La regolamentazione del *deepfake* in Europa, Stati Uniti e Cina*

Alberto Orlando

Abstract

Il vertiginoso aumento dei *deepfake* negli ultimi anni ha portato i regolatori pubblici a interrogarsi sulla necessità di regolamentare il fenomeno, incontrando problematiche simili a quelle che riguardano in generale la regolamentazione dell'IA.
Unione Europea, Stati Uniti e Cina, ossia le potenze che si contendono la posizione dominante in materia di IA, di recente sembrano aver intrapreso la strada della regolamentazione del fenomeno, scegliendo però approcci profondamente diversi tra loro, che riflettono le peculiarità degli ordinamenti e la visione politica e strategica in materia di sviluppo delle nuove tecnologie.
Il presente contributo offre una panoramica delle principali soluzioni normative adottate e mette a confronto i tre approcci, sottolineandone analogie e differenze.

The vertiginous rise of *deepfakes* in recent years has led public regulators to question the need to regulate the phenomenon, encountering similar problems to those affecting AI regulation in general.
The European Union, the United States and China, i.e. the powers that contend for the dominant position in the field of AI, recently seem to have embarked on the path of regulating the phenomenon, choosing, however, profoundly different approaches, reflecting the peculiarities of the legal systems and the political and strategic vision regarding the development of new technologies.
This paper provides an overview of the main regulatory solutions adopted and compares the three approaches, highlighting their similarities and differences.

Sommario

*Su determinazione della direzione, il contributo è stato sottoposto a referaggio anonimo in conformità all'art. 15 del regolamento della Rivista

Keywords

deepfake – intelligenza artificiale – sintesi profonda – regolamentazione dell'IA – disinformazione

1. Introduzione: profili regolatori delle tecnologie di *deepfake*

In una società in cui la circolazione di notizie e contenuti falsi, soprattutto attraverso il *web*, non è certo un fatto nuovo[1], i *deepfake* si ritagliano uno specifico spazio che li distingue da altri fenomeni.

Essi possono essere definiti come output derivanti dall'utilizzo di tecniche di intelligenza artificiale (di seguito, IA) atto a generare audio e video sintetici ma estremamente realistici, soprattutto per quanto riguarda le somiglianze facciali e vocali umane. Poiché la tecnologia è sempre più disponibile per la sperimentazione da parte di chiunque possieda un minimo di competenze, i *deepfake* sono oggi utilizzati in ambiti di applicazione via via crescenti[2], dall'intrattenimento, alla politica, alla pornografia, ecc. Sebbene i contenuti di alta qualità richiedano un'ampia quantità di dati oltre a eccellenti competenze da parte del creatore del contenuto, in realtà anche i *deepfake* di bassa qualità possono rivelarsi ugualmente dannosi[3].

Ben si comprende che, sebbene possano rintracciarsi alcuni utilizzi "positivi"[4], gli impatti negativi sembrano di gran lunga superiori e molto evidenti: ad esempio, danni emotivi, furto di identità, danni alla reputazione, manipolazione politica. Tutti effetti

[1] Si pensi alla questione delle *fake news*, sulla cui regolamentazione il dibattito resta costantemente aperto. Cfr., *ex multis*, M. Bassini – G.E. Vigevani, *Primi appunti su* fake news *e dintorni*, in questa *Rivista*, 1, 2017, 11 ss. e gli altri contributi contenuti nella stessa sezione monografica della rivista.

[2] I contenuti *deepfake* sono generalmente creati utilizzando le reti generative avversarie (GAN), una tecnologia creata da Ian Goodfellow nel 2014. L'ascesa della tecnologia *deepfake* e dell'utilizzo da parte dei consumatori è iniziata nel 2017 sul sito web *Reddit*, quando un utente chiamato appunto "*deepfake*" ha pubblicato materiale pornografico manipolato che scambiava i volti di celebrità e personaggi pubblici con quello di altre persone. Questi post ottenevano grande popolarità, tanto che una pagina *Reddit* specializzata, nota come "*Subreddit*", veniva dedicata esclusivamente ai video *deepfake* e raggiungeva rapidamente centinaia di migliaia di membri della comunità. Nel 2018, un *deepfake* che ritraeva l'ex presidente degli USA Barack Obama intento a utilizzare un linguaggio oltraggioso nei confronti di altri politici fece il giro del mondo ingenerando non poche confusioni sulla sua veridicità e portando alla ribalta le potenzialità e i rischi dei cc.dd. *deepfake* "politici". Dal dicembre 2018 il numero dei deepfake online nei successivi due anni è all'incirca raddoppiato ogni sei mesi, confermando una crescita esponenziale del fenomeno. Cfr. M.B. Kugler – C. Pace, *Deepfake Privacy: Attitudes and Regulation*, in *Northwestern University Law Review*, 3, 2021, spec. 620-621; B. Chesney – D. Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, in *California Law Review*, 6, 2019, spec. 1758 ss.; L. Whittaker – R. Mulcahy – K. Letheren – J. Kietzmann – R. Russell-Bennett, *Mapping the deepfake landscape for innovation: A multidisciplinary systematic review and future research agenda*, in *Technovation*, 125, 2023.

[3] Cfr. Y. Geng, *Comparing "Deepfake" Regulatory Regimes in The United States, the European Union, and China*, in *Georgetown Law Technology Review*, 7, 2023, 158.

[4] Cfr. S. Chandler, *Why Deepfakes are a Net Positive for Humanity*, in *forbes.com*, 9 marzo 2020. Tra gli esempi si possono citare il montaggio di video senza riprese, l'esperienza di cose che non esistono più, l'aumento dell'accessibilità per le persone con disabilità, la possibilità di migliorare le pratiche mediche, ecc.

che con buona probabilità arrivano a coincidere con la commissione di condotte illecite, spesso rilevanti anche sul piano penale[5]. In ogni caso, i *deepfake* contribuiscono a quel processo di erosione della fiducia della società nei confronti della tecnologia, del mondo dell'informazione e delle istituzioni politiche. Questa sperequazione tra rischi e benefici connota i *deepfake* rispetto al *mare magnum* dell'IA, laddove la discussione sul rapporto tra vantaggi e svantaggi appare molto più incerta[6].

In questo quadro, il regolatore pubblico deve quantomeno interrogarsi sull'opportunità di regolare il fenomeno del *deepfake*. Tale esigenza si inserisce perfettamente – al netto delle peculiarità appena evidenziate – nella riflessione sulla regolamentazione delle tecnologie di intelligenza artificiale, di cui evidentemente i *deepfake* fanno parte nel momento in cui la loro realizzazione deriva dall'utilizzo di sistemi di IA. Pensare a nuove forme di regolamentazione per i *deepfake* sembra indispensabile anche perché la tutela accordata alla privacy non appare sufficiente: in effetti, pur nelle esperienze assai differenti prese in considerazione in questo contributo[7], sul piano della privacy le condotte illegali si sostanziano in linea generale nella diffusione di dati protetti dalla legge ma comunque veri. Nel caso dei *deepfake*, invece, si assiste ad una parte del contenuto che è vera ma non privata, come ad es. il volto o la voce della persona rappresentata, unita inscindibilmente ad un'altra parte che potrebbe essere privata ma sicuramente non è vera[8].

Al di fuori del contesto della normativa sulla privacy, resta comunque la strada di impedire o limitare i *deepfake* in quanto dati e notizie falsi, in grado di mettere in pericolo interessi pubblici o diritti dei singoli. In questo senso, si pone un problema di necessario bilanciamento con il diritto alla libertà di espressione e tutti i suoi corollari, compreso il diritto di critica e di satira. Dato che la mera falsità non può operare come condizione sufficiente, si potrebbe scegliere di trattare i *deepfake* come contenuti diffamatori, cioè destinati a essere percepiti come veri e dannosi per la reputazione della persona ritratta: in questo caso, l'etichettatura del *deepfake* verrebbe in soccorso, poiché, svelandone la natura, aiuterebbe a negare almeno l'elemento della apparente veri-

---

[5] Esistono anche potenziali aree grigie: ad esempio, l'uso di *deepfake* di celebrità in pubblicità e video di formazione, con o senza il permesso delle celebrità, ha attirato l'interesse di aziende del settore dei media e del marketing per la possibilità di aumentare i profitti a costi inferiori. Cfr. R. Spivak, *"Deepfakes": The Newest Way to Commit One of the Oldest Crimes*, in *Georgetown Law Technology Review*, 3, 2019, 368-383.

[6] Cfr., *ex multis*, S. Russell – P. Norvig, *Artificial Intelligence: A Modern Approach*, Englewood Cliffs, 2020; W. Barfield – U. Pagallo, *Law and Artificial Intelligence*, Cheltenham, 2020; U. Ruffolo (a cura di), *XXVI lezioni di diritto dell'intelligenza artificiale*, Torino, 2021; A. Santosuosso, *Intelligenza artificiale e diritto. Perché le tecnologie di IA sono una grande opportunità per il diritto*, Milano, 2020; A. D'Aloia (a cura di), *Intelligenza artificiale e diritto*, Milano, 2021; C. Casonato, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Diritto pubblico comparato ed europeo*, Speciale, 2019, 101-130; E. Stradella, *La regolazione della Robotica e dell'Intelligenza artificiale: il dibattito, le proposte, le prospettive. Alcuni spunti di riflessione*, in questa *Rivista*, 1, 2019, 73 ss.

[7] Cfr. A. Di Martino, *Profili costituzionali della privacy in Europa e negli Stati Uniti*, Napoli, 2017; U. Pagallo, *La tutela della privacy negli Stati Uniti d'America e in Europa: modelli giuridici a confronto*, Milano, 2008; D. Clementi, *La legge cinese sulla protezione delle informazioni personali. Un GDPR con caratteristiche cinesi?*, in *Rivista di diritti comparati*, 1, 2022, 189-216; E. Bertolini, *L'"apertura sorveglia": la via cinese alla governance e alla censura di Internet*, in *Diritto pubblico comparato ed europeo*, 3, 2008, 1063-1097.

