Caso Swift, suona l'allarme

IL FENOMENO / Le foto e i video ritoccati della seguitissima cantante americana hanno costretto i colossi della tecnologia ad affrontare la questione dei «deepfake» – Il web è d'altronde pieno di programmi e app ad hoc che sfruttano l'intelligenza artificiale

Stefano Olivari

Da mesi i *deepfake* sono sfuggiti di mano alle piattaforme che ne permettono la creazione, ma ci voleva Taylor Swift perché questo diventasse una notizia. La popstar più famosa del pianeta non è stata infatti la prima vittima dell'intelligenza artificiale ma è stata la prima ad avere il potere, almeno quello mediatico, di reagire e di costringere un colosso come Microsoft a una penosa retromarcia. Ma davvero è così facile creare un'immagine artificiale di chiunque, più vera di quella vera, inserendola nelle situazioni più imbarazzanti?

Taylor Swift, 34 anni, è la star più seguita al mondo. È fidanzata con il giocatore di football Travis Kelce. © AP/ALEX BRANDON

Il ruolo di X

Il caso Swift è partito da una serie di foto porno con protagonista la cantante di *Shake it off*, la fidanzata d'America e di Travis Kelce, che hanno ottenuto decine di milioni di visualizzazioni, generando la reazione degli avvocati della Swift e il terrore di quelli della Microsoft. La quale, temendo una causa miliardaria, ha aggiornato Designer, migliorando (in teoria) il filtraggio dei contenuti e impedendo adesso di associare qualunque nome, anche di sconosciuti, a comandi che usino termini sessualmente espliciti. Designer altro non è che lo strumento di intelligenza artificiale con cui creare immagini a partire da indicazioni scritte o vocali. E aveva un sistema di filtraggio per migliaia di nomi famosi, ad esempio era impossibile scrivere «Taylor Swift nuda». Sistema che però era facilmente aggirabile digitando i nomi con piccoli errori, «Talor Switch», o cose del genere. È quanto hanno fatto gli utenti di alcuni canali Telegram, con il risultato di inondare il web di foto imbarazzanti, costringendo alcuni social network, in particolare X, a inibire per qualche giorno la ricerca generica «Taylor Swift» prima di arrivare alla correzione attuale, nell'attesa di una legge che però non sarà mai più avanti delle possibilità tecnologiche.

Il termine *deepfake*
nasce nel 2017 associato al porno ma diventa ben presto strumento politico

Il giallo

Proprio un ingegnere di Microsoft, Shane Jones, già a inizio dicembre aveva scoperto diversi problemi nel modello DALL-E 3 di OpenAI, quello usato da Designer. Aveva scritto ad OpenAI, senza risposta. E allora ha pubblicato le sue perplessità su LinkedIn, chiedendo a OpenAI di fermare DALL-E 3, perché non impediva la realizzazione di *deepfake* sessualmente espliciti. Il risultato è stato che poche ore dopo Jones è stato diffidato dagli avvocati della Microsoft, che lo hanno per così dire invitato a rimuovere il post. Una storia pesantissima perché Open AI (cioè ChatGPT e tutto il resto) non è un'azienda indipendente ma dopo un percorso travagliato (all'inizio c'era di mezzo anche Elon Musk, che ancora ci sta facendo un pensiero) è controllata da Microsoft, che ne è l'azionista di maggioranza. Da notare che tutto si è svolto nel periodo della rimozione e poi del quasi immediato richiamo del CEO e fondatore Sam Altman, un uomo ascoltato da molti potenti della Terra. Insomma, non sono curiosità da nerd ma situazioni che cambieranno la vita di miliardi di persone.

