DEEPFAKE E REGOLAMENTAZIONE IA NEL DIRITTO EUROPEO CONTEMPORANEO

Teresa D'Aniello*

Estratto della Tesi di Laurea in Diritto internazionale penale,
discussa presso il Dipartimento di Scienze Giuridiche (Scuola di Giurisprudenza)
dell'Università degli Studi di Salerno,
il 18 dicembre 2024
Relatore: Prof.ssa Anna Oriolo

Il lavoro di ricerca ha costituito l'oggetto della Tesi di Laurea in Diritto internazionale penale relativa al tema dei "Crimini transnazionali e cyberspace", discussa presso il Dipartimento di Scienze Giuridiche il 18 dicembre 2024.
L'elaborato si è focalizzato sull'esame delle evoluzioni normative, dottrinali e giurisprudenziali relative al fenomeno della criminalità transnazionale connessa al cd. *cyber spazio*. Segnatamente, negli ultimi decenni, la globalizzazione e le innovazioni tecnologiche hanno cambiato radicalmente le dinamiche politiche, economiche e sociali, dando origine a nuove sfide legate alla criminalità transnazionale. La crescente interconnessione globale ha facilitato la diffusione di crimini che superano i confini nazionali, mettendo in evidenza la difficoltà di contrastarli con strumenti giuridici tradizionali. Muovendo da queste premesse, il lavoro di ricerca esplora in chiave critica la disciplina della prevenzione e repressione criminalità transnazionale con un particolare focus sui crimini informatici l'intelligenza artificiale e i *deep fake* prospettando possibili soluzioni in termini di più efficace contrasto alle nuove minacce di interesse globale.
L'originalità dell'analisi condotta (intimamente connessa alla attualità del tema trattato) il rigore metodologico della ricerca e la ricchezza delle fonti normative e giurisprudenziali diligentemente esaminate rendono il contributo idoneo ad evidenziare la dimensione globale dei cyber crimes e la priorità assunta dal contrasto alla nuova forma di criminalità transnazionale costituita dai *deep fake* e quindi meritevole di pubblicazione.
(Anna Oriolo, Professore Associato di Diritto internazionale, DSG, Relatore)

Presentazione del lavoro:
Il contributo sottoposto alla valutazione per la pubblicazione rappresenta l'esito di un approfondito lavoro di ricerca sviluppato nell'ambito della Tesi di Laurea in Diritto internazionale penale, discussa il 18 dicembre 2024 presso il Dipartimento di Scienze Giuridiche, e dedicata al tema dei "Crimini transnazionali e cyberspace".
L'elaborato affronta con taglio critico e aggiornato l'evoluzione della criminalità transnazionale in rapporto al crescente impiego delle tecnologie digitali e all'emersione di nuove minacce derivanti dall'uso deviante dell'intelligenza artificiale e dal fenomeno del c.d. deep fake. In un contesto segnato da una sempre più accentuata interdipendenza globale, lo studio analizza le trasformazioni del crimine oltreconfine, mettendo in luce le difficoltà degli strumenti giuridici tradizionali nell'intercettare fenomeni criminali in continua mutazione e garantire la veridicità delle informazioni presenti nella infosfera.
La ricerca si distingue per l'accuratezza metodologica, la capacità di selezione e sistematizzazione delle fonti normative, dottrinali e giurisprudenziali, nonché per l'attenzione alle prospettive di riforma e agli scenari di governance globale della cybersicurezza. Il taglio teorico-pratico dell'analisi, unito alla pertinenza del tema trattato, consente di cogliere appieno la portata delle sfide emergenti e di

*Laureata in Giurisprudenza presso il Dipartimento di Scienze Giuridiche (Scuola di Giurisprudenza), Università degli Studi di Salerno.

avanzare proposte ragionate in ordine al rafforzamento delle strategie di prevenzione e contrasto della criminalità digitale.
Per l'attualità del tema, la solidità dell'impianto scientifico e la qualità complessiva dell'argomentazione, ritengo il contributo meritevole di pubblicazione.
(Fabio Coppola, Professore incaricato di Diritto penale e intelligenza artificiale, DSG)

SOMMARIO: 1. - Introduzione: l'Unione europea a confronto con la dualità dell'Intelligenza artificiale tra opportunità economica e minaccia per la sicurezza; 2. - Il Contrasto alla disinformazione: L'intervento giuridico e normativo dell'Unione europea; 3. - *Deepfake* e IA: Le sfide del futuro per la veridicità dell'informazione; 4. - Il ruolo dell'Unione europea nella gestione dei contenuti manipolati; 5. - L'*AI Act* e la nuova *governance* europea dell'Intelligenza artificiale; 6. - Sistemi ad alto rischio: *Deepfake* e regolamentazione preventiva in Europa; 7. - Conclusioni