[8] Cfr. Kugler, *Deepfake Privacy*, cit., 613-614.

dicità. Tuttavia, una tale soluzione non sembra esaurire del tutto le problematiche, dato che il danno alla reputazione o alla dignità della persona potrebbe prodursi ugualmente – per quanto, magari, in misura diversa – nel caso di circolazione di contenuti falsi e dichiarati al pubblico come tali: in questo senso, l'etichettatura non sarebbe risolutiva. Senza contare che, spostandosi dalla ragionevolezza della misura regolatoria alla sua effettività, sembra lecito dubitare della reale efficacia delle etichettature, considerato che queste potrebbero essere comunque eliminate in un momento successivo all'apposizione: ipotesi che, nel mondo di Internet in cui gli utenti sono protetti dall'anonimato e ai *provider* non può essere assegnato con leggerezza il ruolo di censori-gestori dei contenuti, rischia di rendere evanescenti le responsabilità.

In effetti, come accade comunemente in materia di utilizzo di IA[9], proprio la ripartizione delle responsabilità resta un punto controverso. In primo luogo, si potrebbe insistere sulla responsabilità dei creatori di *deepfake*, dato che ad essi può ricondursi la paternità dell'opera e, salvo casi eccezionali, gli effetti conseguenti: tuttavia, tale soluzione rischia di rivelarsi vana a fronte dell'anonimato online e della circolazione virale dei contenuti. In secondo luogo, occorrerebbe riflettere sulle responsabilità di utenti differenti dai creatori, comunque in grado di contribuire alla diffusione, alla trasmissione, al download di contenuti non consentiti. In terzo luogo, il centro del dibattito è probabilmente occupato dalle responsabilità (e dai poteri) che potrebbero essere assegnate ai gestori di servizi online, dovendosi distinguere diverse categorie di operatori, ossia, ad esempio, i fornitori di servizi di sintesi profonda che consentono la creazione di contenuti *deepfake* e le piattaforme online (compresi *social network*) che ne consentono la diffusione. Attribuire responsabilità ai *provider* appare soluzione vantaggiosa perché le competenze tecniche di questi soggetti garantirebbero in maniera efficace un controllo sui contenuti anche in via preventiva. Tuttavia, evidentemente ne scaturirebbero alcuni svantaggi: da un lato, si finirebbe per delegare il delicatissimo controllo sui limiti della libertà di espressione a soggetti privati, spesso operanti in un contesto economico transnazionale; dall'altro, agli stessi soggetti sarebbero richiesti standard e strutture particolarmente gravose da sostenere, a fronte di una esposizione a responsabilità anche rilevanti, in un quadro che potrebbe scoraggiare gli investimenti nel settore (almeno nei contesti geografici caratterizzati da normative più "severe")[10].

Cosa fare, quindi? Verso quale direzione dovrebbero orientarsi i regolatori pubblici? O meglio: quale logica dovrebbero seguire nella ricerca delle soluzioni? Anche qui rischia-

[9] Oltre a rimandare alla bibliografia citata *supra*, nota 6, cfr. anche M. Bassini – L. Liguori – O. Pollicino, *Sistemi di Intelligenza Artificiale, responsabilità e* accountability. *Verso nuovi paradigmi?*, in F. Pizzetti (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino, 2018, 333-371; E. Palmerini – E. Stradella (a cura di), *Law and Technology. The Challenge of Regulating Technological Development*, Pisa, 2013; U. Pagallo, *Intelligenza Artificiale e diritto. Linee guida per un oculato intervento normativo*, in *Sistemi intelligenti*, 3, 2017, 615-636; G. Comandè, *Intelligenza artificiale e responsabilità tra liability e accountability. Il carattere trasformativo dell'IA e il problema della responsabilità*, in *Analisi giuridica dell'economia*, 1, 2019, 173 ss.; S. Lohsse – R. Schulze – D. Staudenmayer (a cura di), *Liability for Artificial Intelligence and the Internet of Things*, Baden-Baden, 2019; R. Brownsword – K. Yeung (a cura di), *Regulating Technologies. Legal Futures, Regulatory Frames and Technological Fixes*, Oxford, 2008.

[10] Cfr. J. Riordan, *The Liability of Internet Intermediaries*, Oxford, 2016; G.E. Vigevani, *Piattaforme digitali private, potere pubblico e libertà di espressione*, in *Diritto costituzionale*, 1, 2023, 41-54; L. Albertini, *Sulla responsabilità civile degli internet service provider per i materiali caricati dagli utenti (con qualche considerazione generale sul loro ruolo di gatekeepers della comunicazione)*, in *Medialaws, Law and Media Working Papers Series*, 4, 2020.

mo di trovarci di fronte ad un dilemma. Da un lato, infatti, il fenomeno appare connotato, come accade per tutto ciò che gravita nel *web*, dal carattere della transnazionalità[11], che rende preferibile approntare soluzioni in grado di valere anche al di là dei confini nazionali: questa strada può essere seguita sia per mezzo di un raccordo preventivo tra i legislatori nazionali, sia attraverso lo studio delle normative già vigenti – o in via di discussione – in altri Paesi. Tale elemento appare ancora più centrale se si considera che attorno a queste opzioni legislative si gioca una parte consistente della partita per il primato economico (e culturale) in materia di sviluppo delle nuove tecnologie e segnatamente dell'IA. D'altro canto, il ricorso a regolamentazioni più o meno restrittive deve fare i conti con l'accettazione sociale delle norme introdotte. Da questo punto di vista, l'analisi del contesto valoriale e sociale assume importanza fondamentale, considerato che sovrastimare o sottostimare la criminalizzazione di determinate condotte può sortire l'effetto di degradare la legittimità della legge agli occhi del pubblico e ridurre il rispetto delle regole stabilite[12].

Alla luce di quanto appena detto, il presente contributo, lungi dal prospettare soluzioni definitive rispetto ai problemi proposti, punta proprio a descrivere e analizzare il modo in cui Unione europea, Stati Uniti e Cina – ossia i maggiori *competitor* a livello mondiale in materia di IA[13] – stanno affrontando il fenomeno. Si anticipa sin da ora che in tutti e tre gli ordinamenti negli ultimissimi anni sono emersi, accanto ad orientamenti e studi sui *deepfake*, anche i primi approcci normativi dedicati più o meno specificamente alla materia, i quali rivelano visioni molto diverse, sulla falsariga di quanto sta avvenendo con riguardo alla regolamentazione generale dell'IA[14].

2. Unione europea: tra *tackling* e *risk-based approach*

La legislatura dell'UE conclusa nel 2024 ha affrontato con decisione le sfide regolatorie poste dalle nuovissime tecnologie, scegliendo un approccio unico a livello globale e molto diverso da quello statunitense e cinese: pur non sopprimendo lo spazio per l'autoregolamentazione privata, le Istituzioni europee hanno convintamente intrapreso la strada della regolamentazione "forte", attraverso l'adozione di strumenti di *hard law*

[11] Cfr. O. Pollicino – M. Bassini, *Internet Law in the Era of Transnational Law*, in *EUI Working Papers*, 24, 2011; C. Marsden, *Transnational Internet Law*, Oxford, 2020; G. Teubner, *Nuovi conflitti costituzionali. Norme fondamentali dei regimi transnazionali*, Milano, 2012.

[12] Cfr. Kugler, *Deepfake Privacy*, cit., 615.

[13] Cfr. D. Castro – M. McLaughlin – E. Chivot, *Who Is Winning the AI Race: China, the EU or the United States?*, in *datainnovation.org*, 19 agosto 2019; E. Pisanelli, *Intelligenza Artificiale: battaglia globale per tre*, in *ispionline.it*, 27 ottobre 2022.

[14] Sui differenti approcci in materia di IA, cfr. H. Roberts – J. Cowls – E. Hine – J. Morley – V. Wang – M. Taddeo – L. Floridi, *Governing artificial intelligence in China and the European Union: Comparing aims and promoting ethical outcomes*, in *The Information Society*, 2, 2023, 79-97; E. Hine – L. Floridi, *Artificial intelligence with American values and Chinese characteristics: a comparative analysis of American and Chinese governmental AI policies*, in *AI & Society*, 1, 2024, 257-278; R. Bal – I.S. Gill, *Policy approaches to artificial intelligence based technologies in China, European Union and the United States*, in *Duke Global Working Paper Series*, 26, 2020.

come i regolamenti: dopo il *Digital Markets Act* (DMA)[15] e il *Digital Services Act* (DSA)[16], al momento in cui si scrive è attesa, dopo lunga gestazione, l'entrata in vigore dell'*Artificial Intelligence Act* (AIA)[17].

Se, quindi, con specifico riferimento al tema dei *deepfake* l'UE non è – o non è ancora – intervenuta con un atto normativo *ad hoc*, il fenomeno non resta comunque sconosciuto alle Istituzioni, tanto che all'interno dei regolamenti appena citati possono trovarsi riferimenti più o meno espliciti, nonché disposizioni più o meno applicabili, le quali si aggiungono alle tutele in materia di privacy stabilite dal GDPR[18].

In effetti, già nel 2021 il *Panel for the Future of Science and Technology* (STOA) rendeva al Parlamento europeo indicazioni sui *deepfake* per mezzo di un report intitolato "*Tackling deepfake in European policy*"[19]. Già dal titolo è possibile evincere il disvalore, cui si accennava in apertura, con cui questi utilizzi di IA sono osservati dalla società e dal regolatore stesso: tanto che essi non vanno regolati, ma "contrastati" (*tackling*). Lo studio dello STOA seguiva di pochi mesi la pubblicazione delle prime proposte del DSA e dell'AIA, delle quali evidentemente teneva conto. In particolare, il testo dell'AIA, nella parte in cui fa riferimento ai *deepfake*, è stato modificato solo leggermente rispetto alla versione del 2021: pertanto, le considerazioni esposte nello studio dello STOA non sono state "superate" dal nuovo regolamento. Inoltre, nel report, accanto ai regolamenti in via di approvazione, vengono considerate anche le disposizioni già vigenti del GDPR.

