Democrazia *deepfake*

di Raluca Besliu, Monica Pelliccia

Sebastian Burduja, ministro dell'Energia rumeno che esorta a investire in una piattaforma di investimenti inesistente. Immagini esplicite della Premier Giorgia Meloni che circolano su siti pornografici statunitensi. L'attivista ed europarlamentare Ilaria Salis in un video che simula la sua detenzione in una gabbia per cani randagi. Si tratta di video deep fake generati tramite intelligenza artificiale, che hanno il potenziale di mettere a rischio l'intero processo democratico manipolando la realtà e diffondendo disinformazione. E i governi europei corrono ai ripari, con l'approvazione dell'AI Act e iniziative nazionali, come nel caso del Belgio che ha convocato un tavolo di discussione cittadino per offrire consigli al Governo.

Le recenti elezioni europee hanno appena visto l'entrata in campo dell'intelligenza artificiale con conseguenze da valutare. Il Centro per il contrasto dell'odio digitale ha provato che gli strumenti di clonazione vocale tramite Ia potrebbero produrre dichiarazioni convincenti di personaggi politici, come l'audio falso con la voce di Macron che avverte di minacce di bombe ai seggi.
Mentre i giganti tecnologici, come Microsoft, sminuiscono il rischio, l'Unione europea ha appena approvato l'Ai Act, in vigore da marzo 2025, che attribuisce ai deep fake un potenziale di «rischio limitato» per le libertà fondamentali. «Per quanto riguarda i deepfake, ciò che ritenevamo essenziale è che ci sia un'etichettache contrassegni contenutidi Ia», ha spiegato Dragos Tudorache, relatore per l'IA al Parlamento europeo. In tal senso, il professor Cristian Vaccari dell'Università di Edimburgo aggiunge: «Una ragione per la classificazione limitata del rischio è che nelle democrazie occidentali ci si preoccupa di non limitare il discorso politico e la libertà di parola».
Tuttavia**,** diversi studi suggeriscono che anche quando i video o i clip audio vengono etichettati come generati dall'intelligenza artificiale, molte persone continuano a condividerli e a crederci, potenzialmente a causa dei messaggi emotivi che trasmettono.
«L'ambiente psicologico e sociale viene totalmente sottovalutato», spiega Ferdinand Gehringer, esperto della Fondazione Konrad Adenauer, «anche quando le persone sanno che le informazioni sono manipolate, le condividono perché sono divertenti o attirano l'attenzione. Ciò crea effetti psicologici e sociali cumulativi, poiché i deep fake non intendono indirizzare gli elettori verso un particolare candidato o partito, ma piuttosto distrarre, distorcere, diffamare e disturbare, convincendo a non votare affatto o a non farlo per un candidato o un partito specifici».

Il garante italiano per la privacy ha sottolineato la necessità di un organo di controllo indipendente e imparziale a livello nazionale per vigilare sull'attuazione della legge sull'AI Act che, spiega Guido Scorza, avvocato e membro dell'Autorità, «non può essere responsabilità del governo stesso o di un ente governativo. È importante lavorare immediatamente affinché le grandi aziende tecnologiche adottino spontaneamente i principi più importanti contenuti nell'AI Act, senza attendere la sua entrata in vigore».
La disinformazione generata attraverso l'intelligenza artificiale può assumere varie forme, amplificando stereotipi preesistenti nella società. «In Italia i video deepfake sono l'ultima frontiera della violenza di genere», spiega Lilia Giugni, ricercatrice alla University City College di Londra. «Sono diventati parte di un problema sistemico nel nostro paese e in Europa, dove osserviamo come questo tipo di episodi siano diretti contro donne con visioni politiche molto differenti».

Nel marzo 2024, la premier Meloni è stata vittima di un video deep fake sessuale con milioni di visualizzazioni nei siti pornografici americani ed ha intentato una causa che potrebbe costituire un precedente legale, chiedendo 100.000 dollari di risarcimento contro i presunti creatori, un padre e figlio sardi. La premier è apparsa anche in altri video IA, come quello che la ritrae nei panni della cantante Shakira, con quasi un milione e mezzo di visualizzazioni su Tik Tok.
Prima delle elezioni europee, Meloni e la segretaria Pd Elly Schlein sono state protagoniste di video deep fake che promuovevano truffe finanziarie. A seguito della candidatura dell'attivista antifascista Ilaria Salis con Alleanza Verdi Sinistra, la violenza online che la riguarda è sfociata in un video su X, probabilmente generato con IA, che mostra la donna al guinzaglio condotta in un furgone da un accalappiacani.

