entered their Google credentials on a non-Google site (perhaps believing that they've entered those credentials on a legitimate Google site that is actually a phishing site). This alert about unusual sign-in helps inform the user that there may be a risk that a bad actor is phishing their account credentials - and it offers timely advice about how to mitigate harm. Password Alert also checks each page the user visits to see if it's impersonating Google's sign-in page, and alerts them if so.
- Advanced Protection: In 2017, we unveiled the Advanced Protection Program (APP), which provides the strongest account protection that Google offers. APP requires the use of security keys, which are resistant to man-in-the-middle attacks where the use of two-factor authentication codes can be phished. APP defends against targeted online attacks on Chrome with even more stringent checks before each download, and also offers extra protections on other Google products like Gmail and Drive.
- Enhanced Safe Browsing: In 2020, we launched Enhanced Safe Browsing protection in Chrome, a new option for users who want a more advanced level of security while browsing the web. After users turn on this feature, Chrome will share additional security data directly with Safe Browsing in order to enable more accurate threat assessments. For example, Chrome will check uncommon URLs in real time to detect whether the site you are about to visit may be a phishing site. It enables more advanced detection techniques that adapt quickly as malicious activity evolves. As a result, Enhanced Protection users are phished 20-35% less than users on Standard Protection.

Annex 3.3 Ads

Billions of people come to Google looking for accurate information, and our business depends upon their trust. Scams undermine that trust. The ad-supported Internet creates a value exchange for advertisers, publishers and users. Businesses large and small can more affordably advertise and can run ads with actionable measurement to accelerate growth. Publishers and content creators generate revenue from running ads next to their content and use measurement to demonstrate effectiveness and maximize their revenue. All of this enables user access to vast amounts of content and information like news, videos, social media, maps and directions – at little or no cost. Ads safety and integrity undergird this ecosystem, making it all possible. At Google, we are not just committed to a healthy ecosystem and the open web, it is critical to our business. Keeping our users safe is our top priority and we invest heavily in the enforcement of our policies. We have teams of thousands working around the clock to create and enforce our policies at scale.

The scams landscape is constantly evolving. Over the past several years, we've increasingly seen bad actors use sophisticated deceptive techniques, such as 'cloaking' to hide from our detection, or to run ads for phone-based scams where the deception takes place offline. We saw an uptick in opportunistic advertising and fraudulent behavior from actors looking to mislead users or deceptively earn money from ads following the COVID-19 pandemic. We have also seen adversarial actors target their campaigns across borders.

Generative AI presents both opportunities and risks. Good advertisers are using generative AI to optimize their campaigns and increasingly generate compelling ad creatives and landing pages. However, the same technology can be used by bad actors to create better looking ads with greater reach on faster timelines. For our part, we are also using generative AI to drive improvements in detection and enforcement.

We meet these challenges through innovation and evolution of our policies and products. We are investing in improvements in our policies, advanced machine learning and large language models (LLMs) as well as human review processes. Our systems take into account network signals, previous account activity, behavior patterns and user feedback. We also invest in technology to detect coordinated adversarial behavior, allowing us to connect the dots across accounts and suspend multiple bad actors at once. Finally, we continue to invest in advertiser verification, which underpins our ads transparency and safety-by-design architecture.

Detect and Respond: developing and enforcing robust ad policies

We detect fraudulent and scam ads through a combination of both AI and human reporting and evaluation. This process helps ensure ads on our platforms adhere to the strict policies we have in place, including policies against misrepresentation and enabling dishonest behavior. In addition, we make it easy for people to report scam ads if they see them.

We regularly review and update our policies to ensure we are protecting users, advertisers and publishers. For example, during the COVID-19 pandemic we enforced our sensitive events policy to prevent behavior like price-gouging on in-demand products like hand sanitizer, masks and paper goods, or ads promoting false cures. As we learned more about the virus and health organizations issued new guidance, we evolved our enforcement strategy to start allowing medical providers, health organizations, local governments and trusted businesses to surface critical updates and authoritative content, while still preventing opportunistic abuse. More recently, toward the end of 2023 and into 2024, we faced a targeted campaign of ads featuring the likeness of public figures to scam users, often through the use of deep fakes. When we detected this threat, we created a dedicated team to respond immediately. We pinpointed patterns in the bad actors' behavior, trained our enforcement models to detect similar ads and began removing them at scale. We also updated our misrepresentation policy to better enable us to rapidly suspend the accounts of bad actors.

Our safety teams have long used machine learning powered by AI to enforce our policies at scale (see the dedicated section on AI below). It's how, for years, we've been able to detect and block billions of bad ads before a person ever sees them. But, while still highly sophisticated, these machine learning models have historically needed to be trained extensively — they often rely on hundreds of thousands of examples of violative content.