Innanzitutto, i *deepfake* sono definiti come media audio o video manipolati o sintetici che sembrano autentici e che mostrano persone che sembrano dire o fare qualcosa che non hanno mai detto o fatto, prodotti utilizzando tecniche di IA, tra cui l'apprendimento automatico e il *deep learning*[20]. Al netto di un dibattito sempre aperto sulla definizione di IA[21], tale enunciazione non sembra porre particolari problemi applicativi.

Lo STOA rende poi una serie di raccomandazioni. Primariamente, consiglia di chiarire i casi in cui i *deepfake* possano ricomprendersi, ai sensi dell'AIA, tra le pratiche vietate o tra le applicazioni ad alto rischio, valutando l'opportunità di includerli come regola tra i sistemi ad alto rischio e di prevedere specifici *ban* per utilizzi particolarmente pe-

[15] Regolamento (UE) 2022/1925 del Parlamento europeo e del Consiglio del 14 settembre 2022 relativo a mercati equi e contendibili nel settore digitale e che modifica le direttive (UE) 2019/1937 e (UE) 2020/1828 (regolamento sui mercati digitali), noto in inglese come *Digital Markets Act*.

[16] Regolamento (UE) 2022/2065 del Parlamento europeo e del Consiglio del 19 ottobre 2022 relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (regolamento sui servizi digitali), noto in inglese come *Digital Services Act*.

[17] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale), noto in inglese come *Artificial Intelligence Act*.

[18] Cfr. B. Van der Sloot – Y. Wagensveld, *Deepfakes: regulatory challenges for the synthetic society*, in *Computer Law & Security Review*, 46, 2022, 7-8.

[19] STOA, *Tackling deepfakes in European policy*, luglio 2021.

[20] Ivi, 2.

[21] Cfr. C. Muller, *The Impact of AI on Human Rights, Democracy and the Rule of Law*, in CAHAI (Comitato ad hoc sull'IA del Consiglio d'Europa), CAHAI(2020)06-fin, 24 giugno 2020, 4 ss.: «AI remains an essentially contested concept, as there is no universally accepted definition».

ricolosi, come quelli riguardanti la pornografia non consensuale o la disinformazione politica[22]. In secondo luogo, con riferimento all'obbligo di etichettare i contenuti per renderne palese la natura, imposto dall'IA ai creatori (*creators*) di *deepfake*, suggerisce di estendere tale obbligo anche ai fornitori (*providers*)[23]. Inoltre, ammonisce dall'utilizzo di eccezioni troppo ampie rispetto a questo obbligo, dato che la proposta dell'IA esclude l'etichettatura per i contenuti usati a fini di prevenzione dei crimini, ma anche artistici e scientifici e comunque quando necessari per la libertà di espressione[24]. In maniera abbastanza singolare, raccomanda di limitare la diffusione delle tecnologie di rilevamento dei *deepfake*, in modo tale da non consentire ai creatori modalità per eludere i controlli, anche se parallelamente si sconsiglia di riservare la *detection* a una cerchia troppo ristretta di attori[25]; infine, sottolinea l'importanza di investire nelle tecnologie di IA in grado di "difendere" dagli attacchi di *deepfake*, oltre che nell'educazione e nella consapevolezza del fenomeno da parte della società[26].

Sebbene il DSA non menzioni direttamente i *deepfake*, l'impostazione generale del regolamento e alcune disposizioni specifiche sembrano riguardare da vicino questo genere di contenuti. In linea generale, alle piattaforme online e ai motori di ricerca di grandissime dimensioni è imposto l'obbligo di effettuare valutazioni specifiche del rischio e mettere in atto misure ragionevoli, proporzionate ed efficaci per prevenire qualsiasi aspetto negativo effettivo o prevedibile sul discorso civico e sui processi elettorali[27]. Si tratta di un obbligo che teoricamente pare interessare almeno quei *deepfake* incidenti sulla disinformazione politica. Agli stessi soggetti il DSA richiede di adottare idonee misure di attenuazione dei rischi e tra queste elenca esplicitamente «il ricorso a un contrassegno ben visibile per fare in modo che un elemento di un'informazione, sia esso un'immagine, un contenuto audio o video, generati o manipolati, che assomigli notevolmente a persone, oggetti, luoghi o altre entità o eventi esistenti e che a una persona appaia falsamente autentico o veritiero, sia distinguibile quando è presentato sulle loro interfacce online e, inoltre, la fornitura di una funzionalità di facile utilizzo che consenta ai destinatari del servizio di indicare tale informazione»[28]. Il legislatore evita di fare riferimento diretto ai *deepfake*, ma evidentemente ragiona sulla loro circolazione. La scelta di utilizzare una perifrasi tanto ampia si giustifica sia per il fatto di non

[22] STOA, *Tackling deepfakes*, cit., 59.

[23] *Ibid.*

[24] Ivi, 60-61.

[25] Ivi, 59.

[26] Ivi, 60.

[27] DSA, art. 34: «I fornitori di piattaforme online di dimensioni molto grandi e di motori di ricerca online di dimensioni molto grandi individuano, analizzano e valutano con diligenza gli eventuali rischi sistemici nell'Unione derivanti dalla progettazione o dal funzionamento del loro servizio e dei suoi relativi sistemi, compresi i sistemi algoritmici, o dall'uso dei loro servizi. […] La valutazione del rischio […] deve comprendere i seguenti rischi sistemici: […] eventuali effetti negativi, attuali o prevedibili, per l'esercizio dei diritti fondamentali, in particolare i diritti fondamentali alla dignità umana […]; eventuali effetti negativi, attuali o prevedibili, sul dibattito civico e sui processi elettorali, nonché sulla sicurezza pubblica; […] qualsiasi effetto negativo, attuale o prevedibile, in relazione alla violenza di genere, alla protezione della salute pubblica e dei minori e alle gravi conseguenze negative per il benessere fisico e mentale della persona».

[28] DSA, art. 35, par. 1, lett. k).

vincolare piattaforme e motori di ricerca alla messa in atto di misure troppo specifiche, sia probabilmente perché all'epoca dell'adozione del regolamento (2022) l'espressione "*deepfake*" non era ancora così sdoganata come oggi e rischiava di ingenerare equivoci interpretativi. Tuttavia, come effetto, una disposizione così ampia finisce con il riferirsi non soltanto all'IA generativa su cui si fondano i *deepfake*, ma permette interpretazioni eccessivamente estensive che potrebbero mettere in difficoltà i soggetti obbligati.

Con il regolamento sull'IA, come noto, l'UE ha imboccato la strada di un approccio olistico e basato sul rischio per regolamentare tutte le pratiche di intelligenza artificiale[29]. L'AIA prende in esame l'intero ciclo di vita dei sistemi e classifica gli stessi sulla base dei rischi che presentano, distinguendo tecnologie (*rectius*: impieghi delle tecnologie) comunque vietate, ad alto rischio e a basso rischio e prevedendo per ognuna di queste categorie obblighi differenti in capo agli *AI actors*: in particolare, i sistemi ad alto rischio sono soggetti a un regime di conformità con requisiti dettagliati[30]. Occorre verificare a quale/i di queste categorie – pratiche vietate, sistemi ad alto rischio, altri sistemi – possano appartenere i *deepfake*.

Preliminarmente, è doveroso evidenziare che questo genere di contenuti non è esplicitamente menzionato tra le pratiche comunque vietate, né tra i sistemi ad alto rischio, mentre – come si vedrà – essi vengono richiamati in altre disposizioni, oltre ad essere definiti esplicitamente dal regolamento come «un'immagine o un contenuto audio o video generato o manipolato dall'IA che assomiglia a persone, oggetti, luoghi, entità o eventi esistenti e che apparirebbe falsamente autentico o veritiero a una persona»[31]. In ogni caso, la disciplina sulle pratiche vietate e sui sistemi ad alto rischio merita di essere analizzata per verificare se possa escludersi completamente l'appartenenza dei *deepfake* a queste categorie.

Con riferimento alle pratiche vietate ai sensi dell'art. 5 AIA, il regolamento mette al bando sia i sistemi che utilizzano «tecniche subliminali che agiscono senza che una persona ne sia consapevole o tecniche volutamente manipolative o ingannevoli aventi lo scopo o l'effetto di distorcere materialmente il comportamento di una persona o di un gruppo di persone, pregiudicando in modo considerevole la loro capacità di prendere una decisione informata, inducendole pertanto a prendere una decisione che non avrebbero altrimenti preso […]»[32], sia i sistemi che sfruttano «le vulnerabilità di una persona fisica o di uno specifico gruppo di persone, dovute all'età, alla disabilità o a una specifica situazione sociale o economica, con l'obiettivo o l'effetto di distorcere materialmente il comportamento di tale persona o di una persona che appartiene a tale gruppo […]»[33]. Non appare impossibile immaginare che i *deepfake* possano presentarsi in queste forme, anche se occorre precisare che il divieto sussiste solo laddove dall'u-

[29] Cfr. per un primo commento della disciplina, nelle more dell'iter legislativo, B. Marchetti – C. Casonato, *Prime osservazioni sulla proposta di Regolamento dell'Unione europea in materia di intelligenza artificiale*, in *BioLaw Journal*, 3, 2021, 415-437; G. Finocchiaro, *The regulation of artificial intelligence*, in *AI & Society*, 3 aprile 2023.

[30] V. AIA, Capo III, spec. sez. 2 (artt. 8-15) e sez. 3 (artt. 16-27).

[31] AIA, art. 3, n. 60.

