Se non vedo, non credo?
L'AI Act e la regolazione del deepfake.

di Lorenzo Berto

Sommario: 1. Introduzione. 2. Il deepfake nell'aia. 3. I deepfake detector. 4. Conclusioni.

"Se non vedo nelle sue mani il segno dei chiodi e non metto il dito nel posto dei chiodi e non metto la mia mano nel suo costato, non crederò" (Gv 20, 25.27).

1. Introduzione

Il termine deepfake descrive contenuti audiovisivi creati o manipolati con tecniche di intelligenza artificiale note come *deep learning*[1]. Più precisamente, nella maggior parte dei casi i programmi di deepfake si basano sulle reti avversarie generative ("GAN"); le GAN sono coppie di reti neurali di apprendimento automatico in grado – tra l'altro – di analizzare una serie di immagini e di crearne di nuove con un livello di qualità comparabile. In estrema sintesi, queste reti imparano cercando di migliorarsi a vicenda: il sistema della coppia c.d. "generatore" crea un output (ad esempio, un'immagine) basato sulla programmazione umana iniziale. L'altra rete, il "discriminatore", è stata programmata per stabilire quale dovrebbe essere l'output corretto; il discriminatore valuta l'output e lo critica. È probabile che gli output iniziali del generatore siano imprecisi; il feedback del discriminatore viene quindi incorporato nel generatore, che continua ad aggiornare i risultati; il ciclo di feedback prosegue finché il generatore non produce dati che il discriminatore ritiene conformi alle aspettative di qualità[2].

Nel corso degli ultimi anni, alcuni sviluppi tecnologici hanno determinato una svolta nelle capacità di manipolazione delle immagini, dando al deepfake una centralità inedita. Da un lato, infatti, i programmatori hanno sviluppato algoritmi in grado di mappare automaticamente – in una data immagine che ritrae una persona – i punti di riferimento del viso, aprendo la strada a tecniche di riconoscimento facciale di cui si è scritto anche su questa rivista[3]. Dall'altro lato, la presenza crescente nelle nostre vite di piattaforme di condivisione di video e foto ha reso disponibili grandi quantità di dati audiovisivi, di cui l'intelligenza artificiale si nutre. Ma il fattore che sovente determina l'ingresso di una tecnologia nel novero

[1] EUROPEAN PARLIAMENTARY RESEARCH SERVICE, *Tackling deepfakes in European policy*, 2021.
[2] B. LEONG, S.R. JORDAN, *The Specturm of Artificial Intelligence*, 2020, Future of Privacy Forum AI Infographic.
[3] M. LO MONACO, J. SCIPIONE, *Anatomia legale del riconoscimento facciale*, 2022, Laboratorio sulla Transizione Digitale della Fondazione Leonardo Civiltà delle Macchine.

delle questioni da risolvere è l'uso pervasivo della medesima, ciò che è puntualmente successo con il deepfake: sono oggi numerosi i siti internet e le app che consentono a chiunque di realizzare un deepfake, senza che sia richiesta particolare dimestichezza con i linguaggi di programmazione informatica[4].

Insomma, oggi il deepfake consente a chiunque di apparire diverso da come è, di pronunciare frasi che non ha mai pronunciato; soprattutto, permette di manipolare quello che altri hanno detto o fatto. Ciò che, evidentemente, minaccia la radicata abitudine a ritenere assai più credibile un fatto perché se ne ha una prova visiva, anziché un racconto[5].

La letteratura in materia raccoglie ormai numerosi esempi di utilizzi di tale tecnologia che si sono rivelati impropri, nel migliore dei casi, o dannosi (quando non illegali) nelle ipotesi peggiori. Così il lettore potrà essersi imbattuto in cronache di *revenge porn* (e cioè di pornografia non consensuale, o addirittura di rappresaglia, compiuta da ex partner), oppure in casi di manipolazione politica[6].