LLMs, on the other hand, are able to rapidly review and interpret content at a high volume, while also capturing important nuances within that content. These advanced reasoning capabilities have already resulted in larger-scale and more precise enforcement decisions on some of our more complex policies. Take, for example, our policy against Unreliable Financial Claims which includes ads promoting get-rich-quick schemes. The bad actors behind these types of ads have grown more sophisticated. They adjust their tactics and tailor ads around new financial services or products, such as investment advice or digital currencies, to scam users.

To be sure, traditional machine learning models are trained to detect these policy violations. Yet, the fast-paced and ever-changing nature of financial trends make it, at times, harder to differentiate between legitimate and fake services and quickly scale our automated enforcement systems to combat scams. LLMs are more capable of quickly recognizing new trends in financial services, identifying the patterns of bad actors who are abusing those trends and distinguishing a legitimate business from a get-rich-quick scam. This has helped our teams become even more nimble in confronting emerging threats of all kinds.

To give an idea of the scale of our efforts on ads safety, in 2023 we blocked and removed 5.5 billion bad ads, and we suspended over 12.7 million ad accounts for policy violations, up from 6.7 million in 2022. With regard to scams in particular, we blocked or removed 206.5 million advertisements for violating our misrepresentation policy, which includes many scam tactics, and 273.4 million advertisements for violating our financial services policy. We also blocked or removed over 1 billion advertisements for violating our policy against abusing the ad network, which includes promoting malware.[2]

2 Ads Safety Report, 2023, https://blog.google/products/ads-commerce/google-ads-safety-report-2023/#enforcement

Advertiser Verification

To provide a safe and trustworthy ad ecosystem for users, Google will require advertisers to complete one or more verification programs. The advertiser verification program can comprise multiple steps:

- About your business: In the first step of the Advertiser verification program, Google will ask advertisers a few basic questions related to their Google Ads account and business under the 'About your business' section. These questions help Google get a better understanding of advertising businesses. For example, Google asks whether a business is an advertising agency, who pays for the ads, whether the advertiser promotes their own products/services or someone else's, and the advertiser's specific industry (or industries).

- Verify your identity: Upon completion of the 'About your business' questions, advertisers may be asked to verify their legal name via the advertiser identity verification or business operations verification process. This verification must be completed by an authorized representative, who is an admin of the Google Ads account and/or the payments profile paying for the ads.

- Verify your business operations: Based on the responses in the About your business questions, Google may ask advertisers to verify details about their business operations (if applicable) along with supporting documentation, such as their business model, business registration information, types of services offered, business practices, and relationships with advertised brands, products or third parties, if applicable.

Once Google verifies the advertiser's information, we will disclose the advertiser's name and geo-location in My Ads Center, which is one-click away from every ad. From there, users can also click through to our searchable Ads Transparency Center to view additional information, including all of the ads the advertiser has run on Google's network.

We are actively verifying advertisers in over 240 countries and regions. If an advertiser fails to complete our verification program after being prompted and given a deadline to complete the process, the account is automatically restricted from serving ads. In risk prone areas, we have higher-friction verification processes in place. For example, in order to advertise financial services in certain countries, advertisers must undergo an additional layer of verification. For most advertisers, this process will entail demonstrating that they are authorized by their local regulator to promote their products and services through ads. As of Q4 2024, we've launched our financial services advertiser verification program in 17 countries and regions.

Advertiser verification also provides a valuable signal that contributes to other safety-by-design product features. For example, in November 2023, we launched Limited Ads Serving, which is designed to protect users by limiting the reach of ads in risk-prone areas from advertisers with whom we are less familiar. The Limited Ads Serving policy introduces what could be described as a "getting to know you period" for advertisers in ad-serving scenarios that have a higher potential of causing abuse. This policy is specific to a certain set of ad-serving scenarios, and in these instances, only qualified advertisers will be able to serve ads without impression limits.

Imagine someone is looking to book their next trip and searches for flights to San Francisco with their favorite airline. Under this new approach, the vast majority of ads they see related to that search would be from advertisers like the airline itself, competing airlines, hotels in the area, and other advertisers with a history of policy compliance and transparency. Advertisers without this record of good behavior might have their impressions limited under this policy as they build their track record on our platform. While we want to allow users the opportunity to interact with relevant and helpful ads, this policy will reduce the chance that they'll see a misleading or confusing ad from an advertiser with an unproven track record.