[32] AIA, art. 5, par. 1, lett. a).

[33] AIA, art. 5, par. 1, lett. b).

tilizzo di questi sistemi derivino, o possano derivare, "danni significativi" a persone[34].
Venendo ai sistemi ad alto rischio, ai sensi dell'art. 6 del regolamento devono essere considerati tali quei sistemi utilizzati in determinati settori e per specifici impieghi, stabiliti tassativamente dall'allegato III[35]. Sebbene in questo elenco non vi sia alcun riferimento esplicito ai *deepfake*, resta comunque possibile che essi possano rientrare tra i settori e gli impieghi ad alto rischio, laddove siano utilizzati nell'ambito di settori "sensibili", quali istruzione e formazione professionale, occupazione lavorativa, servizi essenziali, attività di contrasto ai crimini, processi democratici. A ben vedere, è proprio quest'ultimo il settore che più realisticamente potrebbe essere interessato dai *deepfake*, considerato che l'allegato classifica come ad alto rischio «i sistemi di IA destinati a essere utilizzati per influenzare l'esito di un'elezione o di un referendum o il comportamento di voto delle persone fisiche nell'esercizio del loro voto alle elezioni o ai referendum»[36].
Tuttavia, lo stesso art. 6, al par. 3, sembra mitigare gli effetti di questa classificazione, poiché esclude che un sistema, pur indicato nell'allegato III, debba considerarsi ad alto rischio qualora non presenti «un rischio significativo di danno per la salute, la sicurezza o diritti fondamentali delle persone fisiche, anche nel senso di non influenzare materialmente il risultato del processo decisionale»[37]. Pertanto, se anche i *deepfake* rientrassero tra i sistemi elencati nell'allegato, comunque potrebbero essere "esentati" dalla stringente disciplina riservata ai sistemi ad alto rischio nel caso in cui si riuscisse a provare l'assenza di un rischio significativo.
D'altronde, tale disciplina si applicherebbe ai soli *deepfake* utilizzati in settori "sensibili": questa scelta appare certamente ragionevole poiché motivata sulla base della maggiore astratta pericolosità. Tuttavia, i *deepfake* utilizzati in questi settori potrebbero essere rivolti al raggiungimento di scopi "meritevoli" (ad es., il contrasto ai crimini), ben più di altri *deepfake*, che, per il sol fatto di essere utilizzati in settori non sensibili e a prescindere dalla "meritevolezza" del loro impiego, sarebbero liberi dai vincoli imposti per i sistemi ad alto rischio.
Comunque la si veda, da questa analisi emerge un assunto chiaro: per il regolatore unionale, i *deepfake* non costituiscono di per sé sistemi ad alto rischio, a meno che non riguardino particolari impieghi in settori sensibili.
Tuttavia, il disvalore associato a questo genere di contenuti ha indotto lo stesso legislatore a ritagliare per essi uno spazio specifico, in modo che possano essere distinti dai semplici sistemi a basso rischio (*rectius*: a minimo rischio), per i quali il regolamento

[34] V. art. 5, par. 1, lett. a) e b), identici nella parte finale della disposizione: «[…] in un modo che provochi o possa ragionevolmente provocare a tale persona, a un'altra persona o a un gruppo di persone un danno significativo».

[35] L'allegato III indica otto settori, nell'ambito dei quali l'utilizzo di sistemi di IA deve essere considerato ad alto rischio: biometria, infrastrutture critiche, istruzione e formazione professionale, occupazione e gestione dei lavoratori, servizi essenziali, contrasto all'illegalità, migrazione e asilo, giustizia e processi democratici. Per ognuno di questi settori sono segnalati specifici impieghi di IA considerati ad alto rischio.

[36] AIA, allegato III, n. 8, lett. b).

[37] AIA, art. 6, par. 3, primo comma. Tuttavia, come è specificato successivamente, questa disposizione si applica solo qualora sia soddisfatta almeno una di determinate condizioni: v. art. 6, par. 3, secondo comma.

non impone particolari obblighi. Proprio con riferimento ai *deepfake* (e a pochi altri sistemi di IA[38]), l'AIA immagina una categoria intermedia tra i sistemi ad alto e i sistemi a minimo rischio. L'art. 50, infatti, stabilisce specifici obblighi di trasparenza per i fornitori e i *deployer* di «determinati sistemi di IA»: in particolare, i *deployer*[39] di un sistema di IA che genera o manipola immagini o contenuti audio o video che costituiscono un «deep fake» sono obbligati a rendere noto che il contenuto è stato generato o manipolato artificialmente[40]. Tale obbligo subisce comunque delle mitigazioni: intanto, non si applica, come logico, se l'utilizzo è autorizzato dalla legge per il contrasto di reati[41]; inoltre, nel caso in cui il contenuto faccia parte di un'opera o un programma «manifestamente artistici, creativi, satirici o fittizi», allora è sufficiente rivelare l'esistenza dei contenuti generati o manipolati «in modo adeguato», senza che risultino ostacolati «l'esposizione o il godimento» dell'opera[42]. La lettera di questa disposizione lascia qualche dubbio interpretativo. Innanzitutto, non si comprende in cosa differirebbe l'obbligo in questione rispetto a quello previsto per tutti i *deepfake*: in particolare, quando la rivelazione (*disclosure*) potrebbe dirsi effettuata «in modo adeguato»? Sembra che la natura del *deepfake* debba essere rivelata senza che possa essere preclusa «l'esposizione o il godimento» dell'opera: quindi sarebbe possibile rimandare la *disclosure* ad un momento successivo? Inoltre, occorre delimitare il perimetro delle opere e dei programmi «manifestamente artistici, creativi, satirici o fittizi»: alla luce di alcuni di questi aggettivi non sembra potersi escludere del tutto che all'interno di questo perimetro possano essere ricompresi anche contenuti pubblicati sui *social* da semplici utenti, in quanto "protetti"

[38] L'art. 50 AIA stabilisce obblighi per i fornitori di «sistemi di IA destinati a interagire direttamente con le persone fisiche» (par. 1) e per i *deployer* di sistemi «di riconoscimento delle emozioni o di […] categorizzazione biometrica» (par. 3).

[39] Mentre la versione in italiano della proposta del 2021 imponeva l'obbligo agli «utenti» (art. 52, par. 3, Proposta di regolamento COM(2021) 206 final, 21 aprile 2021), il regolamento approvato preferisce fare riferimento ai «*deployer*», scegliendo in modo singolare di mantenere il termine inglese anche nella versione italiana, probabilmente considerandolo di problematica traduzione. Stando all'art. 3, n. 4, il *deployer* deve essere inteso come «una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che utilizza un sistema di IA sotto la propria autorità, tranne nel caso in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale».

[40] AIA, art. 50, par. 4. Con riferimento agli obblighi per i fornitori, il par. 2 dispone: «I fornitori di sistemi di IA, compresi i sistemi di IA per finalità generali, che generano contenuti audio, immagine, video o testuali sintetici, garantiscono che gli output del sistema di IA siano marcati in un formato leggibile meccanicamente e rilevabili come generati o manipolati artificialmente. I fornitori garantiscono che le loro soluzioni tecniche siano efficaci, interoperabili, solide e affidabili nella misura in cui ciò sia tecnicamente possibile, tenendo conto delle specificità e dei limiti dei vari tipi di contenuti, dei costi di attuazione e dello stato dell'arte generalmente riconosciuto, come eventualmente indicato nelle pertinenti norme tecniche».

[41] Tuttavia, in questo caso, i *deepfake* potrebbero essere assoggettati alla disciplina prevista per i sistemi ad alto rischio alle condizioni stabilite dall'art. 6 e dall'allegato III, n. 4.

[42] Si riporta integralmente il contenuto dell'art. 50, par. 4, AIA: «I deployer di un sistema di IA che genera o manipola immagini o contenuti audio o video che costituiscono un «deep fake» rendono noto che il contenuto è stato generato o manipolato artificialmente. Tale obbligo non si applica se l'uso è autorizzato dalla legge per accertare, prevenire, indagare o perseguire reati. Qualora il contenuto faccia parte di un'analoga opera o di un programma manifestamente artistici, creativi, satirici o fittizi, gli obblighi di trasparenza di cui al presente paragrafo si limitano all'obbligo di rivelare l'esistenza di tali contenuti generati o manipolati in modo adeguato, senza ostacolare l'esposizione o il godimento dell'opera».

dalla loro natura "creativa, satirica o fittizia"[43]. In particolare, il riferimento alle opere "manifestamente fittizie" rischia di ampliare enormemente – e forse eccessivamente – il campo dell'eccezione. Infine, pur tacendo i dubbi che attengono alle modalità di divulgazione dei contenuti, il regolamento non prevede sanzioni chiare in caso di mancata osservanza degli obblighi di trasparenza[44].

Le Istituzioni europee, rompendo gli indugi rispetto a GDPR e DSA, hanno più convintamente avviato il percorso di regolamentazione dei *deepfake*, considerandoli nell'AIA come tecnologie (*rectius*: prodotti di tecnologie) di IA che possono essere classificate sulla base del rischio. Tuttavia, il fatto di aver inserito nel regolamento una disciplina specifica (art. 50) rivela la difficoltà di inquadrare questo genere di contenuti secondo i criteri di rischio previsti per l'IA in generale; in aggiunta, tale disciplina non sembra brillare per chiarezza e completezza. Siamo probabilmente agli albori di una regolamentazione che potrebbe necessitare di nuovi approfondimenti già nel prossimo futuro.