La nascita del fenomeno

Il caso Swift ha fatto notizia per la notorietà dell'artista, unita al discorso porno, visto che è diventato senso comune associare l'uso di questa tecnologia al porno o comunque a qualcosa al confine della legge. E in effetti lo stesso termine, *deepfake*, nasce proprio associato al porno. Usato per la prima volta da un utente di Reddit nel 2017, e subito cavalcato da Vice, ai tempi sito di tendenza, è poi entrato nel linguaggio per i tanti casi di video o foto a contenuto sessuale riguardanti celebrità dello spettacolo, per poi diventare strumento politico di disinformazione grazie anche all'audio (ugualmente fake) e ad abili sceneggiature, che rendono certi falsi più veri delle persone a cui sono ispirati. Da Emma Watson a Donald Trump, da Katy Perry a Tom Hanks, da Scarlett Johansson a Joe Biden, negli Stati Uniti il fenomeno è molto più avanti che da noi, e del resto è lì che il *deepfake* è stato creato e studiato a livello accademico prima ancora che si chiamasse *deepfake*: il famoso progetto «Sintethizing Obama» ha segnato una svolta nella ricostruzione non soltanto delle immagini ma anche della mimica facciale.

I programmi

I *deepfake* si basano su un tipo di rete neurale chiamato *autoencoder* e sono costituiti da un *encoder*, che riduce l'immagine a uno spazio latente di dimensioni inferiori, e da un *decoder*, che ricostruisce l'immagine partendo dalla rappresentazione latente. Traduzione: già la rappresentazione latente dei Trump o Taylor Swift della situazione contiene gran parte delle caratteristiche fondamentali di viso e corpo, e su questa si innestano caratteristiche o informazioni ad hoc, continuamente rielaborate dall'intelligenza artificiale. Importante è quindi sapere che i *deepfake* per così dire moderni sono in continua evoluzione, per correggere le differenze rispetto all'originale, quindi sarà sempre più difficile distinguere il vero dal falso. Quando tutto questo è uscito dalle università, diventando patrimonio dello smanettone medio, la situazione è diventata ingestibile. Perché il web è pieno di programmi e app per creare *deepfake* facilissimi da usare. Da Reface, molto usata per i meme, a Zao, fra le migliori per sostituire i volti degli attori in un video, dalla famosa FaceApp, usata da chi è concentrato sulle foto, a SpeakPic, per chi queste foto le vuole far parlare, le app amatoriali di ottimo livello sono davvero tantissime. E quindi figurarsi cosa si può fare con quelle professionali.

L'esperto

Ale Agostini: «È un far west, ed è solo l'inizio»

«Facili da creare»
È davvero così facile usare la faccia di una persona facendole dire e fare di tutto? Lo abbiamo chiesto ad Ale Agostini, amministratore delegato di avantgrade.com, attiva dal 2011 e con sede a Balerna, la prima agenzia di digital marketing a integrare l'intelligenza artificiale. «È abbastanza facile – spiega Agostini – e i due grossi casi di questi giorni lo dimostrano. Non soltanto i video di Taylor Swift ma anche la chiamata vocale di un finto Biden a migliaia di elettori democratici per invitarli a non votare alle primarie. E siamo solo all'inizio».

«Leggi da riscrivere»
Le misure prese dopo il caso Swift sono insufficienti? «Funzionano al livello di un neofita, ma non a quello di chi voglia apposta fare male. Non saprei nemmeno come gli organismi sovranazionali potrebbero scriverla, una legge sull'intelligenza artificiale. La tecnologia va oltre le capacità di limitarla, lo abbiamo visto anche con i recenti scam bancari in Ticino». È quindi impossibile tutelare cinema, musica e in generale la creatività: «Per chi non ha un brand alla Taylor Swift difendersi sarà impossibile, mentre le fasce basse saranno sostituite dall'industria, non soltanto a Hollywood. Di sicuro tutte le leggi sul diritto d'autore andranno riscritte, ora siamo al Far West». Viene da chiedersi quale sia il limite dell'AI. «Il tetto è quello di cui parla Pedro Domingos nel suo *The Master Algorithm*: l'idea di un'intelligenza artificiale generale, in grado di prendere decisioni per il bene comune». **SO**