The importance of transparency, user choice and control

User confidence in Google's products and services is essential. We want users to be empowered to make informed decisions about the ads they see online. Trust in advertisers on our platforms helps us deliver a smart and useful web experience for everyone. This means providing transparency about who our advertisers are, where they are located, and which creatives they have served through Google. We have a history of ads transparency:

- Since 2011, we have published an annual Ads Safety report detailing the actions we've taken to prevent malicious use of our ads platforms.
- In 2018, starting in the US, we began requiring advertisers who wish to run election ads on our platforms to go through a verification process and include an in-ad disclosure that clearly shows who paid for the ad.
- In 2020, the introduction of advertiser identity verification provided users with additional transparency, helping them learn more about the company behind a specific ad and also differentiate credible advertisers in the ecosystem while limiting the ability of bad actors to misrepresent themselves.
- In 2022, we launched My Ads Center, which gives people more control over their ad experience on Google's sites and apps. Within My Ads Center, people can block sensitive ads and learn more about the information used to personalize your ad experience.
- In 2023, we announced the Ads Transparency Center, a searchable hub of verified advertisers where users can view basic information about the advertiser and see the other ads they are running on our platforms.

Google is no stranger to adversarial actors who attempt to hack our defense systems, and there will always be sophisticated bad actors who seek to game or circumvent our systems. We want to make it difficult and expensive for bad actors to use Google Ads products. Increasing the cost and friction through policy changes and verification requirements helps to achieve this.

Globally, we have seen a reduction in abuse where we have been able to deploy robust advertiser verification in high risk areas. It's also important to remember that verification is one layer of defense – it is not a "silver bullet" that will work in all circumstances – and we rely upon multiple enforcement processes and policies on top of verification. We are constantly investing in new processes to stop fraudsters before they reach our platform, as well as reviewing and updating our policies to capture their changing tactics.

Annex 3.4 YouTube

YouTube is a video sharing platform where users can upload and share videos. Responsibility is YouTube's number one priority and reflected in everything we do, including our approach to scams. Every day, users look to YouTube for information and educational resources to enrich their lives, which includes learning about financial matters.

Google's work on Ads Safety extends to all our products: Google Ads Financial Products and Services policy is enforced across all Google Ads, including Ads on YouTube. But YouTube's protections against scams go beyond Ads on the platform: YouTube's community guidelines also prohibit scams, such as exaggerated promises to get rich quickly, pyramid schemes, and cash gifting schemes in any content shared by users on the platform, including in videos, video descriptions, and comments posted by other users.

YouTube's Community Guidelines are designed to enable free and open exchange of ideas while keeping our community safe.

YouTube's Community Guidelines prohibit users from posting content that encourages dangerous or illegal activities. We list below some of our Community Guidelines:

- Spam, scams, deceptive practices policy: YouTube does not allow spam, scams, or other deceptive practices that take advantage of the YouTube community. We prohibit content offering cash gifts, "get rich quick" schemes, or pyramid schemes (sending money without a tangible product in a pyramid structure). We also don't allow content where the main purpose is to trick others into leaving YouTube for another site, including scam sites. Related policies include the External Links policy which covers links to websites or apps that install malware, or enable phishing; and our Video Spam policy.
- Harmful & Dangerous policies: Prohibit content promoting products and services such as counterfeits, illegal sales (for example bank account passwords or stolen credit cards) and digital security (phishing, hacking, etc). User channels may be terminated on such grounds.
- Impersonation policy: Under this policy, we do not allow content that is intended to impersonate a person or channel.
- Misinformation policy: Certain types of misleading or deceptive content with serious risk of egregious harm are not allowed on YouTube. This includes content that has been technically manipulated or doctored in a way that misleads users (usually beyond clips taken out of context) and may pose a serious risk of egregious harm.

We take action to remove content that violates our policies as quickly as possible, using a combination of people and machine learning to detect and enforce on violative content at scale. For repeated violations or a single-instance of severe abuse, we may terminate a user's channel or account.

Machine learning enables us to proactively identify and flag harmful content to our human reviewers, and in some instances, automatically remove certain types of content very similar to what has been previously removed, such as spam. This allows us to take action on violative content often before it is widely viewed by users.

The impact of this approach is evident. From April through June 2024, our Violative View Rate was 0.9-0.11%, meaning that for every 10,000 views of content on YouTube, 0-11 of those are of content that violates our Community Guidelines. During the same period, YouTube removed over 8.4 million videos for violating our Community Guidelines, and terminated over 3.2 million channels for violating our Community Guidelines. The overwhelming majority were terminated for violating our spam policies. More details are in the YouTube Community Guidelines enforcement report, which is published every quarter.

Monetisation: YouTube has a higher bar for creators to earn revenue for content on YouTube via the YouTube Partner Program. Over the last few years, YouTube has taken steps to strengthen the requirements for monetisation to limit spammers, impersonators and other bad actors from accessing monetization privileges and harming the creator ecosystem. For example, to qualify for content monetization via our YouTube Partner Program, channels must meet eligibility thresholds related to watch time and subscribers. Following a creator's application, Creators may also have their privileges suspended for repeated violations of our content and monetization policies. YouTube reviews the channel to identify any violations of our monetisation, content and copyright policies. More details in the YouTube Partner Program overview & eligibility.