Va detto che anche l'aneddotica sugli utilizzi vantaggiosi della tecnologia deepfake non è trascurabile. *In primis*, l'industria cinematografica può beneficiarne in molti modi: si potranno creare nuovi film con attori morti da tempo, utilizzare effetti speciali e un editing avanzato che non richiede di ripetere le scene. Il deepfake consente, inoltre, il doppiaggio automatico e realistico delle voci nei film in qualsiasi lingua, permettendo così a un pubblico eterogeneo di fruire meglio di film e media educativi[7].

---

[4] È sufficiente una banale ricerca su Google per accorgersene; si possono citare, in via esemplificativa, i seguenti software: Deep Art Effects e MyVoiceYourFace, disponibili rispettivamente su Apple/Play Store e online.

[5] K. GEDDES, *Ocularcentrism and Deepfakes: Should Seeing Be Believing?*, 2021, Fordham Intellectual Property, Media and Entertainment Law Journal, Vol. 31 N. 4; W. GALSTON, *Is Seeing Still Believing? The Deepfake Challenge to Truth in Politics*, 2020, Brookings.

[6] Manipolazione che può essere bonaria, quale il deepfake del 2018 creato dal regista hollywoodiano Jordan Peele che mostrava l'ex presidente degli Stati Uniti Barack Obama discutere dei pericoli delle fake news e prendere in giro Donald Trump, in allora presidente in carica: https://www.youtube.com/watch?v=cQ54GDm1eL0; o pericolosa, quale il video del 2019 in cui l'eloquio della *speaker* della Camera USA Nancy Pelosi era stato artificialmente rallentato, in modo da metterne in dubbio la lucidità: https://www.youtube.com/watch?v=sDOo5nDJwgA.

[7] A tal proposito, si ricordi la campagna globale di sensibilizzazione sulla malaria del 2019, in cui l'icona sportiva e di eleganza David Beckham – grazie a una tecnologia di alterazione visiva e vocale – si rivolgeva alle persone in moltissime lingue diverse, senza doppiaggio: https://www.youtube.com/watch?v=QiiSAvKJIHo. Questo utilizzo della tecnologia deepfake può tornare utile agli sviluppatori di metaversi, intenti a creare un universo digitale in cui abiteremo, almeno parzialmente, per mezzo di un avatar, che potrebbe interagire con altri avatar provenienti da tutto il modo senza dover ricorrere a una lingua franca, alla traduzione o all'interpretariato (si possono leggere in questo senso gli sforzi massicci di Meta nello sviluppo di sistemi di AI di traduzione simultanea: https://about.fb.com/news/2022/07/new-meta-ai-model-translates-200-languages-making-technology-more-accessible/).

E ancora: i deepfake possono aiutare ad affrontare la perdita di persone care riportandole "in vita" digitalmente[8]. Inoltre, possono permettere alle persone transgender di vedersi meglio nel loro genere; possono aiutare a creare voci digitali per coloro che hanno perso la propria a causa di una malattia; la tecnologia deepfake può persino aiutare le persone affette da Alzheimer a interagire con un volto più giovane che potrebbero ricordare[9].

Con la speranza di non frenare l'ottimismo sul potenziale di questa tecnologia, corre l'obbligo di ricordare che, almeno sino al 2020, il 69% dei contenuti deepfake consisteva in *revenge porn*[10]. Fioccano, poi, gli utilizzi del deepfake a scopo di disinformazione o misinformazione, particolarmente preoccupanti nel contesto occidentale (per ovvie ragioni punto di vista adottato qui) lacerato da minacce al normale svolgimento delle elezioni[11], dalla diffusione online di notizie false – soprattutto durante i primi anni di pandemia da Covid-19 – foriera di effetti assai tangibili anche offline[12], e ora anche dalla guerra. Non a caso, dopo pochi giorni di conflitto in Ucraina, è comparso un video deepfake che ritraeva il presidente ucraino Zelensky nell'atto di annunciare la resa all'invasore russo[13].