3. Stati Uniti: interventi (statali) per ambiti di utilizzo

Come noto, l'approccio statunitense in materia di regolamentazione dell'IA è profondamente diverso rispetto a quello europeo, soprattutto per il fatto che non pare registrarsi – al netto di alcune proposte di legge presentate al Congresso – la volontà di intervenire con atti di *hard law* operanti a livello federale[45]. Questa scelta ha prodotto l'emersione di normative di livello statale, che sempre più tentano di colmare il vuoto regolatorio con riferimento agli utilizzi di alcune nuove tecnologie[46]. Come sta accadendo, ad esempio, per le tecnologie di riconoscimento facciale[47], anche per i *deepfake*, in assenza di un intervento federale, alcuni Stati sono intervenuti per disciplinare il fenomeno. Prima di dare conto di queste normative, sembra opportuno soffermarsi su alcune recenti proposte di legge in materia presentate a livello federale.

Nel settembre 2023, al fine di proteggere la sicurezza nazionale e garantire mezzi di tutela alle vittime, è stato presentato alla Camera dei rappresentanti il *Deepfakes Accountability Act*[48], che stabilirebbe per i creatori di *deepfake* alcuni obblighi comuni e altri spe-

[43] Vero è che l'avverbio «manifestamente» potrebbe servire a limitare l'ampiezza dell'eccezione.

[44] Cfr. per un commento sulla disciplina dei *deepfake* contenuta nell'AIA, già dai tempi della proposta: A. Fernandez, *"Deep fakes": disentangling terms in the proposed EU Artificial Intelligence Act*, in *UFITA Archiv für Medienrecht und Medienwissenschaft*, 2, 2022, 392-433; M. Łabuz, *Regulating deep fakes in the artificial intelligence act*, in *Applied Cybersecurity & Internet Governance*, 1, 2023, 1-42; M. Łabuz, *Deep fakes and the Artificial Intelligence Act – An important signal or a missed opportunity?*, in *Policy & Internet*, 2024, 1-18.

[45] Cfr. B. Marchetti – L. Parona, *La regolazione dell'intelligenza artificiale: Stati Uniti e Unione europea alla ricerca di un possibile equilibrio*, in *DPCE online*, 1, 2022, 237 ss.; E. Chiti – B. Marchetti, *Divergenti? Le strategie di Unione europea e Stati Uniti in materia di intelligenza artificiale*, in *Riv. reg. merc.*, 1, 2020.

[46] S. Parinandi – J. Crosson – K. Peterson – S. Nadarevic, *Investigating the politics and content of US State artificial intelligence legislation*, in *Business and Politics*, 2, 2024, 240-262.

[47] Cfr. J. Spivack – C. Garvie, *A taxonomy of legislative approaches to face recognition in the United States*, in A. Kak (a cura di), *Regulating biometrics: Global approaches and urgent questions*, New York, 2020, 86-95.

[48] *Deepfakes Accountability Act*, H.R. 5586, 20 settembre 2023. Allo stato, questa proposta risulta ferma al momento dell'introduzione, non essendo stata discussa dal Congresso.

cifici a seconda del contenuto (audio/video, solo video o solo audio): in linea generale, si imporrebbe l'inserimento di tecnologie, come quelle di provenienza dei contenuti (*content provenance technologies*), sufficienti a identificare chiaramente il contenuto come composto di elementi alterati o come interamente creato attraverso IA generativa o simili[49]. In particolare, a seconda della natura del contenuto, si pretenderebbero dichiarazioni verbali, scritte o icone da integrare nel *deepfake* atte a prevenire qualsiasi fraintendimento[50]. Alla violazione di questi obblighi sarebbe associata una sanzione differente a seconda della offensività del contenuto, stabilendo la pena della reclusione fino a cinque anni nei casi in cui il *deepfake* sia volto ad arrecare molestie attraverso contenuti sessuali, a interferire in un procedimento ufficiale (comprese le elezioni) purché si tratti di minaccia credibile, a porre in essere frodi o furti di identità, a influenzare un dibattito pubblico interno nell'interesse di una potenza straniera[51]. Le eccezioni risulterebbero abbastanza circoscritte: oltre a consentire contenuti utilizzati dalle forze dell'ordine per la tutela della pubblica sicurezza, verrebbero esentati dalla disciplina i contenuti pubblicati in un contesto tale che una persona ragionevole non potrebbe confondere l'attività falsificata con l'attività effettiva della persona esposta, come nel caso di parodie, rievocazioni storiche o programmi radiofonici, televisivi o cinematografici manifestamente fittizi[52]. Interessanti anche gli strumenti di prevenzione prospettati: Intanto, l'istituzione di una *task force* all'interno del Dipartimento di sicurezza nazionale[53]; inoltre, l'obbligo imposto agli sviluppatori di tecnologie utili alla creazione di *deepfake* di garantire la capacità tecnica del prodotto di inserire la provenienza dei contenuti, unitamente al dovere di rivolgere all'utente una informativa sugli obblighi attinenti alla creazione di *deepfake*[54]; infine, la pretesa che i fornitori di piattaforme online garantiscano non solo la capacità tecnica per l'indicazione della provenienza del contenuto, ma che si dotino di un sistema per il rilevamento dei *deepfake*[55].

All'inizio del 2024, è stato presentato un altro progetto di legge noto come *DEFIANCE Act*[56], approvato dal Senato nel mese di luglio con emendamenti, ai sensi del quale sarebbe garantita la tutela in via giudiziaria alle vittime di *deepfake* "intimi" diffusi senza il loro consenso: si tratta di estendere anche al caso dei *deepfake* la tutela già prevista per la divulgazione non consensuale di *intimate images*[57].

Recentissima, infine, la proposta nota come *COPIED Act* (*Content Origin Protection and*

[49] Ivi, sec. 2(b).

[50] Ivi, sec. 2(c)(d)(e). Singolare la disposizione che per i deepfake audio di durata superiore a due minuti obbligherebbe i creatori a inserire una dichiarazione verbale per chiarire la natura del contenuto ad intervalli di due minuti.

[51] Ivi, sec. 2(f)(1). Per le sanzioni civili e i mezzi a tutela delle vittime, v. ivi, sec. 2, §1041(f)(2)(g)(h).

[52] Ivi, sec. 2(j).

[53] Ivi, sec. 7(a).

[54] Ivi, sec. 10(a).

[55] Ivi, sec. 10(b).

[56] *Disrupt Explicit Forged Images and Non-Consensual Edits Act of 2024* (*DEFIANCE Act*), S. 3696, 30 gennaio 2024.

[57] Cfr. 15 USC § 6851.

Integrity from Edited and Deepfaked Media Act)[58], volta a semplificare l'identificazione dei contenuti generati con l'IA e la tutela del diritto d'autore. Il progetto, abbastanza ambizioso, riporta la questione sul piano generale dei sistemi di IA e richiede la presenza di elementi distintivi che consentano di identificare il lavoro dell'IA da quello di un essere umano in modo rapido e standardizzato, ad esempio attraverso l'apposizione obbligatoria di filigrane digitali inerenti all'origine e all'autenticità del contenuto. A protezione del diritto d'autore, la legge vieterebbe l'addestramento di modelli di IA attraverso materiali protetti da copyright senza il consenso dei titolari dei diritti, introducendo sanzioni per le violazioni. L'iter di questo progetto merita di essere osservato con attenzione in riferimento al nostro tema, dato che tali disposizioni, pensate per tutti i contenuti IA, riguarderebbero in primo luogo proprio i *deepfake*.

Se sul piano federale si resta nel campo delle proposte di legge, come accennato, alcuni Stati si sono già dotati di una disciplina in materia. Da questo punto di vista, si possono rintracciare due chiare tendenze: alcuni interventi normativi, infatti, si focalizzano sui *deepfake* pornografici, mentre altri prendono in esame la questione dei *deepfake* politici.

La California ha guidato la carica, approvando nel 2019 due leggi in materia, dedicate appunto alle due applicazioni specifiche anzidette, ossia l'influenza delle campagne politiche[59] e l'utilizzo in ambito pornografico[60]. Occorre evidenziare che la legge sui *deepfake* politici era, in realtà, sottoposta a una *sunset clause* che ne prevedeva l'automatica abrogazione al 1° gennaio 2023, salvo ulteriori determinazioni del legislatore, che non sono però intervenute. Questa legge vietava, nei sessanta giorni precedenti una elezione, la distribuzione di "supporti audio o visivi materialmente ingannevoli", avvenuta con effettiva malizia (*actual malice*) e con l'intento di danneggiare la reputazione di un candidato o di ingannare gli elettori[61]. Le sanzioni non trovavano comunque applicazione nel caso di contenuti satirici o parodistici né per gli operatori dell'informazione che, anche attraverso internet, diffondevano i *deepfake* affermandone però con chiarezza la natura ingannevole[62]. Come detto, tale legge non risulta più in vigore, mentre resta pienamente applicabile la legge sui *deepfake* pornografici, che prevede il diritto di agire in giudizio contro chi distribuisce intenzionalmente *deepfake* di foto o video aventi natura intima o sessuale senza il consenso della persona ritratta[63]. Le eccezioni, pur previste, sembrano abbastanza difficili da realizzarsi, dato che dovrebbe dimostrarsi l'interesse pubblico alla divulgazione, oppure il valore politico o giornalistico del contenuto o, in

[58] La presentazione del *Content Origin Protection and Integrity from Edited and Deepfaked Media Act of 2024 (COPIED Act)* è stata annunciata in data 11 luglio 2024 con un comunicato stampa reperibile su *commerce.senate.gov*.

[59] California, AB no. 730, chap. 493, 10 aprile 2019.

[60] California, AB no. 602, chap. 491, 3 ottobre 2019.

[61] California, AB 730, cit., sec. 4(a). In particolare, per «materially deceptive audio or visual media» si intendeva un'immagine o registrazione audio o video dell'aspetto, del discorso o della condotta di un candidato intenzionalmente manipolata in modo tale che fossero soddisfatte entrambe le seguenti condizioni: il contenuto doveva apparire falsamente autentico a una persona ragionevole; poteva indurre una persona ragionevole ad avere una comprensione o un'impressione sostanzialmente diversa rispetto a quella che avrebbe avuto ascoltando o vedendo la versione inalterata.