YouTube Shopping

YouTube Shopping is a set of features to build more authentic, trusted, and enjoyable shopping experiences on YouTube[3]. With YouTube Shopping, creators can feature and tag their own products, products from other brands, and easily track their revenue and performance.

YouTube has established criteria for what types of videos and creators can utilise YouTube Shopping features, including a dedicated YouTube Shopping Affiliate program with specific criteria for eligibility. This helps to build a more responsible platform for creators, brands, and viewers. As part of our responsible approach, product tags will not be shown to viewers if the content contains copyright claims, has an audience that is predominantly "Made for Kids", or if the creator has active community guideline strikes (although the active community strikes may be re-examined).

YouTube Shopping has a number of policies, terms and conditions in place to address many concerns related to consumer protections, but we will continue to engage with policymakers in order to achieve optimal policy solutions for consumers, creators, and brands. While YouTube Shopping is working to create a seamless experience for viewers, currently, transactions do not take place on YouTube. In addition to our eligibility criteria, we work closely with a number of partners that have their own terms and conditions for using their platforms. For example, policymakers may be looking at rules on fake reviews and influencer disclosures, fraudulent transactions and counterfeit goods, refunds and returns, "dark patterns", and hidden fees.

Viewers often trust shopping recommendations from creators they follow on YouTube. The quality and authentic information from YouTube creators helps consumers make more informed and confident purchasing decisions.

3 YouTube Shopping Guide - go/yt-shopping-guide

Annex 3.5 Gmail and Workspace apps suite

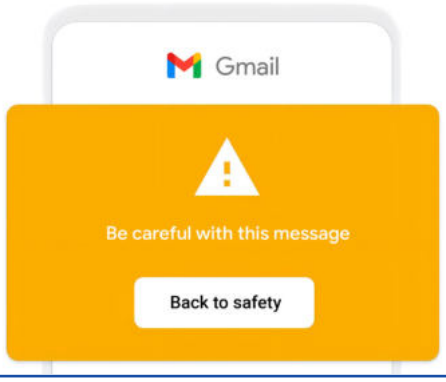
Google's email service, Gmail, has been known since its creation in 2004 for its effectiveness in protecting users from spam and phishing attempts thanks to AI-enhanced filtering. Similar efforts have been extended to the entire suite of Google Workspace applications, such as Google Docs, Google Forms, etc.

Spam, phishing, and malware continue to be serious threats to all email users. We have adapted to more sophisticated phishing campaigns, while also prioritizing phishing protections that are most immediately threatening to users' data and credentials.

- Gmail blocks 99.9% of dangerous emails before they reach users every day (this includes emails containing spam, phishing links or harmful malware).
- 63% percent of the malicious documents we block in Gmail differ from day to day.
- 68% of the phishing emails blocked by Gmail today are new variations that were never seen before.

Faced with a significant increase in spam and phishing during the Covid pandemic, we doubled down on our efforts at in-built protective features. For example, our Gmail malware scanner now processes more than 300 billion attachments each week to block harmful content.

Machine learning helps us with upwards of 95% of percent of all spam and phishing identification in Gmail. This is an area where more data enhances the protections we're able to offer to Internet users. Our improving technology in this area thwarts many account hijacking efforts, including phishing campaigns, from ever reaching the inboxes of users.



We added in-product technical protections in 2023 when we started requiring that emails sent to a Gmail address must have some form of authentication. As a result, we have seen the number of unauthenticated messages Gmail users receive plummet by 65%, helping to declutter inboxes while blocking billions of malicious messages with higher precision. Since early 2024, we have required bulk Gmail senders to authenticate their emails, allow for easy unsubscription, and stay under a reported spam threshold.

In addition, Google's Threat Analysis Group, a dedicated team of security professionals, further detects, prevents, and mitigates government-backed threats. They share their main findings regularly, in a quarterly bulletin and ad hoc reports.

Google continues to issue warnings to users when we believe they may be the targets of government-backed phishing attacks. We have issued these warnings, which include advice about ways to improve the security of users' Google accounts, since 2012.

We've built new systems that detect suspicious email attachments and submit them for further inspection by Safe Browsing (see section of Chrome above). This protects all Gmail users, including enterprise Workspace customers, from malware that may be hidden in attachments and which can be used as a vector to defraud users, for example as a way to steal their login and other personal identification details.

Annex 3.6 Search

Every search matters. That is why, whenever users come to Google Search to find relevant and useful information, it is our ongoing commitment to make sure that they receive the highest quality results possible. Unfortunately, on the web there are some disruptive behaviors, including scams and other threats to online user safety.

Online scams and fraud affecting search results might include deceptive sites impersonating a business or service provider to trick users into paying money to the wrong party.

