

**Întrebare: La începutul lunii iunie, jurnalistul Dragoș Pătraru a descoperit pe Google reclame care promovează videoclipuri ( [Video 1](#) și [Video 2](#) ) deepfake realizate cu AI, în care îi este folosită imaginea pentru a promova niște escrocherii. Deși a raportat ambele videoclipuri acum o lună, și și-a îndemnat și comunitatea să o facă, videoclipurile încă sunt acolo, iar jurnalistul ne spune ca nu a primit niciun fel de notificare din partea Google. Cum explicați această situație?**

Răspuns: Deepfake-urile sunt o problemă care a luat amploare în ultima perioadă, inclusiv în România. Pentru Google, situația este de cea mai mare prioritate, o abordăm cu proactivitate și ne încercăm să o rezolvăm cât mai repede posibil. Ne străduim să identificăm și să eliminăm rapid acest tip de conținut din platforma noastră, folosind o abordare multiplă care combină politici stricte, tehnologii avansate și echipe dedicate de experți.

Protejarea utilizatorilor este prioritatea noastră principală. Avem [politici](#) stricte care guvernează ce tipurile de reclame permitem să ruleze pe platforma noastră. Investim masiv în aplicarea politicilor noastre printr-o combinație de evaluări umane și sisteme automate. Avem o echipă de mii de persoane care lucrează non-stop pentru a crea, actualiza și aplica politicile noastre la scară largă. Continuăm să investim resurse semnificative pentru a opri actorii rău intenționați, a evalua și actualiza constant politicile noastre și a ne îmbunătăți tehnologia pentru menținerea utilizatorilor în siguranță. În mod constant analizăm ecosistemul publicitar și, dacă găsim reclame care încalcă politicile noastre, luăm măsurile adecvate.

**Întrebare: Ce măsuri aveți în vigoare pentru a preveni afișarea reclamelor care sunt înșelătoare sau care conțin deepfake-uri?**

Răspuns: Politicile noastre interzic clar deepfake-urile și reclamele false, oferind un cadru clar pentru a identifica și elimina acest tip de conținut. Tehnologiile avansate, bazate pe algoritmi de învățare automată, ne permit să detectăm rapid reclamele false. De asemenea, avem echipe dedicate de experți care analizează manual conținutul suspect. Pe lângă detectare și eliminare, luăm și măsuri severe împotriva celor care creează și difuzează deepfake-uri și reclame false. Investigăm amănunțit sursa acestor materiale pentru a identifica actorii rău-intenționați și aplicăm sancțiuni severe, inclusiv suspendarea permanentă a conturilor. Pentru a completa eforturile noastre interne, colaborăm cu experți din domeniul securității cibernetice și al inteligenței artificiale,

beneficiind de cunoștințele și perspectivele lor valoroase. De asemenea, avem inițiative de educare a utilizatorilor despre cum să identifice și să raporteze deepfake-urile și reclamele false.

În detaliu, avem diferite politici care acoperă tot ecosistemul de reclame. Una dintre ele interzice utilizarea conținutului media manipulat, inclusiv deepfake-uri și alte forme de conținut falsificat, cu scopul de a înșela sau induce în eroare utilizatorii. Această politică interzice, de asemenea, difuzarea de reclame care sugerează afilierea cu, sau aprobarea de către, o altă persoană sau organizație fără cunoștința sau consimțământul acestora.

**Întrebare: Care este procesul de aprobare a reclamelor și cum detectați conținutul înșelător?**

Răspuns: Avem [politici](#) stricte privind reclamele, care guvernează tipurile de reclame și publicitari care pot funcționa pe platformele noastre. De exemplu, nu permitem reclame care folosesc tactici de tip clickbait sau texte sau imagini senzaționaliste pentru a atrage trafic ([Politica privind informațiile false](#)).

Nu împărtășim detalii specifice despre sistemele noastre de aplicare a politicilor, deoarece actorii rău intenționați folosesc adesea aceste informații pentru a ne submina eforturile și a evita aplicarea. Folosim o combinație de evaluări umane și sisteme automate pentru a ne aplica politicile și monitorizăm constant rețeaua noastră pentru abuzuri. Adoptăm o abordare cu mai multe direcții pentru a combate comportamentul rău intenționat, cum ar fi verificarea identității agenților de publicitate și identificarea activității coordonate între conturi utilizând semnale din rețeaua noastră. Avem mii de oameni care lucrează non-stop pentru protejarea ecosistemului publicității digitale. De asemenea, folosim puterea LLMs (Large Language Motors) pentru siguranța reclamelor. Am început anul acesta să aducem capacitățile sofisticate de raționament ale lui Gemini (cel mai capabil AI al Google) în eforturile noastre de siguranță și aplicare a reclamelor.