«Il deep fake riguardante Ilaria Salis è emblematico per l'intersezionalità della sua candidatura tra attivismo, politica e lotta di genere», spiega Silvia Brena, fondatrice di Vox Diritti, associazione che mappa l'odio online da quasi un decennio. «Nella nostra ricerca le donne sono vittime di un messaggio d'odio su due: la misoginia è trasversale anche se gran parte del discorso d'odio globale proviene dall'estrema destra».

Sull'uso politico dei deep fake, la ricercatrice Giugni aggiunge: «Osserviamo leader di destra come Salvini che hanno costruito la loro avventura politica su un certo tipo di retorica macho antifemminista e non hanno avuto scrupoli nell'usare attacchi digitali contro nemici politici e qualsiasi donna visibile. Giorgia Meloni e altre leader di destra si trovano in una situazione complessa perché sono cresciute in partiti in cui la violenza digitale contro le donne viene utilizzata anche ad alti livelli».
Gli stereotipi perpetrati emergono già dalla programmazione dell'Ia, come evidenzia un report dell'Unesco: le donne vengono descritte come impiegate in ruoli domestici molto più spesso degli uomini, associati invece alla carriera.
«La misoginia è usata come arma nei deep fake e nei contenuti generati dall'intelligenza artificiale. Le donne e le giovani sono incredibilmente scoraggiate in Italia e in altri paesi quando si tratta di prendere in considerazione la carriera politica a causa della preoccupazione legata alla loro immagine sui social», spiega Lucina Di Meco, co-fondatrice di *ShePersisted*, un'iniziativa dedicata ad affrontare la disinformazione di genere contro le donne in politica.

Ciononostante**,** i deep fake non hanno ancora rivelato il loro potenziale nel dibattito politico italiano come spiega Lorenzo De Sio, direttore del Centro Italiano Studi Elettorali: «Abbiamo osservato una campagna elettorale europea senza episodi particolari sui social. Alcuni commentatori hanno segnalato un effetto importante delle politiche restrittive attuate dal gruppo Meta che ha messo a tacere i contenuti politici per coloro che non avevano espresso esplicitamente il consenso a vederli».
Raquel Serrano, ricercatrice di DisinfoEu, aggiunge: «Non siamo ancora arrivati al punto di rottura in termini quantitativi per quanto riguarda l'uso delle tecnologie di intelligenza artificiale nella creazione di bufale, compresi i deep fake». Non al punto di rottura, ma il problema in Europa sta prendendo sempre più piede.

Truffe finanziarie, immagini sessualmente esplicite ed ex-premier che tornano in vita. I video *deep fake* generati tramite intelligenza artificiale hanno il potenziale di mettere a rischio l'intero processo democratico manipolando la realtà, diffondendo disinformazione, perpetuando stereotipi di genere in modo da alterare la fiducia dell'elettorato.

In paesi europei come l'Italia, il Belgio e la Romania si sono già verificati casi di video manipolati durante le recenti tornate elettorali.

La Romania, in particolare, affronta quest'anno quattro tornate elettorali: dalle europee alle politiche, le amministrative e infine le elezioni presidenziali, il 24 novembre. Un anno ricco di appuntamenti, in cui figure politiche di spicco sono state già vittime di *deepfake*.
In Romania la fiducia nelle istituzioni è in caduta libera: quella nel governo è crollata dal 34,8% al 19,4% così come quella nel parlamento è scesa dal 26,7% al 17,4% nell'ultimo decennio.
Quest'anno centinaia di migliaia di giovani della generazione Z rumena votano per la prima volta. Dato il loro massiccio utilizzo dei social media potrebbero essere particolarmente vulnerabili a tattiche manipolative digitali.
L'esperta rumena in *deepfake* e sicurezza online Agnes Venema crede invece che: «A rischio non siano le generazioni più giovani che sono esperte di tecnologia, bensì i giovani boomer che non capiscono appieno come operano i social e che potrebbero essere più vulnerabili ai pericoli dei *deep fake*».
Il video deepfake più famoso è stato quello del febbraio 2024, che ritrae il primo ministro Marcel Ciolacu mentre promuove falsi investimenti in Hidroelectrica, la principale compagnia elettrica della Romania.
A distanza di pochi giorni, il ministro dell'Energia Sebastian Burduja è apparso in un altro video manipolato per promuovere la vendita di azioni false di Enel. Burduja ha intrapreso un'azione legale contro i creatori e ha precisato che si è trattato di «un messaggio veicolato per