For example, many fraudsters pretend to be offering customer support phone numbers to popular services and products, only to mislead users who call in into paying them via bank transfers or gift cards. Fraudsters may create many low quality websites with so-called ‘keyword stuffing’, logos of brands they're imitating, and a phone number they want you to call. These sites can trick people into disclosing sensitive personal information, losing money, or infecting their devices with malware. Commonly known as 'customer support scam' or 'tech support scam', this type of scam has been reported by hundreds of thousands of users and can lead to losses of hundreds of dollars to fraudsters in each case.

Our efforts to counter fraud and scams are based on:

- research which leads to scaled automated solutions and
- learnings from specific cases that we have collected over the past several years. For instance, based on phishing attacks and scams reported by users via Google reporting tools for scam and phishing, we identified the most common scam practices, then developed solutions and have taken action to address them.

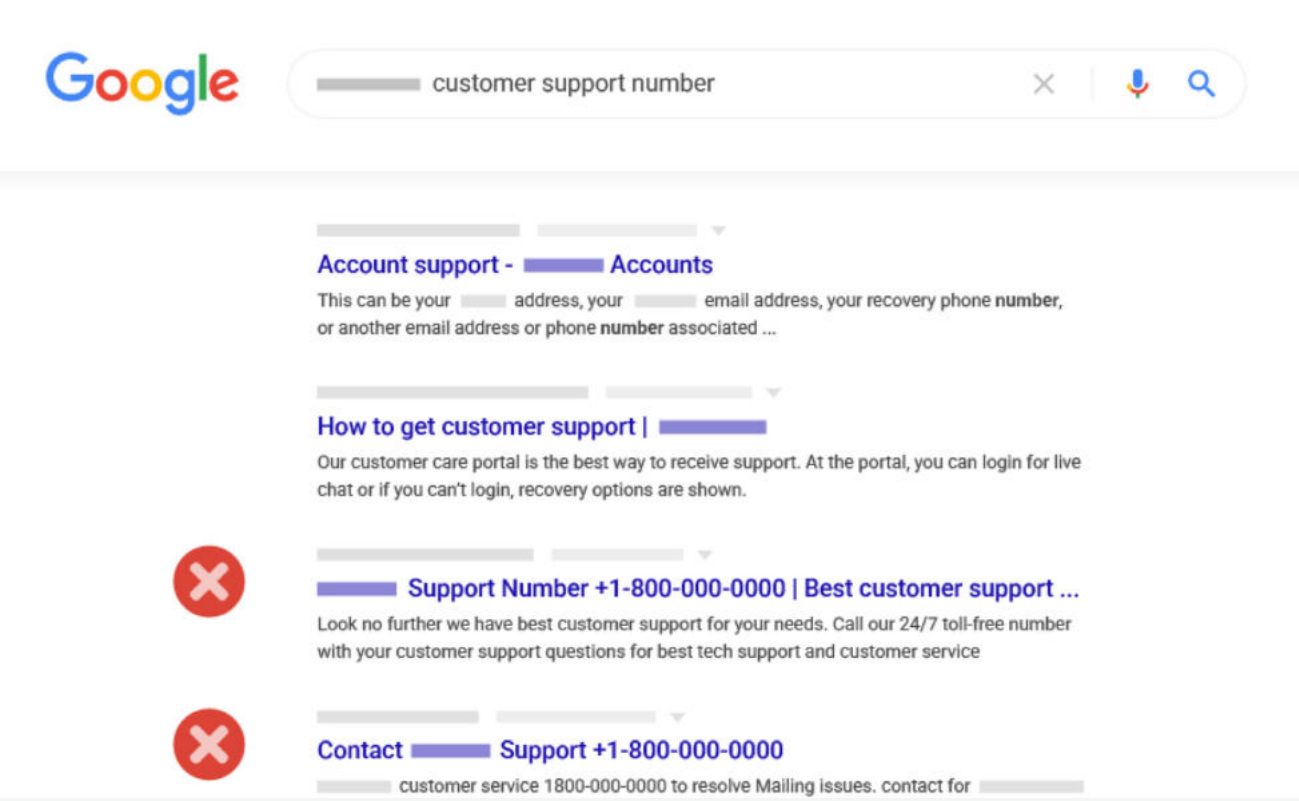
Prevent

We have been closely monitoring queries that have a higher likelihood of returning scammy pages within organic search results, and we've worked to stay ahead of spam tactics in those spaces to protect users. Since 2018, our systems have been able to protect hundreds of millions of searches a year by detecting potentially scammy sites and preventing them from showing up in Google Search results.

Our ability to identify disruptive and malicious behaviors among billions of web pages has allowed us to keep more than 99% of searches spam-free. We're also taking our approaches, best practices, and lessons learned to address potential fraud and scams that may not be addressed through our extensive protections against spam. For example, our improved algorithmic approach has resulted in a 40% reduction of certain spam results in 2021 in comparison to 2020, such as customer support scams.

Of course, there are many other forms of spam, and part of our work is to ensure our defenses stay up to date.