IMMAGINI E VERITÀ

L'EDUCAZIONE VISIVA

Michele Amadò

Rispetto alla potenza delle immagini cui siamo costantemente bombardati, siamo disarmati. Come distinguere quelle fasulle da quelle autentiche? Immagini finte e manipolate che, anche grazie agli sviluppi dell'intelligenza artificiale, si presentano come più credibili e vere di quelle originali. Le guerre, ad esempio, sono anche contese di immagini. Immagini che suscitano emozioni che contribuiscono, spesso in modo irriflessivo e non mediato, a formare l'opinione pubblica. Siamo bersagliati da foto, video di immagini crude, atroci, di persone esposte alla crudeltà altrui, a sostegno di questa o quella parte dei belligeranti. Spesso si tratta di video fatti altrove, in altri tempi, se non fasulli. Una questione rilevante sorge per il fatto che le immagini autentiche, naturali, reali come quelle ricostruite, artificiali o falsificate sono assunte dal nostro sistema visivo, dall'occhio al cervello, nella stessa maniera, senza alcuna distinzione. Entrambe poi sono potentissime, e ci prendono alla gola, come detto, in modo istantaneo, non mediato.

Già Platone denunciava questo potere manipolatorio delle immagini che fa sì che prendiamo spesso lucciole per lanterne. Per questo motivo il filosofo bandiva gli artisti dalla sua Repubblica. Ma non è questa la via che vogliamo seguire. Il paradosso è che nella nostra epoca, che ormai da qualche secolo, in modo acuito col Protestantesimo, è logocentrica, ossia è un ragionare che affida al linguaggio (verbale e scritto) la ricerca e l'espressione privilegiata del senso, siamo invece molto più influenzabili dalle immagini e meno capaci di interpretarle e distinguerle piuttosto che, ad esempio, in epoca medioevale. In quell'epoca, gran parte delle persone erano analfabete, o meglio non sapevano leggere, ma sapevano ben meglio di noi interpretare le immagini, ed erano in grado di comprendere il senso trasmesso da esse.

Chi è educato a leggerle, e sa che sono una finzione (anche le foto e i video, i reality show e i documentari sono fiction), sa altrettanto bene che una narrazione, una fiction, spesso racconta il vero. Il sapere che un'immagine è una finzione permette al suo lettore quella giusta distanza per interpretarne il senso, oltre all'immediatezza dell'impatto emotivo che essa è capace di suscitare. Di certo bisognerebbe armarsi della conoscenza della retorica, che non riguarda solo il linguaggio verbale, ma anche quello visivo. Laddove esiste una dittatura, lo studio della retorica è meglio abolirlo, perché permette di difendersi da argomentazioni fasulle, e lo stesso vale anche per le immagini. Il fatto che vi siamo immersi come una foglia in un torrente in piena non dimostra affatto che siamo in grado di comprenderle. Fondiamo i nostri giudizi su di esse, e chi ben sa usarle a questo fine vince decisive battaglie. Sia nella guerra in Ucraina, sia in quella a Gaza, siamo vittime di tali strategie, e in buona parte l'esito delle stesse battaglie dipende molto anche dal sapiente utilizzo delle immagini. Ad esempio, sulla guerra in Ucraina sono sparite, per lasciare il posto a quella di Gaza, rappresentazioni diverse, ma in realtà le stesse. Immagini utilizzate dai diversi fronti per sconfessare gli avversari. Questo vale anche in altri settori, non solo nell'informazione giornalistica, anche nella diffusione scientifica, su argomenti come il clima, ma pure in altri importanti campi della comunicazione del senso. Siccome, come dicevamo, percepiamo indistintamente tanto le immagini autentiche quanto quelle fasulle, quelle reali e quelle artificiali (ma in un certo senso sono tutte artificiali), è opportuno ricominciare a imparare a pensare non solo con la lingua ma anche con le immagini.

Già Platone
denunciava il potere manipolatorio delle immagini

Il contributo del professor Michele Amadò è il primo di una rubrica che ci accompagnerà lungo l'intero 2024 e che si dedicherà, in collaborazione con USI e SUPSI, proprio alla questione – attualissima – della verità legata alle immagini.