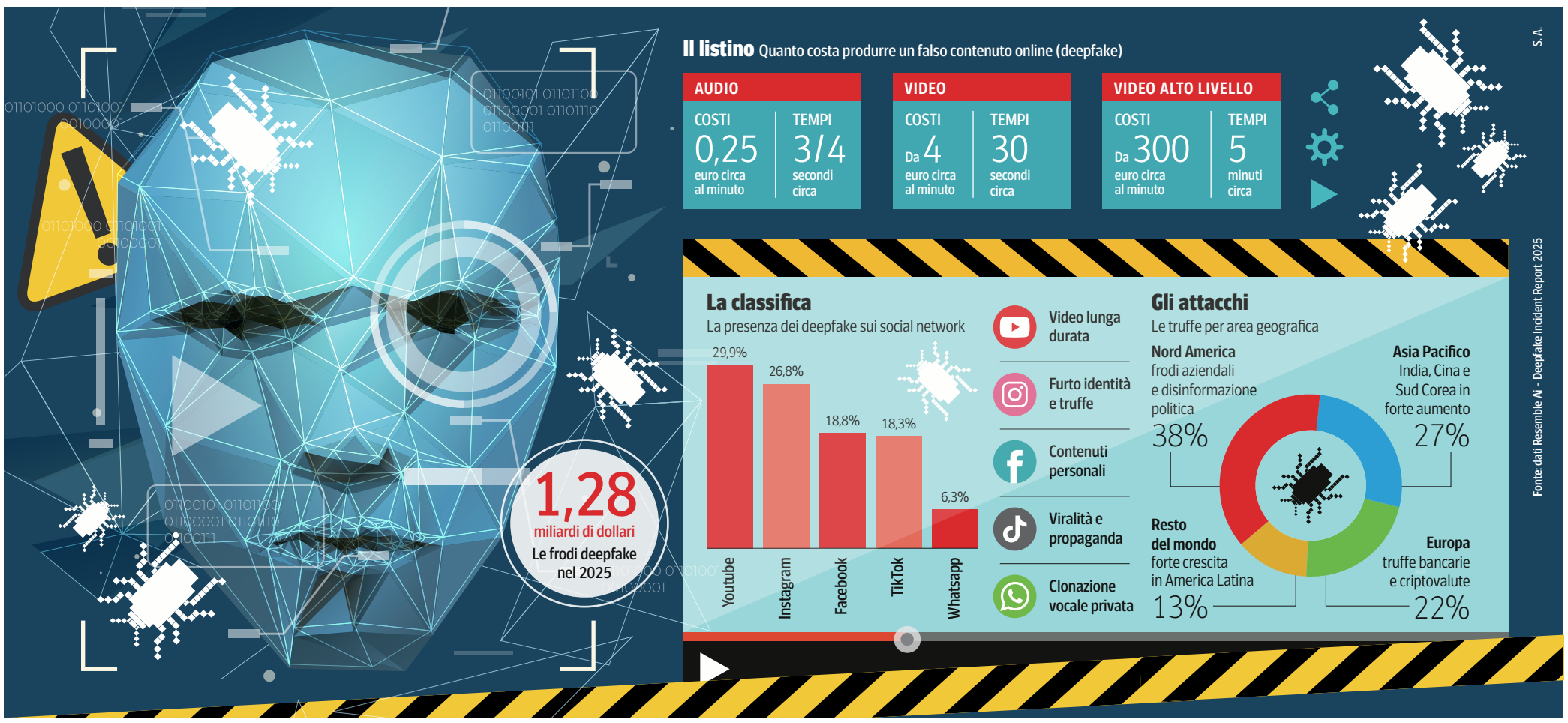
ALLARME DEEPFAKE: SONO 8 MILIONI POCHI EURO PER UN VIDEO FALSO E UN ATTACCO SU 3 È SU YOUTUBE



Prezzi bassi e tempi di produzione brevi. Le notizie audio e video generate dall'intelligenza artificiale si stanno moltiplicando sui social Bastano 25 centesimi al minuto per una voce contraffatta, quattro euro per un filmato. Dita, ombre, orologi: come riconoscere le trappole

di **UMBERTO TORELLI**

La linea di demarcazione tra le notizie audio e video autentiche e quelle create con i sistemi di intelligenza artificiale si sta assottigliando, ormai è diventata quasi invisibile. Secondo il Deepfake Incident Report 2025 di Resemble Ai, il mercato globale delle deepfake — termine nato dalla fusione tra deep learning, apprendimento profondo, e fake, falso, che indica la tecnica per la sintesi dell'immagine umana — negli ultimi 12 mesi ha fatto incassare ai truffatori 1,28 miliardi di dollari. Il fenomeno dei falsi contenuti si è amplificato alla fine dello scorso febbraio, con la guerra dichiarata da Stati Uniti e Israele all'Iran: i social network sono stati invasi da una valanga di immagini e video fasulli, diffusi allo scopo di generare distorsione informativa.