Using algorithms, our dedicated classifiers help to detect and action on spammy pages, which enables us to filter over 2 billion search results daily, and to catch 96% of cloaking spam.



Ranking protections: Reducing low-quality, unoriginal results

Across our products, Google seeks to provide the most relevant and authoritative results possible. We use ranking algorithms to ensure we are meeting users' expectations of surfacing relevant and high quality sources, as well as minimizing low quality or harmful content from appearing prominently in search features or search results, where users are not actively seeking out such content. The design of these systems is our greatest defense against harmful low quality content.

For example, where bad actors are seeking illegal content, we deliberately uprank content that is aimed at user safety. So in this context, if a user searches for "find financial credentials darkweb", we will deliver results for users about how to protect their account credentials from being compromised, how to find out if they've been leaked onto the dark web, and what you can do about it.

In 2022, we began tuning our ranking systems to reduce unhelpful, unoriginal content on Search and keep it at very low levels. We brought what we learned from that work into a later update.

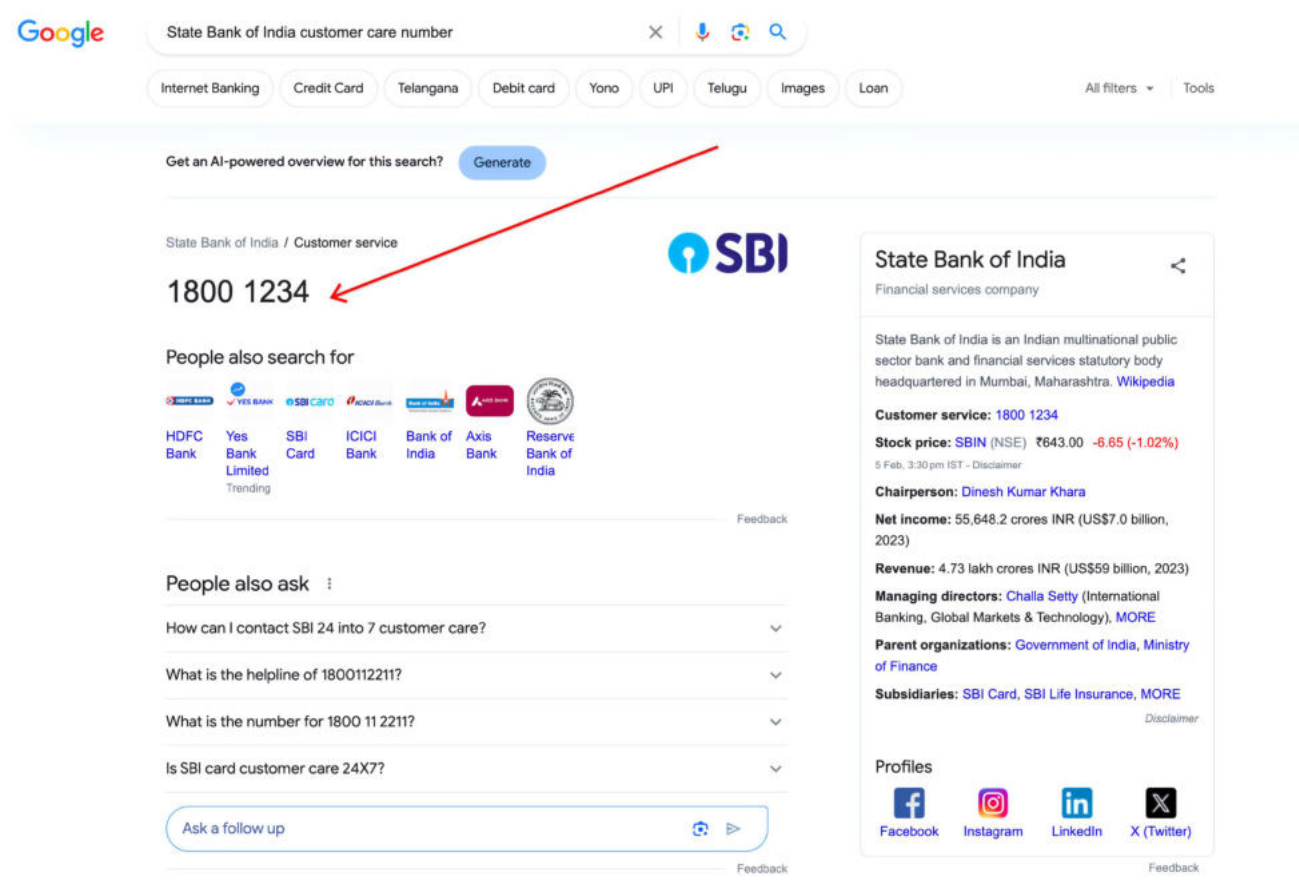
This update involves refining some of our core ranking systems to help us better understand if webpages are unhelpful, have a poor user experience or feel like they were created for search engines instead of people. This could include sites created primarily to match very specific search queries.