[62] California, AB 730, cit., sec. 4(d)(4)(5).

[63] California, AB 602, cit., sec. 1(b). Come altri legislatori statali, anche quello californiano preferisce utilizzare il termine "*digitization*" rispetto a "*deepfake*" nell'ambito della pornografia non consensuale.

generale, la protezione costituzionale dell'attività realizzata[64].

Altri Stati stanno hanno seguito l'esempio californiano, nel senso di una o dell'altra direzione regolatoria. Con riferimento ai *deepfake* pornografici, nel 2019 in Virginia è stata approvata una legge che, nel tentativo di prevenire il c.d. *revenge porn*, sanziona penalmente la distribuzione di *deepfake* pornografici se idonei a costringere, molestare o intimidire una persona[65]. In Florida, dal 2022 le sanzioni penali relative alla pedopornografia e alla pornografia non consensuale sono state estese, fatti esplicitamente salvi gli internet *provider*, a "chi promuove" («*A person who […] promotes*») *deepfake*, a nulla valendo eventuali filigrane o etichettatura del contenuto: sebbene il termine *deepfake* non sia utilizzato esplicitamente, è chiaro che contenuti di tal genere risultino inclusi nella amplissima definizione di "immagini create, alterate, adattate o modificate con mezzi elettronici, meccanici o di altro tipo"[66]. In senso simile hanno provveduto anche Louisiana[67], South Dakota[68] e Washington[69].

In relazione ai *deepfake* politici, il Texas ha approvato nel 2019 una legge, simile alla AB730 della California, che impedisce la distribuzione di *deepfake* politici entro trenta giorni dalle elezioni[70]. Dal 2024 in Mississippi è previsto il reato di diffusione di "*digitization*" (termine preferito dal legislatore rispetto al più comune deepfake) nei novanta giorni precedenti le elezioni, nel caso in cui, mancando il consenso della persona ritratta, chi diffonde il materiale ne conosce la natura e mira a influenzare il dibattito elettorale[71]. In New Mexico, nel 2024, è stata introdotta una legislazione, molto simile alla proposta federale nota come *Deepfakes Accountability Act*, che impone di etichettare i contenuti in modo da renderne chiara la natura, oltre a sanzionare penalmente la condotta di chi diffonde o si accorda con altri per diffondere "*materially deceptive media*"[72]. Simili discipline, con riferimento all'obbligo di etichettatura dei contenuti, sono state approvate anche in Indiana[73] e in Oregon[74].

A margine della bipartizione appena analizzata, merita di essere segnalato l'*Ensuring Likeness, Voice and Image Security (ELVIS) Act*, recentemente approvato in Tennessee, che aggiorna e sostituisce il *Personal Rights Protection Act* del 1984, prevedendo una sanzione civile a carico di chi rende disponibile al pubblico "*voice or likeness*" senza autorizzazione del titolare del diritto[75]. Sebbene concepito per la tutela della proprietà intellettuale, tale provvedimento impatta sulla legittimità della creazione di *deepfake*. Proprio per questo sono previste importanti eccezioni alla disciplina per utilizzi, protetti dal

64 California, AB 602, cit., sec. 1(c)(1).

65 Virginia, H.B. 2678, 18 marzo 2019.

66 Florida, S.B. 1798, 24 giugno 2022.

67 Louisiana, S.B. 175 (Act 457), 28 giugno 2023.

68 South Dakota, S.B. 79, 13 febbraio 2024.

69 Washington, S.B. 1999, 6 giugno 2024.

70 Texas, S.B. 751, 25 maggio 2019.

71 Mississippi, S.B. 2577, 30 aprile 2024.

72 Nex Mexico, H.B. 182, 5 marzo 2024.

73 Indiana, H.B. 1133, 12 marzo 2024.

74 Oregon, S,B, 1571, 27 marzo 2024.

75 Tennessee, H.B. 2091, 26 marzo 2024 (*ELVIS Act*).

Primo emendamento, connessi a fini giornalistici, informativi, educativi, satirici e parodistici[76].

L'analisi del quadro normativo federale e statale negli Stati Uniti lascia la sensazione di una tendenza a regolamentare solo debolmente i *deepfake*. Per quanto debba essere notato che il dibattito e gli interventi normativi siano in clamorosa crescita negli ultimi anni, comunque l'attenzione è circoscritta a contesti specifici, come i procedimenti elettorali o la pornografia. Questo orientamento risente probabilmente della notoria rilevanza che nel diritto costituzionale statunitense è riconosciuta al Primo emendamento, del quale i *deepfake* costituirebbero in linea generale una forma di espressione. Infatti, alla luce del caso *Alvarez*, deciso nel 2012 dalla Corte Suprema[77], anche le dichiarazioni false possono ricevere la protezione del Primo emendamento, a meno che non provochino danni gravi legalmente riconoscibili: in questo senso si spiega l'attenzione dei legislatori statali per le interferenze elettorali o la pornografia non consensuale. Resta pertanto controverso un intervento normativo che limiti il ricorso a questi contenuti, a meno che essi non mettano seriamente in pericolo beni di innegabile rilevanza, come la sicurezza pubblica, la dignità umana, il pudore sessuale e la democraticità delle elezioni.

In aggiunta, il diritto statunitense, più marcatamente rispetto al diritto dell'UE e a quello cinese, si caratterizza per l'alto grado di protezione che assicura agli *internet service provider*, ai motori di ricerca e alle piattaforme online, attraverso cui circolano comunemente i *deepfake*: questi soggetti sono infatti generalmente esentati da responsabilità per i contenuti presenti sui loro siti e app. Tale orientamento sembra confermato dalle leggi statali sui *deepfake*, mentre a livello federale occorre evidenziare il differente atteggiamento del *Deepfake Accountability Act* – comunque fermo al momento della presentazione della proposta –, che quantomeno immagina determinati obblighi per questi soggetti. Sul punto era attesa la pronuncia della Corte suprema nel recente caso *Gonzalez v. Google*[78], che avrebbe potuto rivedere l'orientamento sulla irresponsabilità dei *provider*: invece, nonostante le grandi aspettative, i giudici hanno concisamente ribadito che questi soggetti non possono rispondere come responsabili per i contenuti offensivi pubblicati sui propri siti, manifestando una sorta di disagio istituzionale per essere stati chiamati a pronunciarsi su un tema che richiederebbe l'intervento del Congresso.

4. Cina: *deepfake, deep synthesis* e *"deep control".*

Come noto, il governo cinese ha sviluppato una visione strategica mirata a consolidare la propria "sovranità informatica", esercitando un controllo rigoroso sul cyberspazio

[76] *Elvis Act*, cit., sec. 10(a).

[77] Corte Suprema USA, *United States v. Alvarez*, 567 U.S. 709 (2012).

[78] Corte Suprema USA, *Gonzalez v. Google LLC*, 598 U.S. 617 (2023). A seguito dell'attentato terroristico avvenuto al Bataclan di Parigi nel 2015, i parenti di una vittima statunitense, Nohemi Gonzalez, intentavano causa contro *Google* sostenendone la responsabilità indiretta, in quanto l'organizzazione terroristica ISIS aveva potuto diffondere i propri pericolosi messaggi sulla piattaforma *Youtube*, gestita da *Google*.

e sulle tecnologie emergenti, in particolare l'intelligenza artificiale (IA)[79]. Questo approccio si manifesta attraverso l'implementazione di normative che regolano l'uso e lo sviluppo dell'IA, con l'obiettivo di garantire che tali tecnologie siano allineate con gli interessi nazionali e i valori sociali cinesi. Così si giustificano anche i noti controlli governativi sulle aziende tecnologiche, atti ad evitare, tra le altre cose, che possano emergere figure imprenditoriali con potenziale influenza politica[80].

La determinazione della Cina nel consolidare la propria sovranità informatica, perseguendo una politica che integra sviluppo economico, sicurezza nazionale e influenza geopolitica nel contesto digitale globale, non risparmia neanche il fenomeno dei *deepfake*, a cui, anzi, il regolatore pubblico ha rivolto presto la propria attenzione.

Anche in Cina, come in California, il fermento intorno alla regolamentazione dei *deepfake* ha avuto inizio nel 2019, probabilmente a causa dell'aumento della popolarità delle app – in particolare, *ZAO* – che consentivano la creazione di questi contenuti, soprattutto attraverso lo "scambio di volti", e suscitavano polemiche per i profili relativi alla raccolta dei dati[81]. Le autorità cinesi, non soddisfatte dell'adeguamento delle proprie politiche privacy portato avanti dalla stessa ZAO, hanno presto cominciato a ragionare sulla possibilità di regolamentare, se non vietare drasticamente, i *deepfake*. Così, entro tre mesi dal lancio di ZAO, la *Cyberspace Administration of China* pubblicava diversi documenti in cui si discuteva della necessità di regolamentare l'IA e gestirne lo sviluppo[82]. Entro la fine dell'anno, con entrata in vigore a partire dal 2020, sono stati approvati i "*Regulations on the Administration of Online Audio and Video Information Services*", che hanno stabilito nuove regole e responsabilità, anche sul piano penale, sia per i fornitori che per gli utenti, introducendo di fatto un divieto ampio e generalizzato con riguardo all'uso di immagini, audio e video generati attraverso tecnologie di sintesi profonda per creare o diffondere notizie false[83]. Tale scelta è stata giustificata sulla base della considerazione che l'utilizzo di nuove tecnologie può turbare l'ordine sociale e violare gli interessi delle persone, generando rischi politici e un impatto negativo sulla sicurezza nazionale e la stabilità sociale. Queste norme si rivolgono in larga parte alle piattaforme online, chiamate a rafforzare l'autoregolamentazione nel settore, a istituire un sistema di responsabilità editoriale, a garantire la sicurezza informatica, a verificare la reale identità degli utenti, a segnalare adeguatamente contenuti non reali, a interrompere la circolazione dei contenuti non consentiti e, in generale, ad accettare consapevolmente il controllo sociale.