Non stupisce, dunque, che da molte parti siano giunti appelli per la messa al bando immediata di tale tecnologia, prima della sua diffusione di massa, escludendo così in principio il tentativo di regolarla.

A parere di chi scrive le (numerose e serie) minacce che il deepfake pone ai diritti soggettivi e alla collettività non costituiscono una buona ragione per astenersi dal tentare di affrontarle.

La materia è ampia e piuttosto sfaccettata: la lesione dei diritti soggettivi (diritti della personalità, della proprietà intellettuale, della reputazione, etc.) differisce dalle minacce alla collettività (sicurezza nazionale, tenuta della democrazia, salute dell'ecosistema dell'informazione, etc.). Nel prosieguo, abbiamo scelto di concentrarci sulla risposta che l'Unione Europea intende offrire alla pericolosità, per la società tutta, del deepfake. Il momento è peraltro fecondo: il più importante testo legislativo che affronta il tema oggetto

[8] Non si tratta di deepake, ma un progetto di sviluppo riguardante il sistema di assistenza vocale Alexa, sviluppato da Amazon, è parso voler andare in questo senso: https://www.ilpost.it/2022/06/24/alexa-sintesi-vocale-deepfake-audio-amazon/. Il tema sembra quindi non trascurabile.
[9] EUROPEAN PARLIAMENTARY RESEARCH SERVICE, *Tackling deepfakes in European policy*, 2021, pag. 28.
[10] N. SCHICK, *Deepfakes: The Coming Infocalypse*, 2020, Twelve.
[11] C. TENOVIE, J. BUFFIE, S. MCKAY, D. MOSCROP, *Digital Threats to Democratic Elections: How Foreign Actors Use Digital Techniques to Undermine Democracy*, 2018, Centre for the Study of Democratic Institutions, University of British Columbia.
[12] S. VÉRITER, C. BJOLA, J. A. KOOPS, *Tackling COVID-19 Disinformation: Internal and External Challenges for the European Union*, 2020, The Hague Journal of Diplomacy.
[13] Il video è disponibile qui: https://www.youtube.com/watch?v=X17yrEV5sl4.

di questo approfondimento è ancora in fase di lavorazione, sicché la riflessione può assumere anche una dimensione di *policy*.

2. Il deepfake nell'AIA

La proposta di regolamento dell'Unione Europea sull'intelligenza artificiale, nota come AI Act[14] (di seguito "AIA") è, probabilmente, il più avanzato tentativo globale di regolare in maniera onnicomprensiva l'intelligenza artificiale, di cui viene offerta una definizione: "*un software sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono*" (cfr. art. 3.1 AIA).

L'AIA, che ha già ricevuto più di 3.000 emendamenti da parte dei membri del Parlamento Europeo, di cui si discuterà in autunno, adotta un approccio basato sul rischio; la AI è infatti divisa in quattro gruppi di rischio, cui conseguono divieti di utilizzo, di vendita, o differenti gradi di *compliance*, a seconda dei rischi che, secondo l'UE, una certa AI può comportare. Più precisamente, i livelli di rischio sono: rischio inaccettabile; rischio alto; rischio limitato; rischio minimo[15].

Per quanto qui rileva, l'attenzione va posta sull'art. 52.3 dell'AIA, ove si stabilisce che "*Gli utenti di un sistema di IA che genera o manipola immagini o contenuti audio o video che assomigliano notevolmente a persone, oggetti, luoghi o altre entità o eventi esistenti e che potrebbero apparire falsamente autentici o veritieri per una persona ("deep fake") sono tenuti a rendere noto che il contenuto è stato generato o manipolato artificialmente*".

La norma prosegue disponendo che "*Tuttavia il primo comma non si applica se l'uso è autorizzato dalla legge per accertare, prevenire, indagare e perseguire reati o se è necessario per l'esercizio del diritto alla libertà di espressione e del diritto alla libertà delle arti e delle scienze garantito dalla Carta dei diritti fondamentali dell'UE, e fatte salve le tutele adeguate per i diritti e le libertà dei terzi*".