We believe these updates will reduce the amount of low-quality content on Search and send more traffic to helpful and high-quality sites. Based on our evaluations, we expect that the combination of this update and our previous efforts will collectively reduce low-quality, unoriginal content in search results by 40%.

User empowerment

We also help users in seeing authoritative content, protecting them further. An example is the advice we provide to web publishers on best practices that help ensure Search results are showing the most accurate information for their business or service. This helps ensure that people are directly contacting a business, rather than reaching a third-party claiming to have an authorized relationship that cannot be verified, in an effort to defeat fake customer support scams.

We also use information panels appearing around Search results, known as Trigger Featured Snippets or Knowledge Panels, for most businesses and service providers. These information panels help users get a quick snapshot of the authoritative information and customer care number for a given business.



Detect

Our protection efforts on Search focus on ranking and automated detection:

- First, we are identifying scams via automated detection, using a variety of signals.
- Second, we train our ranking algorithms to recognize low quality webpages. This includes URLs that are not authoritative or trustworthy.

As a result - we are using these signals to downrank sites that are likely to be low quality, in keeping with our broader efforts and principles – by which we surface high authority content on Search.

Respond

(i) Legal removals

We restrict content that violates the laws of a country in which our products and services are operating.

We provide tools to help users report content that they believe should be removed from Google's services based on applicable domestic laws. For example, a website selling stolen financial data would be illegal under UK law, and therefore subject to our legal removals policy.

The content for these removal requests is removed in the specific country, except for copyright related issues, which are removed globally.

(ii) Policy removals

As with other Google products, we also develop and maintain policies - which outline what types of content and behaviors are not acceptable for each product or service - Known as 'content policies'. We aim to make them clear and easily accessible to all users.

Specifically for Google Search, our Search policies explain what types of content are not allowed, and the process by which a piece of content may be removed from the Google Search results.

For example, our policy on Personally Identifiable Information (PII) removal policy provides user protections for personal information that creates significant risks of identity theft, financial fraud, or other specific harms. Under the policy, the user or an authorized representative, can submit a request to remove links to the content containing certain PII (such as Confidential national identification numbers, bank account numbers, etc) from Google search results.

We are continually seeking to enhance these policies and their enforcement. That's why in 2024 Search introduced a number of updates to combat spam and improve the overall usefulness and quality of results. This includes changes to the guidelines that inform the way we detect and make enforcement decisions around spam, along with enhancements to our ranking systems to help us better understand if webpages are unhelpful and have a poor user experience—but which might not meet our formal definition of spam.

Alongside the enhancements to our ranking systems, there are three key elements to our spam updates:

- Scaled content abuse - We've long had a policy against using automation to generate low-quality or unoriginal content at scale with the goal of manipulating search rankings. These efforts have become more sophisticated and our policy will now focus on the abusive behavior — producing content at scale to boost search ranking — whether automation, humans or a combination are involved.
- Site reputation abuse - Some websites permit third parties to post low-quality content as a form of monetization (for example, a third party might publish payday loan reviews on a trusted educational website). We'll now consider very low-value, third-party content produced primarily for ranking purposes and without close oversight of a website owner to be spam, and will start enforcing the policy on May 5 to ensure webmasters have time to understand and make any necessary changes to their practices.
- Expired domain abuse - Occasionally, expired domains are purchased and repurposed to boost search ranking of low-quality or unoriginal content, and to potentially mislead users into thinking the new content is part of the older site. Expired domains that are purchased and repurposed to manipulate the search ranking of low-quality content are now also considered spam.

These changes are just part of the widespread efforts to keep low-quality content on Search to low levels, so we can better connect people with helpful information.

For more details on how Search protects users, you can visit How Search Works, our dedicated website outlining how we help users access relevant and useful information at-scale, how we fight spam, and how we continually improve Search through rigorous testing.

Annex 3.7 Shopping

We have a number of tools and features in place to make sure users can trust what and who they are buying from. There are three key ways we help users shop safely on Google:

(i) Automation helps us vet listings quickly and accurately

Products and merchants go through in-depth safety reviews before they can list on Google. Thanks to the Shopping Graph, our data set of the world's products and sellers, our systems can quickly review whether a business is legitimate, the products you see are accurate and their content follows our policies.

Like any community, we need policies to keep things trustworthy. Our shopping policies, which cover product listings and shopping ads, outline what is and isn't permitted on Google, including any products crawled from the web and shown in shopping results. These policies can give you confidence that the product you see is what you'll get, and that you won't have to filter through things we don't allow, like violent weapons, merchants misrepresenting their businesses or hateful content.

We use a combination of automated and human evaluation to ensure that product listings comply with our policies. Our enforcement technologies use algorithms and machine learning, modelled on human reviewers' decisions, to help protect our customers and keep our platforms safe. Automation has helped us more efficiently and accurately review a significant amount of products. More complex, nuanced or severe cases are often reviewed and evaluated by our specially-trained experts.