Si tratta di atti normativi in perfetta sintonia con la direzione intrapresa dal governo cinese, completamente opposta a quella statunitense, di esercitare il proprio controllo

---

[79] Cfr. G. Santoni, *La Cina e lo spazio digitale. Questioni di governance nello spazio digitale globale*, in *OrizzonteCina*, 3, 2020, 70-75; M. Kolton, *Interpreting China's pursuit of cyber sovereignty and its views on cyber deterrence*, in *The Cyber Defense Review*, 2, 2017, 1 ss.; H. Gu, *Data, big tech, and the new concept of sovereignty*, in *Journal of Chinese political science*, 2023, 591-612; L. Formichella, *La disciplina normativa sulla protezione dei dati e delle informazioni personali in Cina: emersione di un nuovo paradigma nel contesto internazionale*, in *Mondo Cinese*, 3, 2021, 85-94.

[80] Cfr. A.H. Zhang, *High Wire: How China Regulates Big Tech and Governs Its Economy*, Oxford, 2024.

[81] Cfr. Geng, *Comparing "Deepfake" Regulatory Regimes*, cit., 167 ss.

[82] *Ibid.*

[83] *Administrative Regulations on Online Audio and Video Information Services*, 18 novembre 2019. Cfr. *China issues regulation for online audio, video services*, in *english.www.gov.cn* 30 novembre 2019.

sulle nuove tecnologie e sui loro effetti attraverso interventi statali che richiedono una "cooperazione", spesso sotto forma di veri e propri obblighi, alle grandi imprese operanti nel settore[84].

Nel pieno perseguimento di questa politica, nel corso del 2022, a seguito di una procedura di consultazione pubblica e prevedendo l'entrata in vigore per l'anno successivo, la stessa *Cyberspace Administration of China* ha introdotto un'altra serie di norme, note come *Regulations on Deep Synthesis Management of Internet Information Services*, volta a regolamentare le attività di sintesi profonda su Internet in un'ottica di promozione dei "valori socialisti fondamentali", oltre che di protezione della sicurezza nazionale e dei diritti dei cittadini[85]. Queste norme, quindi, riguardano più specificamente i *deepfake*, anche se la definizione utilizzata – "tecnologie di sintesi profonda" – sembra riferirsi a un più ampio spettro di tecnologie, tra cui potrebbero rientrare, ad esempio, ambienti virtuali e *chatbot*. Di fatto, al fine di operare un controllo più incisivo, l'attenzione è posta non sui risultati dell'utilizzo della tecnologia (quali sono i *deepfake*), ma sulla tecnologia stessa. Vero è che il regolamento non stabilisce un divieto generalizzato alla creazione di *deepfake*: tuttavia, i confini dei contenuti non consentiti sembrano particolarmente vaghi. Da un lato, i servizi di sintesi profonda non devono essere utilizzati da alcuna organizzazione o individuo per produrre, riprodurre, pubblicare o trasmettere informazioni proibite da leggi o regolamenti amministrativi, o per intraprendere attività proibite da leggi e regolamenti amministrativi, come quelle che mettono a repentaglio la sicurezza e gli interessi nazionali, danneggiano l'immagine della nazione, danneggiano l'interesse pubblico della società, disturbano l'ordine economico o sociale o danneggiano i diritti e gli interessi legittimi di altri[86]. Dall'altro, come rinforzo al divieto, è precisato che i fornitori e gli utenti di servizi di sintesi profonda non devono utilizzare servizi di sintesi profonda per produrre, riprodurre, pubblicare o trasmettere informazioni di notizie false[87]. Se già i contenuti della prima disposizione lasciano aperti dubbi interpretativi – ad esempio, quali contenuti, stando ai valori promossi dal governo cinese, potrebbero dirsi dannosi per l'immagine della nazione o l'ordine economico e sociale? –, la seconda precisazione sembra bandire qualsiasi contenuto "falso": in questo senso, astrattamente qualsiasi *deepfake* potrebbe essere considerato come tale.

In piena coerenza con l'orientamento politico sopra descritto, anche la normativa sulla sintesi profonda è incentrata sugli obblighi imposti ai *provider*, i quali sono chiamati ancora una volta, salvo incorrere in sanzioni sul piano sia civile che penale, a verificare l'identità degli utenti, a etichettare adeguatamente i contenuti, a rafforzare la gestione dei

---

[84] Cfr. E. Hine – L. Floridi, *New deepfake regulations in China are a tool for social stability, but at what cost?*, in Nature Machine Intelligence, 4, 2022, 608-610. Come spiegato dagli autori, si tratta di una tendenza consolidata, poiché il governo fa sempre più affidamento sulle aziende tecnologiche per applicare nuove normative su Internet per facilitare la visione del Partito Comunista Cinese (PCC) di una società stabile e prospera, prevedendo conseguenze in caso di mancata collaborazione. A riprova di ciò, ripercorrono la vicenda del presidente esecutivo della piattaforma *Alibaba*, Jack Ma, addirittura scomparso per mesi agli occhi dell'opinione pubblica a seguito di dissidi con il governo.

[85] *Regulations on Deep Synthesis Management of Internet Information Services*, 25 novembre 2022 (per la traduzione in lingua inglese, cfr. *chinalawtranslate.com*).

[86] Ivi, art. 6.

[87] *Ibid.*

dati e, soprattutto, a segnalare agli utenti l'obbligo di ottenere il consenso da parte delle persone interessate prima della produzione del contenuto di sintesi profonda (con specifico riferimento alla alterazione di volti e voci)[88]. Si tratta di disposizioni particolarmente gravose per gli operatori, che, a conti fatti, potrebbero dimostrarsi comunque insufficienti. In effetti, fermo restando che non risultano chiare le modalità con cui dovrebbero essere segnalati o etichettati i contenuti, i fornitori dovrebbero provvedere in tal senso per qualsiasi prodotto di sintesi profonda, a prescindere dallo scopo e dalla pericolosità più o meno elevata. Inoltre, le più comuni modalità di etichettatura – come l'apposizione di filigrane – potrebbero comunque essere rimosse mediante l'utilizzo di altre tecnologie, magari anche queste basate sulla *deep synthesis*, con frustrazione di tutti gli sforzi effettuati dai gestori delle piattaforme, che rimarrebbero ancora esposti alle responsabilità conseguenti alla circolazione dei contenuti. Infine, l'obbligo di interruzione immediata della trasmissione di contenuti non etichettati sembra particolarmente velleitario se consideriamo che, una volta creato, il contenuto può essere facilmente disaccoppiato dal servizio su cui è stato creato e diffuso indipendentemente da esso, oltre alla possibilità che lo stesso contenuto venga ricaricato o catturato tramite *screenshot*, quindi sottratto dal controllo dell'autore e del fornitore di servizi originario.

5. Tre approcci rivelatori di tendenze, visioni e obiettivi

Nell'affrontare le sfide della particolare categoria di output generati dall'IA noti come *deepfake*, Unione europea, Stati Uniti e Cina hanno adottati approcci diversi, che, come detto in apertura, ricalcano in buona parte le scelte finora intraprese in materia di regolamentazione generale dell'IA. Gli approcci si distinguono per quanto riguarda le regole proposte, la ripartizione delle responsabilità tra i soggetti coinvolti e le conseguenze derivanti dalle violazioni.

In via preliminare, occorre osservare un dato comune: in nessun ordinamento è operativo un divieto generalizzato dei *deepfake*. Nonostante la recente normativa cinese in materia, specialmente se letta in combinato disposto con i precedenti regolamenti relativi ai servizi online, presenti contorni particolarmente sfumati che rendono di fatto a rischio illegalità qualsiasi contenuto *deepfake*, resta almeno astrattamente accettata l'impraticabilità di un divieto generalizzato, che costituirebbe una illegittima compressione della libertà di espressione[89]. All'opposto, proprio facendo leva sul Primo emendamento, negli Stati Uniti le limitazioni ai *deepfake* continuano a latitare a livello federale e restano abbastanza circoscritte a livello statale. Al di là della indubbia valenza dei principi, la regolamentazione più o meno stringente non può che incidere anche sulla libertà di impresa degli operatori privati del settore tecnologico, i quali contribuiscono alla creazione e alla diffusione dei *deepfake*: da questo punto di vista, mentre nel sistema cinese il controllo sui privati è connaturato all'organizzazione socioeconomica del pae-

[88] Ivi, artt. 14-18.

[89] Sul tema della libertà d'espressione nel mondo digitale, cfr. P. Tanzarella, *La trasformazione della libertà d'espressione dal mondo liberale al mondo digitale. Quale futuro per i principi classici del costituzionalismo?*, in V. Faggiani – G.B. Sales Sarlet (a cura di), *Retos del derecho ante la IA*, Barcellona, 2024, 103-137.

se, nei sistemi europeo e statunitense restrizioni eccessive, magari unite all'attribuzione di obblighi e responsabilità, rischiano di disincentivare gli investimenti nel settore, con tutto ciò che ne consegue anche in termini di posizione economica dell'UE e degli USA in relazione allo sviluppo dell'IA.