Ci troviamo nel titolo IV, dedicato ai sistemi a rischio limitato.

L'articolo si rivolge agli utenti professionali dei sistemi deepfake e cioè, ai sensi dell'art. 3.4 AIA, a "*qualsiasi persona fisica o giuridica, autorità pubblica, agenzia o altro organismo che utilizza un sistema di IA sotto la sua autorità, tranne nel caso in cui il sistema di IA sia utilizzato nel corso di un'attività*

[14] La bozza è disponibile qui: https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:52021PC0206&from=EN.

[15] L'approccio basato sul rischio è ben presentato nel seguente contributo: M. VEALE, M. Z. BORGESIUS, *Demystifying the Draft EU Artificial Intelligence Act*, 4/2021, Computer Law Review International.

personale non professionale". Si tratta quindi del soggetto che si serve professionalmente del software, non del programmatore.

Dunque, ai soggetti così individuati è consentito utilizzare il deepfake, a patto che rispettino alcuni obblighi di trasparenza, primo fra tutti quello di comunicare che il contenuto audiovisivo è artefatto[16].

Ai fini del presente contributo, vi è un punto che merita di essere commentato: il perimetro soggettivo di applicazione delle norme dedicate ai contenuti deepfake.

Ci si chiede, in particolare, se l'impostazione scelta dall'UE sia condivisibile. Infatti, la pericolosità data dalla diffusione dei deepfake dipende dal fatto che pressoché chiunque è in grado di creare e diffondere in rete materiale artefatto potenzialmente pericoloso per la collettività. Ci pare che l'ipotesi in cui i contenuti deepfake siano creati professionalmente con lo scopo di inquinare il dibattito pubblico sia rara, mentre assai più frequente è il caso in cui a farlo siano gruppi più o meno organizzati e motivati da ragioni politiche, che solo tangenzialmente incontrano la logica di profitto. Non sembra che tale ultima fattispecie ricada nell'ambito di applicazione dell'art. 52.3, in forza del requisito dell'utilizzo professionale, con il risultato che – ad esempio – gli operatori dell'industria cinematografica sono soggetti all'onere di trasparenza, mentre un gruppo antidemocratico organizzatosi in rete e concentrato sulla destabilizzazione di un Paese, anziché sul profitto personale, non lo sarebbe.

Non ci risulta che la dottrina si sia occupata di questo specifico aspetto, ma a parere di chi scrive sarebbe opportuno avviare un dibattito in merito all'ambito di applicazione soggettiva della disposizione in commento, spostando gli oneri di *compliance* sulle app e sui siti internet che permettono a chiunque di utilizzare la tecnologia deepfake, e cioè sui fornitori[17], soluzione peraltro adottata dall'AIA con riferimento ai bot[18]; sfuggono le ragioni che hanno portato alla differenza di trattamento.

---

[16] L'obbligo di dichiarare l'utilizzo del deepfake si pone peraltro in linea con un suggerimento della dottrina: Pasquale, infatti, riprendendo le leggi della robotica di Asimov, ha formulato i seguenti enunciati: "*Rule Two: Robotic systems and AI should not counterfeit humanity. Rule Four: Robotic systems and AI must always indicate the identity of their creator(s), controller(s), and owner(s)*". Cfr. F. PASQUALE, *New Laws of Robotics*, 2020, Harvard University Press.

[17] I quali sono definiti dall'AIA, all'art. 3, punto 2, come "*una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che sviluppa un sistema di IA o che fa sviluppare un sistema di IA al fine di immetterlo sul mercato o metterlo in servizio con il proprio nome o marchio, a titolo oneroso o gratuito*".