In January 2023 alone, we stopped over 100 million product offers from being shown and disapproved nearly 300,000 accounts for having quality issues or not following our policies.

(ii) Store badges and other visual cues point out quality businesses

For extra reassurance about a merchant, you can look for store badges, which we give to businesses on Google where you may expect a positive shopping experience. Stores which show this badge have earned it by having, for example, fast shipping, easy returns, high quality websites and good user ratings.

We'll also show you ratings for both an item and its various sellers, so you can learn about other shoppers' experiences with those products and businesses. And since our product listings bring you right to merchant websites, you can do even more research about a store directly on their site.

(iii) Our automated and human review teams monitor live merchants and listings

Our safety efforts don't stop once a product listing goes live. Our automated systems are always monitoring for violating activity, and our team of human reviewers is on standby to review issues that might need a more nuanced perspective.

For example, our systems might pick up that an electronics company has dropped its prices by 70% and removed its shipping information from its website. Or maybe they notice that a merchant that was originally selling sweaters is now listing household appliances. These might be signals for our team to take a closer look at whether something misleading is happening.

Additionally, after they're onboarded, we keep an eye on merchants and their listings, making sure nothing has suspiciously changed since they first came to Google. For example, if a merchant met country-specific rules for selling alcohol when they onboarded, we'll check in periodically to make sure those qualifications are still met.

We take different types of actions when we see odd behavior — from removing listings that seem suspicious or violate our policies, to banning a merchant from listing on Google. If users see something that doesn't look right or seems out of sync with our policies, like a super high price or a violent weapon, they can report it to us via the "report a listing" link on the bottom right of a product page or through our Help Center.

Annex 3.8 Payments

Privacy and security needs to be a base-level utility for all payments systems. A top Google priority is to ensure safe and secure payments, regardless of whether you're using Google Wallet or Google Pay.

We use world-class machine learning and fraud detection algorithms to:

- Ensure transactions are fast, secure and easy, both online and with tap & pay in store and
- Make sure that your money remains safe.

We also provide fraud management tools to the ecosystem. During enrollment of token, GPay provides issuers with a toolkit to verify account ownership of users, and at transaction time, GPay provides token and cryptograms so partners make their own risk evaluation. Device tokenisation protects card information from leakage throughout the chain, and two-factor confirmation of online or high-value payments provides gold-standard security step-up seamlessly.

We store all data on encrypted, secure servers.

We've designed and developed our payments products around strong privacy principles of transparency, choice and control. For example, you can save or remove a payment method from your account at any time, and easily turn autofill on or off for your cards. Plus, all of the security measures we take to protect your Google Account apply to your payment cards, too. For example, we confirm your identity before authorizing payments from a device with the same technology used for other Google features. For recommendations and guidance to keep your Google Account safe and secure, users are encouraged to regularly take a Security Checkup.

In addition to Google's security infrastructure, users get all the protections they are used to getting from their card service provider. Contactless payments are not only convenient, they're also more secure than swiping your card.

How we prevent unauthorized payments via Google Wallet:

- Google Wallet doesn't store your cards on the phone
- The data is stored on secure and encrypted servers
- If your device gets stolen, you can rest assured that your payment methods are safe. Contactless payments with Wallet must be recently authenticated with a secure screen lock — like a PIN, password, facial recognition or biometric
- Users can remotely wipe their devices if it's lost or stolen using Google Find My Device.

Additionally, when you add a payment card to Google Wallet, it creates a device-specific virtual account number, known as a device token, so your real card number isn't stored on your device or shared with merchants.

Annex 3.9 Maps

Every day, we receive around 20 million contributions from people using Maps. Those contributions include everything from updated business hours and phone numbers to photos and reviews.

As with any platform that accepts contributed content, we must stay vigilant in our efforts to fight abuse and make sure this information is accurate. We take a robust approach towards fake reviews and given the volume of reviews we receive, we invest in machine learning and human moderators to build trust between customers, businesses, and users. As soon as someone posts a review, we send it to our moderation system to make sure the review doesn't violate any of our policies.

Thanks to a combination of machine learning and human operators, we continue to decrease the amount of content seen on Maps that is fraudulent or abusive – in fact, it's less than one percent of all the content that is viewed on Maps.

Machines are our first line of defense because they're good at identifying patterns. These patterns often immediately help our machines determine if the content is legitimate, and the vast majority of fake and fraudulent content is removed before anyone actually sees it. Our team of human operators works around the clock to review flagged content. When we find reviews that violate our policies, we remove them from Google and, in some cases, suspend the user account or even pursue litigation. In addition to reviewing flagged content, our team proactively works to identify potential abuse risks, which further reduces the likelihood of successful abuse attacks.