Venendo a individuare alcuni criteri classificatori delle discipline analizzate nei paragrafi precedenti, sembra innanzitutto opportuno soffermarsi sulla definizione della categoria. Sul punto non sembrano sussistere particolari divergenze tra i vari sistemi, atteso che il *deepfake*, accogliendo la definizione dello STOA dell'UE, può essere inteso come media audio o video manipolato o sintetico che sembra autentico e che mostra persone che sembrano dire o fare qualcosa che non hanno mai detto o fatto, prodotto utilizzando tecniche di IA, tra cui l'apprendimento automatico e il *deep learning*. Sebbene non possa considerarsi questa definizione come standardizzata a livello mondiale, comunque le caratteristiche di questo genere di contenuti rimangono identificate in maniera pressoché analoga. Tutt'al più, può notarsi come difficilmente si opti per una regolamentazione circoscritta al fenomeno dei *deepfake* specificamente intesi; soprattutto, appare singolare il fatto che le normative non facciano così frequentemente riferimento al termine "*deepfake*". Nell'UE, la disciplina è integrata all'interno della regolamentazione olistica dell'IA e comunque sembra particolarmente invischiata nell'impianto generale del regolamento, tanto da lasciare dubbi sulla classificazione del grado di rischio attribuito a queste tecnologie; negli USA, ad eccezione della proposta federale nota appunto come *Deepfakes Accountability Act*, la regolamentazione statale sceglie spesso di riferirsi al fenomeno con delle perifrasi o addirittura con altri termini alquanto inconsueti (ad es., *digitization*); in Cina, al centro della regolamentazione sono finiti i contenuti prodotto di *deep synthesis*. Tale ritrosia ad utilizzare il termine "*deepfake*" svela, da un lato, una non completa padronanza sul piano semantico, che evidentemente preoccupa i legislatori; dall'altro, denota pure come sembri ancora oggi maggiormente opportuno regolamentare il fenomeno in un'ottica più ampia, ragionando su discipline che, mentre pensano ai *deepfake*, immaginano quadri regolatori anche per altri utilizzi dell'IA magari ancora non così diffusi. Tale atteggiamento potrebbe "allungare la vita" delle discipline che stanno entrando in vigore, ma potrebbe prestare il fianco ad una regolamentazione dei *deepfake* intesi in senso stretto ancora incerta e poco incisiva.

Un altro criterio classificatorio può essere rintracciato guardando al rapporto tra intervento normativo e fasi specifiche del ciclo di vita dei *deepfake*. Come per ogni sistema di IA, questo può essere scomposto in varie fasi: intanto, esiste il momento della nascita della tecnologia di IA attraverso cui i contenuti saranno prodotti; dopodiché, si può distinguere la fase di creazione del *deepfake* da quella della sua circolazione e ancora da quella della visualizzazione da parte del pubblico. In ognuna di queste fasi si trovano coinvolti soggetti diversi, che possono sovrapporsi e giocare un ruolo in momenti differenti: soprattutto, possiamo distinguere i fornitori dei servizi che consentono la creazione dei contenuti, i fornitori dei servizi che ne permettono la pubblicazione e la trasmissione, i soggetti (sostanzialmente, utenti) che creano i contenuti, quelli che li fanno circolare e quelli che li visualizzano. Detto che una efficace regolamentazione dei *deepfake* dovrebbe tenere in conto tutte le fasi del ciclo di vita del sistema, si possono comunque notare alcune chiare differenze tra gli ordinamenti presi in esame con rife-

rimento all'enfasi riservata ad alcune fasi rispetto ad altre. Nell'UE e in Cina il focus è sulle prime fasi del ciclo di vita (tecnologia utilizzata per la produzione, creazione e circolazione del contenuto), mentre negli Stati Uniti le normative tendono a enfatizzare sulle fasi successive (effetti del contenuto sul pubblico). Ciò non significa che le altre fasi vengano trascurate del tutto: nelle leggi statali statunitensi, per esempio, si ritrovano obblighi di etichettatura dei contenuti attinenti alle prime fasi, ma l'attenzione è posta molto spesso sul piano dei rimedi da accordare ai soggetti danneggiati, cui si cerca di assicurare una tutela legale sul piano riparatorio/risarcitorio che eventualmente si affianchi alle sanzioni penali.

La differente enfasi regolatoria sulle fasi del ciclo di vita dei *deepfake* ha i suoi effetti sulla ripartizione delle responsabilità tra i soggetti coinvolti. Mentre in Cina e nell'UE l'attenzione per le prime fasi pone l'accento sulla responsabilità dei fornitori di servizi, negli Stati Uniti i rimedi garantiti alle vittime finiscono con l'obbligare soprattutto chi diffonde i contenuti dannosi. Anche in questo caso la differenza appare sfumata, dato che sia in Cina che nell'UE le responsabilità si espandono quantomeno ai creatori di contenuti; tuttavia, di fondo resta vero che negli USA la responsabilità dei *provider*, anche alla luce della recente giurisprudenza, è ancora ipotesi poco percorsa, mentre i *creators* continuano a essere protetti, finché possibile, dal Primo emendamento sulla libertà di espressione.

Un secondo criterio classificatorio attiene alla scelta di regolamentare il fenomeno con riferimento agli specifici utilizzi (ed effetti che ne conseguono) o con riferimento al livello di rischio, considerato che una valutazione del secondo tipo ragiona su fattispecie ed effetti maggiormente astratti e potenziali, ma potrebbe garantire una prevenzione più efficace dei danni. D'altronde, non si tratta di orientamenti perfettamente contrapposti: nell'UE, infatti, laddove la scelta è chiaramente votata alla classificazione sulla base del rischio (*risk-based approach*), questa scaturisce comunque dall'individuazione di settori e utilizzi particolarmente sensibili. Il tema è collegato alla scelta di regolamentare il fenomeno singolarmente o come parte di una strategia olistica, magari costruita sull'IA in generale. Da questo punto di vista, appare innegabile che siamo di fronte a tre impostazioni differenti: in Cina, la regolamentazione è abbastanza generalizzata se si osservano le norme di applicazione generale in materia di servizi e media online, ma deve notarsi l'introduzione di una disciplina dedicata nello specifico alle tecnologie di sintesi profonda, insieme comunque più ampio rispetto alla categoria dei *deepfake*; nell'UE, se è vero che la disciplina in materia si ritrova all'interno del regolamento generale sull'IA e che la classificazione dei *deepfake* sulla base del rischio appare ancora incerta e nebulosa, comunque deve essere riconosciuto che un riferimento ad essi fa capolino in una parte a sé dell'AIA (art. 50); negli Stati Uniti, infine, la tendenza abbastanza netta è quella di regolamentare il fenomeno esclusivamente in relazione a due specifiche aree di utilizzo, ossia l'influenza sui procedimenti elettorali e la pornografia non consensuale: in questo modo, la valutazione sulla pericolosità delle tecnologie è compiuta a monte dal legislatore, circoscrivendo chiaramente i settori che necessitano di intervento.

Da ultimo, possiamo classificare i tre approcci nel quadro della visione strategica che le rispettive autorità politiche stanno portando avanti in materia di gestione e sviluppo

delle nuove tecnologie, specialmente dell'IA. Partendo dal presupposto che tutte e tre le potenze puntano a migliorare o consolidare la propria posizione in ordine allo sfruttamento e allo sviluppo dell'IA[90], l'analisi delle prime discipline sui *deepfake* confermano le tendenze che valgono per il settore dell'IA in generale. Per quanto riguarda l'UE, sebbene con maggiore timidezza rispetto ad altre applicazioni di IA, il legislatore prevede obblighi specifici per i *deepfake*, senza che si possa escludere l'applicabilità della più restrittiva disciplina congegnata per i sistemi ad alto rischio: gli operatori privati ne risultano certamente responsabilizzati e lo spazio per l'autoregolamentazione, pur esistente, risulta comunque compresso. Inoltre, la scelta di introdurre regole attraverso uno strumento di *hard law* direttamente applicabile in tutti gli Stati membri rivela la netta preferenza per l'armonizzazione normativa, nel tentativo di offrire un quadro regolatorio chiaro e definito sia ai cittadini sia agli operatori del settore. La visione strategica degli USA, probabilmente forti di una posizione maggiormente dominante nel settore dell'IA, non coincide con quella europea. Come per altre applicazioni di IA[91], anche per i *deepfake* la scelta segue una duplice direzione: sul piano dell'imposizione di obblighi e divieti, si continua a preferire, finché possibile, la strada della "*no regulation*" o al più ad incoraggiare l'autoregolamentazione privata, limitando l'intervento del regolatore pubblico a casistiche eccezionali; sul piano della competenza, all'opposto della strada seguita dall'UE, si consolida, almeno per ora, la tendenza del legislatore federale ad evitare interventi normativi, a fronte di legislatori statali che sempre più frequentemente optano per una regolamentazione delle nuove tecnologie. Infine, in Cina l'intervento del governo in materia di nuove tecnologie assume contorni sempre più evidenti, all'interno di un sistema in cui gli operatori del settore sono chiamati a condividere integralmente la visione e i valori promossi dal partito, con aumento del controllo e degli obblighi cui sono sottoposti.

Tuttavia, e ferme restando tutte le differenze evidenziate, vale la pena evidenziare che il tema dei *deepfake* sembra unire i tre approcci, nella misura in cui comunque la diffusione di questo genere di contenuti, almeno in alcuni settori particolarmente sensibili, è guardata con grande sospetto dai regolatori pubblici, tanto da comportare l'introduzione di normative *ad hoc*: in particolare, l'esigenza di preservare le elezioni da indebiti condizionamenti o influenze è avvertita da tutte le autorità di governo, quale che sia l'orientamento politico o ideologico. In questo senso, fatte salve le differenti concezioni sul piano filosofico, giuridico e sociale, la riflessione sulla regolamentazione dei *deepfake*, soprattutto se portata avanti in un contesto di cooperazione tra le autorità politiche delle tre potenze, potrebbe costituire un "ponte" per avvicinare le visioni sul futuro e sui rischi dell'IA.

[90] Cfr. A. Bradford, *Digital empires: The global battle to regulate technology*, Oxford, 2023.

[91] Il riferimento è, ad esempio, alle tecnologie di riconoscimento facciale. Cfr., per una comparazione in materia tra i tre sistemi, W. Chen – M. Wang, *Regulating the use of facial recognition technology across borders: A comparative case analysis of the European Union, the United States, and China*, in *Telecommunications Policy*, 2, 2023.