[18] I bot sono programmi automatizzati che eseguono compiti ripetitivi; solitamente con tale termine ci si riferisce a software utilizzati nella interazione scritta tra uomo e macchina, giacché i bot sono in grado di rispondere – in maniera più o meno sofisticata – alle domande che vengono loro poste. A tal proposito, l'art. 52.1 AIA recita: "*I fornitori garantiscono che i sistemi di IA destinati a interagire con le persone fisiche siano progettati e sviluppati in modo tale che le persone fisiche siano informate del fatto di stare interagendo con un sistema di IA, a meno che ciò non risulti evidente dalle circostanze e dal contesto di utilizzo*".

Ad ogni buon conto, tali oneri di *compliance* dovrebbero a nostro avviso spingersi sino alla regolazione parziale del codice, imponendo l'obbligo di includere nel software misure tecnologiche, quali il *watermarking* e l'utilizzo di metadati, che segnalino la natura artefatta del contenuto, in modo da rendere agevole (e – viene da dire – concretamente possibile) il rilevamento dei deepfake e la loro rimozione tramite l'adozione di altri strumenti tecnologici, che verranno analizzati nel paragrafo successivo (sempre, naturalmente, nella misura in cui la legge consenta tale attività di *content policing*, tema altrettanto scottante in questo momento nell'Unione Europea).

Infatti, l'onere di indicazione della natura artefatta del contenuto multimediale, preso singolarmente, è ritenuto insufficiente, anche sulla scorta dell'esperienza delle c.d. *label* utilizzate su base volontaria dalle piattaforme digitali per contrassegnare notizie false o ingannevoli riguardanti la pandemia da Covid-19[19]. D'altronde, se lo scopo è rintracciare i contenuti che non sono stati correttamente segnalati come deepfake, non ci si può basare sulle etichette apposte sui contenuti medesimi.

3. I deepfake detector

Come si è accennato poc'anzi, oltre alle misure giuridiche di contrasto alla diffusione di contenuti deepfake, vi sono quelle tecnologiche, che dovrebbe integrarsi con le prime.

Infatti, la minaccia che il deepfake pone alla ecologia dello spazio in cui le conversazioni online si svolgono rende cruciale l'attività di riconoscimento di tali contenuti, al fine di rimuoverli o di prevenirne la diffusione. Il monitoraggio può essere eseguito manualmente, e cioè grazie al lavoro di esseri umani, o automaticamente, tramite l'utilizzo dell'AI (o ancora, naturalmente, grazie all'ibridazione tra i due modelli).

Schematizzando, possiamo affermare che le soluzioni manuali possono essere impiegate per un numero circoscritto di contenuti; al contempo, offrono i benefici di un approccio sartoriale, capace di comprendere il contesto (ad esempio: un deepfake è satirico o una fake news? Si tratta di un'opera artistica o di un contenuto offensivo?). Le soluzioni automatiche, d'altro canto, consentono di gestire le mastodontiche quantità di contenuti audiovisivi caricati dagli utenti (basti pensare che, sulla sola piattaforma YouTube, ogni minuto vengono caricate – si stima – 400 ore di video) e, ne consegue, non offrono la stessa sensibilità semantica della soluzione manuale.

[19] A. FERNANDEZ, Regulating Deep Fakes in the Proposed AI Act, 2022, MediaLaws: https://www.medialaws.eu/regulating-deep-fakes-in-the-proposed-ai-act/.

La letteratura su questi strumenti è vasta e impossibile da riassumere in poche righe. Ai nostri fini è sufficiente tenerne a mente alcuni nei loro tratti generali e ricordare che, data la quantità di contenuti deepfake che ci troveremo a fronteggiare, non si può prescindere dal loro utilizzo, in qualche misura.

