

Detectarea imaginilor false devine o miză majoră: cursa globală împotriva conținutului generat de AI

Internetul este tot mai invadat de fotografii, videoclipuri și meme-uri generate cu inteligență artificială, multe dintre ele aproape imposibil de distins de realitate. În acest context, guvernele, companiile tehnologice și centrele de cercetare dezvoltă sisteme menite să identifice conținutul "sintetic" și să limiteze efectele fraudelor digitale.

Potrivit unei analize realizate de agenția EFE, pierderile provocate de fraude bazate pe conținut generat sau modificat de IA au depășit 1,5 miliarde de dolari la nivel global în ultimul an.

O industrie în plină expansiune:
instrumente pentru
detectarea
imaginilor false

Numeroase aplicații sunt deja disponibile, unele gratuite, și promet identificarea imaginilor generate de inteligența artificială prin analiza pixelilor, a distorsiunilor sau a amprentelor digitale ale sistemelor de generare. Printre acestea se numără IMGDetector.ai, RealReveal, AI or Not sau Vericta. Compania spaniolă Vericta a dezvoltat un sistem capabil să determine în câteva secunde dacă o imagine sau un videoclip este autentic ori fabricat digital.



Încrederea în imagini, tot mai fragilă

Co-fondatoarea Vericta, Georgina Viaplana, susține că imaginile nu și-au pierdut valoarea, dar oamenii tind să le accepte ca fiind reale fără verificări.

Ea avertizează că tocmai puterea unei imagini o face o țintă ideală pentru manipulare. În mediul de afaceri, verificarea autenticității a devenit deja esențială, mai ales în: asigurări, unde fotografiile pot fi falsificate pentru despăgubiri; comerț electronic, pentru solicitări frauduloase de rambursare; sectorul imobiliar, pentru validarea proprietăților.

Hervé Lambert, Panda Security: "O cursă a înarmării digitale"

Experții din securitate cibernetică avertizează însă că instrumentele de detectare nu vor fi niciodată infailibile. Potrivit lui Hervé Lambert, reprezentant al companiei Panda Security, evoluția tehnologiei generează o competiție permanentă între sistemele de detectare și cele de generare a conținutului fals. "Este o veritabilă cursă a înarmării digitale: fiecare progres în detectare duce la progrese în falsificare", afirmă acesta. Lambert consideră că eficiența acestor instrumente depinde în mare măsură de cultura digitală

a utilizatorilor. Statisticile arată că majoritatea oamenilor nu verifică autenticitatea informațiilor, iar un procent ridicat consideră că dispozitivele lor nu au nevoie de protecție. Specialiștii indică trei direcții esențiale pentru combaterea conținutului fals: Reglementarea utilizării inteligenței artificiale; Responsabilizarea platformelor online; Creșterea alfabetizării digitale și a gândirii critice. În opinia lor, etichetarea obligatorie a imaginilor generate de IA ar putea ajuta, dar nu reprezintă o soluție completă, deoarece metadatele și filigranele pot fi manipulate.