Deși este treaba noastră să facem tot posibilul pentru a bloca reclamele dăunătoare, uneori actorii rău intenționați pot evita temporar detectarea noastră. În aceste circumstanțe, există lucruri pe care oamenii le pot face pentru a se proteja online. Primul lucru pe care o persoană îl poate face este să raporteze o reclamă dăunătoare. Dacă o persoană vede o reclamă pe platforma noastră despre care consideră că ar putea încălca politicile noastre sau ar putea dăuna utilizatorilor, o poate raporta foarte ușor. Evaluatorii

noștri vor analiza reclama și o vor elimina de pe platforma noastră dacă încalcă politicile noastre.

**Întrebare: Câte reclame au fost eliminate anul acesta din cauza conținutului deepfake sau înșelător?**

Răspuns: Lupta împotriva reclamelor înșelătoare este un efort continuu, în condițiile în care atacatorii operează cu tot mai multă sofisticare, la o scară mai mare, folosind noi tactici, cum ar fi acestea de deepfakes. În total, în 2023 am blocat sau eliminat 206,5 milioane de reclame pentru încălcarea politicii noastre privind informațiile false, care include multe tactici de înșelătorie, între care și deepfake, și 273,4 milioane de reclame pentru încălcarea politicii noastre privind serviciile financiare. De asemenea, am blocat sau eliminat peste 1 miliard de reclame pentru încălcarea politicii noastre împotriva abuzului de rețea publicitară, care include promovarea de malware.

În total, anul trecut am blocat sau eliminat peste 5,5 miliarde de reclame, puțin mai mult decât în anul precedent, și am suspendat 12,7 milioane de conturi de agenți de publicitate, aproape dublu față de anul precedent. În mod similar, lucrăm pentru eliminarea reclamelor noastre de pe paginile și site-urile care încalcă politicile noastre, cum ar fi conținutul explicit sexual sau produsele periculoase. În 2023, am blocat sau restricționat difuzarea de reclame pe mai mult de 2,1 miliarde de pagini, în creștere marginală față de 2022. În plus, am luat măsuri punitive la nivelul unui site întreg, pentru încălcărilor grave sau repetate ale politicilor, în cazul a aproape 400.000 de site-uri, în creștere semnificativă față de 2022.

Pentru a pune în perspectivă impactul AI-ului asupra acțiunilor de protejare a mediului reclamelor: anul trecut, peste 90% din măsurile noastre de aplicare a politicilor la nivel de pagină au început cu utilizarea modelelor de învățare automată, inclusiv cele mai recente LLM-uri ale noastre.

**Întrebare: Care este timpul mediu de răspuns pentru evaluarea și eliminarea unei reclame raportate?**

Răspuns: Nu putem vorbi despre un timp mediu de răspuns: odată ce o reclamă este raportată, sistemele noastre (automate sau cu element umane) o vor analiza. Dacă găsim reclame sau agenți de publicitate care încalcă regulile care guvernează sistemul nostru de publicitate, luăm imediat măsuri conform politicilor noastre.

**Întrebare: Ce sancțiuni sunt aplicate agenților de publicitate care încalcă politicile referitoare la deepfake-uri?**

Răspuns: Pentru a asigura o experiență sigură și pozitivă pentru utilizatori, Google solicită agenților de publicitate să respecte [toate legile și reglementările aplicabile](#), în plus față de [politicile Google Ads](#). Dacă găsim reclame sau agenți de publicitate care încalcă aceste reguli, luăm măsuri rapide în conformitate cu politicile noastre. Toate aceste politici sunt disponibile public și le aplicăm în mod consecvent și fără părtinire.

Politicile noastre sunt perfecționate și actualizate continuu pentru a aborda riscurile noi și emergente pentru utilizatorii și agenții noștri de publicitate. În 2023, am lansat 31 de politici noi pentru a aborda direct probleme noi sau emergente.

**Întrebare: Ce se întâmplă dacă reclama a fost plătită și este data jos înainte de termen? Trebuie rambursată o parte din suma încasată? Cât ați rambursat în România?**

Răspuns: În general, sumele rezultate astfel sunt investite înapoi în sistemele noastre de aplicare regulilor și politicilor, pentru a le consolida și a opri actorii rău intenționați în viitor. Investim foarte mult pentru menținerea ecosistemului publicitar sănătos, în sistemele care identifică și elimină elementele rău-intenționate. Sute dintre oameni de la Google (inginerii, experții și specialiști în diferite domenii) și-au dedicat cariera în această direcție.