Tra le tecnologie idonee a scovare contenuti modificati artificialmente ricordiamo, dunque[20]:

- Automatic Speaker Verification: è una tecnologia che può verificare l'autenticità di una voce, comparando l'audio sottoposto con un frammento originale con cui la macchina è stata "allenata".
- Voice Liveness Detection: si tratta di una tecnologia in grado di verificare se un campione audio deriva da una persona che sta parlando in diretta o se la voce è registrata.
- Facial Recognition: così come questa (controversa[21]) tecnologia è in grado di riconoscere, secondo un certo grado di probabilità, i cittadini confrontando immagini catturate da videocamere con un database fotografico, allo stesso modo possibile utilizzare sistemi analoghi per verificare se la persona che compare in un video è autentica.
- Rilevatori di incoerenze: sovente un video deepfake è ottenuto modificando pezzi di filmati preesistenti, sicché un sistema automatico può rilevare improvvisi cambiamenti nella postura, movimenti innaturali, labiale disallineato.
- Analisi dei metadati: la lavorazione del filmato di partenza lascia tracce, metadati, che possono essere utilizzati per distinguere un contenuto autentico da uno deepfake.

Ebbene, il trattamento riservato alle AI che possono svolgere il ruolo di deepfake detector è assai diverso da quello riservato ai deepfake generator. L'Allegato III all'AIA classifica come ad alto rischio le tecnologie di *deepfake detection*, se impiegate dalle c.d. autorità di contrasto (autorità pubbliche o incaricate dal potere pubblico di eseguire attività di indagine o perseguimento di reati: cfr. art. 3, para. 40), perché l'utilizzo di sistemi automatici o – in taluni casi – non trasparenti[22] di rilievo e rimozione di contenuti deepfake postati in rete dagli utenti pone delle delicate questioni di tutela della liberà di espressione.

[20] EUROPEAN PARLIAMENTARY RESEARCH SERVICE, *Tackling deepfakes in European policy*, 2021, pagg. 18.

[21] Si veda, ancora, *inter alia*, M. LO MONACO, J. SCIPIONE, *Anatomia legale del riconoscimento facciale*, 2022, Laboratorio sulla Transizione Digitale della Fondazione Leonardo Civiltà delle Macchine.

[22] È noto il problema della cosiddetta *black-box* dei sistemi di *machine learning*: semplificando molto, un dato sistema di *machine learning* risolve il compito affidato sviluppando soluzioni autonome, che partono senz'altro dai dati che e dalle istruzioni di partenza forniti dal programmatore, ma che non sono state programmate una volta per tutte. Si può dire, insomma, che la macchina apprenda da sé le regole necessarie alla soluzione di un dato problema, e la soluzione raggiunta non è del tutto intelligibile dal programmatore stesso, che non è in grado di sapere come il sistema "ragioni".

Per le AI ad alto rischio l'AIA prevede principalmente un sistema di gestione del rischio medesimo, di *governance* dei dati utilizzati dalla macchina, di documentazione dell'attività e di sorveglianza umana[23].

Ne consegue che l'impiego di deepfake detector da parte di soggetti privati non è sottoposto alle medesime restrizioni; si tratterà in questo caso di una AI a rischio limitato[24].

A nostro giudizio si tratta di una scelta coerente con i numerosi obblighi di *content moderation* che l'Unione Europea ha introdotto o sta per introdurre. Si pensi, ad esempio, all'art. 28 *ter* della Direttiva 2018/1808 sui servizi audiovisivi, recepito negli Stati Membri, che impone ai fornitori di piattaforme per la condivisione di video di adottare misure adeguate per tutelare, anche in via preventiva, i minori e il grande pubblico da alcuni contenuti ritenuti inidonei; o alle ambizioni del Digital Services Act, approvato dal Parlamento Europeo il 5 luglio ultimo scorso, ai sensi del quale le "*piattaforme online che prestano i loro servizi a un numero medio mensile di destinatari attivi del servizio nell'Unione pari o superiore a 45 milioni*" (cfr. art. 25 DSA) devono adottare "*misure di attenuazione ragionevoli, proporzionate ed efficaci, adattate ai rischi sistemici specifici*" (art. 27), tra cui la diffusione di contenuti illegali e "*la manipolazione intenzionale del servizio, anche mediante un uso non autentico o uno sfruttamento automatizzato del servizio, con ripercussioni negative, effettive o prevedibili, sulla tutela della salute pubblica, dei minori, del dibattito civico, o con effetti reali o prevedibili sui processi elettorali e sulla sicurezza pubblica*" (art. 26). O, ancora, al dibattutissimo art. 17 della c.d. Direttiva Copyright (2019/790), recepito negli Stati Membri, che impone ai fornitori di servizi di condivisione di contenuti online (o piattaforme) – sotto la propria responsabilità – di procurarsi dal titolare dei diritti l'autorizzazione a condividere le opere protette caricate dagli utenti, così imponendo un onere di verifica circa la liceità degli *upload*, dal punto di vista del diritto d'autore[25].