Il caso Dubai

Nelle prime settimane di conflitto, il *New York Times* ha individuato oltre 110 video di esplosioni mai avvenute, di strade e piazze devastate, con la popolazione civile e i soldati colpiti da armi da fuoco. E poi filmati di navi militari americane raggiunte dai missili, fino all'immagine emblematica del Burj Khalifa di Dubai, il grattacielo più alto del mondo, in fiamme, colpito da un missile iraniano. Tutto falso.

Era una campagna di disinformazione e dietro il forte realismo che ha tratto in inganno molte persone c'è l'ultima generazione di deepfake: falsi contenuti non solo più sofisticati di prima, ma anche generati da sistemi d'intelligenza artificiale ormai alla portata di tutti, invece che da potenti supercomputer. Come Sora, Nano Banana, Kling e altri software «chiavi in mano» presenti sul dark web. I prezzi e i tempi di creazione sono calati drasticamente.

Il Report 2025 di Resemble Ai calcola che un audio falso si possa acquistare ormai a partire da 25 centesimi di euro al minuto. E bastano pochi secondi di campionatura per assemblare la voce da contraffare. Mentre un video fake si può comperare per quattro euro al minuto.

Secondo DeepStrike, nel 2025 sono stati creati più di otto milioni di deepfake nel mondo: 16 volte in più rispetto al 2023. E i principali vettori di diffusione restano i social network. Il primo posto spetta a YouTube che ha veicolato il 29,9% degli attacchi nel mondo. I suoi video, di solito entro 10-15 secondi, mostrano falsi messaggi di celebrità che promuovono investimenti fraudolenti. Anche, come accade in Italia, con noti personaggi del mondo accademico e medico, che appaiono sullo schermo per pubblicizzare medicine con imprecisati poteri di guarigione.

Invece Instagram (26,8% degli attacchi) e Facebook (18,8%) si distinguono per i video con furto dell'identità digitale e le cosiddette «truffe sentimentali» (romance scams). Qui abili truffatori fingono attenzioni romantiche nei confronti della vittima, per indurla a inviare denaro con qualche pretesto. Su TikTok (18,3% degli attacchi) il panorama dei deepfake è vario. Si va dai discorsi politici manipolati ai video degli influencer con pubblicità ingannevoli, fino ai contenuti di pornografia che appaiono senza consenso.

L'anno scorso sono anche aumentati i deepfake su WhatsApp (6,3%). Non si tratta di contenuti pubblici, bensì privati. Spesso sono file audio inviati a familiari e dipendenti aziendali per indurli a trasferimenti urgenti di denaro.

Gli indizi

Ma come riconoscere un falso? «Nel caso dei file audio — dice Sabrina Curti, marketing director di Eset Italia — bisogna prestare attenzione ai rumori di fondo: il sospetto deve nascere quando sono assenti e uniformi». Invece per le immagini vanno guardati con cura i dettagli. Ad esempio, le mani. Normalmente, le dita sono irregolari. L'Ai può fonderle assieme, creando l'«effetto polpo». O le orecchie, gli orologi. Gli orecchini tendono a entrare nei lobi anziché pendere, sui quadranti compaiono numeri senza senso e lancette uscite dal nulla. Anche la pelle va osservata: nelle immagini generate dall'Ai spesso è troppo levigata, priva di pori e rughe, effetto «bambola di cera». E indizi di immagini fake arrivano anche dalle ombre, che sovente non cadono nella direzione corretta rispetto alla fonte di luce.

Vero è che online esistono rilevatori come Illuminarty, Hive, InVid. Sono software in grado di capire se un'immagine, o parte di essa, è falsa. Nessuno, però, ne assicura l'efficacia totale.

Fabio Ugolini è cofondatore e ceo di TrueScreen, azienda bolognese che ha sviluppato una piattaforma per garantire l'autenticità delle immagini. «I deepfake di ultima generazione — dice — vengono progettati per ingannare, così ogni volta che un software rilevatore impara a riconoscere immagini false, il modello generativo seguente mette in atto nuove strategie per aggirare i controlli».

Perciò i veri indizi da cercare, quelli più nascosti, non sono nei contenuti dell'immagine, bensì nei suoi metadati: chi l'ha generata, quando e dove, con quale tecnologia. Soprattutto, va verificato se esiste un flusso controllato di dati dalla creazione alla condivisione del contenuto. Se queste informazioni sono assenti o non certificate, il file digitale non può essere considerato attendibile.

È una nuova era di insicurezze digitali? Probabilmente sì, si tratta di imparare ad affrontarla. Oggi vediamo un'immagine e tendiamo a crederla vera, domani il primo pensiero sarà controllare se è autentica. È un passaggio simile a quello attraversato con le email. «Vent'anni fa ci fidavamo di qualsiasi messaggio — dice Ugolini —. Oggi sappiamo che può trattarsi di phishing».

29,9

Per cento

La quota di falsi contenuti su Youtube: video con celebrità che propongono acquisti fraudolenti. Seguono Instagram e Facebook con le truffe sentimentali