compromettere e danneggiare» la sua immagine e «che può certamente essere considerato diffamazione o influenzare i risultati elettorali».
Il governo rumeno ha proposto una legge che prevede da 6 mesi a 2 anni di reclusione per la creazione di *deepfake*, criticata dalla società civile perché simile al rigido "modello cinese" dove si criminalizzano anche contenuti legali come la satira politica. Inizialmente approvato al Senato nel 2023, è stato rinviato per la revisione alla Camera dei deputati nel febbraio 2024.
Il ministero rumeno per la Digitalizzazione ha annunciato l'istituzione di uno sportello che consente alla cittadinanza di segnalare gli episodi di *deep fake* elettorali tramite il proprio sito web. Un team di dieci professionisti esaminerà queste segnalazioni e, se ritenute valide, le inoltrerà a piattaforme come Facebook e TikTok per la rimozione.
Un'operazione non priva di polemiche: Bogdan Manolea, direttore esecutivo dell'Associazione per la tecnologia e internet (Apti) ha criticato il processo di selezione di questi professionisti sottolineando che «non è stato seguito il giusto processo di trasparenza e informazione pubblica sulla loro identità».
In questo anno strategico per la Romania, le discussioni con le piattaforme tech per la rimozione dei deep fake sono in corso. Marian Andrei, conduttore del popolare programma televisivo di approfondimento digitale *I like IT*, ha osservato che «i deepfake vengono promossi sui social con le sponsorizzazioni, il che significa che aziende come Facebook, Google e Tik Tok guadagnano soldi fianco a fianco con i creatori di deepfake».
Per i giganti tecnologici continuare a pubblicare questi video manipolati rimane un affare che genera profitti e rende ancora più difficile il contrasto alla disinformazione online.

Come la Romania, anche il Belgio riconosce l'urgenza di contrastare i deepfake. Il Paese è stato uno dei primi in Europa ad assistere a casi di video politici manipolati con intelligenza artificiale. Nel 2018, il partito social democratico Sp.a ha pubblicato un video manipolato di Donald Trump che esortava il Belgio a lasciare gli accordi di Parigi sul clima. E l'ambiente è stato il tema al centro anche del *deepfake* pubblicato da Extinction Rebellion Belgio nel 2020 modificando un discorso della premier Sophie Wilmès sulla diffusione del Covid.
Alla fine del 2023, il partito cristiano-democratico fiammingo CD&V ha pubblicato un *deepfake* elettorale con protagonista il defunto ex primo ministro Jean-Luc Dehaene, riportandolo in vita in un video concordato con la famiglia.
Questa serie di casi hanno svelato il potenziale disinformativo dei *deepfake*, per questo il Belgio ha intrapreso diverse strategie per il loro contrasto in un quadro di governance multilivello.
Il governo ha coinvolto la cittadinanza nel dibattito, convocando un comitato sull'Ia. Per tre settimane, sessanta cittadini e cittadine selezionate in modo casuale, con diversi livelli di conoscenza tecnologica, hanno discusso di *deepfake* e altre questioni legate all'Ia insieme a importanti accademici ed esperti del settore per consigliare direttamente il governo.
Il loro lavoro è culminato in un report consegnato all'esecutivo, ai politici e alla stampa. In merito ai deepfake, il gruppo ha chiesto un cambiamento importante, sottolineando che dovrebbero essere considerati «ad alto rischio» nell'AI Act, non l'attuale designazione di «rischio limitato», ed esortando i governi e l'Ue a porre la questione in cima alle loro agende.
Inès da Camara Santa Clara Gomes, addetta alla presidenza belga dell'Ue nel 2024 e organizzatrice del tavolo di discussione sull'IA, ha precisato che «la visione presentata nella relazione del *panel* ha influenzato il contributo del Belgio all'agenda strategica dell'Ue e probabilmente modellerà l'approccio e le posizioni future sull'intelligenza artificiale».
Il governo belga ha inoltre nominato un comitato sull'etica dei dati e sull'Ia, con l'obiettivo di fornire consulenza scientifica su questioni etiche, legali, economiche, sociali e ambientali legate a questa tecnologia.
La discussione cittadina è stata pioniera a livello europeo. Nel continente il contrasto della disinformazione online tramite deepfake rimane in evoluzione, specialmente in termini di diritti umani e discriminazione.

Costanza Nardocci, professoressa associata di diritto costituzionale all'Università degli Studi di Milano e parte del progetto Humanall4AI, che studia le relazioni tra IA e diritti, sostiene: «L'Ai Act si coordina poco con le direttive di diritto antidiscriminatorio dell'Ue ed è meno sensibile all'approccio *human-rights-based* (basato sui diritti umani, *ndr*) del trattato sull'IA del Consiglio d'Europa». «Servirebbe maggiore attenzione a come la 'macchina' interferisce con il principio di

non discriminazione, con le persone e i loro diritti, analizzando le correlazioni che l'Ia instaura tra caratteristiche protette e gli elementi che la 'macchina' utilizza per compiere le sue decisioni. È fondamentale riconoscere, per poi contenere, ogni effetto discriminatorio».

Questo articolo ha ricevuto il sostegno del Display Europe Grant promosso dalla European Cultural Foundation

il manifesto, 9-10 luglio 2024