Orbene, se le piattaforme fossero sottoposte agli stessi limiti che l'AIA impone – nell'utilizzo dei deefake detector – alle autorità di controllo, si verificherebbe una situazione paradossale: da un lato, le norme che si occupano della liceità dei contenuti che circolano online imporrebbero oneri di controllo e rimozione, talvolta preventivi, i quali sono possibili – dato il volume dei contenuti caricati ogni secondo dagli utenti – solo se realizzati anche con l'aiuto di strumenti di intelligenza artificiale; dall'altro lato, le disposizioni che regolano tali tecnologie ne limiterebbero fortemente l'utilizzo[26].

---

[23] Cfr. Titolo III dell'AIA.

[24] A. FERNANDEZ, *Regulating Deep Fakes in the Proposed AI Act*, 2022, MediaLaws: https://www.medialaws.eu/regulating-deep-fakes-in-the-proposed-ai-act/.

[25] Ci siamo occupati di questo tema in un recente articolo: L. BERTO, L'art. 17 della direttiva copyright è legittimo, 2022, CyberLaws: https://www.cyberlaws.it/2022/art-17-direttiva-copyright-legittimo/.

[26] Con ciò non si vuole dire che l'utilizzo di tali strumenti sia lineare e privo di problemi anche giuridici; semplicemente, si sottolinea che un profilo di incoerenza è stato eliminato grazie all'impostazione adottata

4. Conclusioni

I deepfake sono presenti in maniera significativa da qualche anno; la sempre crescente quantità di dati (e, nella fattispecie, di immagini e file audio personali) disponibile e la costante semplificazione degli strumenti di manipolazione di tali dati, tuttavia, renderà le sfide poste da questa intelligenza artificiale sempre più pressanti.

Da questo punto di vista, il fatto che la discussione sull'AIA stia per entrare nel vivo offre un'occasione per affrontare alcune di queste sfide.

L'AIA, infatti, regola la generazione del deepfake e l'utilizzo dei deepfake detector. In questo contributo, abbiamo evidenziato alcuni punti critici che riguardano la regolazione della creazione dei contenuti deepfake: segnatamente, la scelta dell'Unione di imporre obblighi di trasparenza agli utenti di software di manipolazione dei contenuti multimediali, anziché ai fornitori di questi, e il fatto che l'obbligo di trasparenza, di per sé, non sia adeguato ai fini dell'*enforcement* delle regole dell'AIA.

Abbiamo, invece, apprezzato la coerenza dell'Unione nel trattare diversamente l'utilizzo dei deepfake detector nel pubblico e nel privato, riconoscendone la necessità alla luce degli altri testi legislativi varati o in discussione a livello dell'Unione.

L'AIA è un testo ambizioso, innovativo e di avanguardia. Come è naturale, ha raccolto molti suggerimenti e critiche, da fuori e da dentro alle istituzioni. All'interno della complessità della materia, un aspetto specifico come il deepfake è stato forse meno osservato dai commentatori. Ci sembra invece che, data la rilevanza attuale e futura del fenomeno, la negoziazione sull'AIA sia l'occasione perfetta per sviluppare un dibattito stimolante e idoneo a guidare l'iter legislativo.

dall'Unione Europea, a vantaggio del perseguimento di un obiettivo prefissato, e a prescindere da ciò che si pensi sull'obiettivo medesimo.