1. - Introduzione: l'Unione europea a confronto con la dualità dell'Intelligenza artificiale tra opportunità economica e minaccia per la sicurezza

Nel contesto delle politiche di *cybersecurity* adottate dall'Unione europea (UE) nel 2023, l'attenzione verso l'Intelligenza artificiale (IA) emerge come uno degli elementi chiave nella ridefinizione della sovranità digitale e della sicurezza cibernetica. A partire dall'insediamento della Commissione von der Leyen nel dicembre 2019, l'UE ha posto una crescente enfasi sulla sovranità digitale come risposta alle vulnerabilità percepite nei confronti di attacchi esterni, che provengono non solo da Stati non appartenenti all'UE, ma anche da attori privati il cui comportamento potrebbe non allinearsi ai valori fondamentali dell'Unione[1]. In questo contesto, la *governance* tecnologica diventa un pilastro centrale dell'azione esterna dell'UE, mirando a ridurre la dipendenza nell'approvvigionamento di prodotti per l'*high-tech* dalle potenze straniere e a rafforzare la propria c.d. autonomia strategica[2]. Al contempo, la crescente integrazione della *cybersecurity* con le misure volte a garantire la resilienza tecnologica dell'UE è il risultato di un'evoluzione del quadro

[1] I principali documenti in cui si riflette questa linea di pensiero includono la Comunicazione congiunta al Parlamento europeo e al Consiglio, del 16 dicembre 2020, *La strategia dell'UE in materia di cibersicurezza per il decennio digitale*, JOIN(2020) 18 final; il 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*, in GUUE L 277, del 27 ottobre 2022; ed il 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*, in GUUE L 265 del 12 ottobre 2022.
[2] F. Crudele, *CyberWar. L'invisibile guerra digitale. Strategie e soluzioni per i nuovi campi di battaglia*. 2024.

geopolitico globale, in cui minacce ibride[3], disinformazione e attacchi cibernetici[4] sono visti come strumenti attraverso i quali attori esterni possono compromettere la stabilità dell'UE[5].

In questo scenario, l'Intelligenza artificiale gioca un ruolo doppio: se da un lato rappresenta uno strumento potenzialmente utile per migliorare le difese cibernetiche; dall'altro comporta significativi rischi, se utilizzata in modo offensivo[6]. La guerra in Ucraina, ad esempio, ha messo in luce diverse vulnerabilità informatiche e l'importanza di rafforzare le capacità difensive contro attacchi cibernetici, che si sono intensificati con l'*escalation* del conflitto. Le attuali dinamiche geopolitiche hanno reso evidente come le guerre moderne non si svolgano più solamente sui campi di battaglia fisici, ma anche nel cyberspazio. Pertanto, l'impiego dell'Intelligenza artificiale nell'ambito della *cybersecurity* dell'Unione europea non si limita alla protezione delle infrastrutture critiche, ma include anche la risposta a minacce sempre più sofisticate, come quelle provenienti dalla *cyberwar*[7]. Segnatamente, gli attacchi cibernetici, che variano dall'interruzione delle reti alle operazioni di disinformazione mirate, sono diventati una forma di guerra ibrida sempre più utilizzata da Stati e gruppi sovranazionali per minare la stabilità politica ed economica di altri Paesi[8]. La crescente

[3] Il *Tallinn Manual 2.0* è un documento non vincolante elaborato da esperti in diritto internazionale, che fornisce linee guida sull'applicabilità delle norme di diritto internazionale agli attacchi cibernetici. Secondo la Regola 92 del *Tallinn Manual 2.0*, un *cyber attack* è definito come: "*a cyber operation, whether offensive or defensive, that is reasonably expected to cause injury or death to persons or damage or destruction to objects*.". Questa definizione considera come attacchi cibernetici sia le operazioni di carattere offensivo sia quelle difensive, purché siano ragionevolmente destinate a causare danni a persone o beni materiali. D'altro canto, il Regolamento (UE) 2019/796 del 17 maggio 2019, che ha natura vincolante per gli Stati membri dell'Unione europea, definisce i *cyber-attacks* in maniera più ampia e con riferimento alle minacce dirette alla sicurezza dell'Unione. L'Articolo 1, comma 5 specifica: "*Cyber-attacks constituting a threat to the Union include those carried out against its institutions, bodies, offices and agencies, its delegations to third countries or to international organisations, its common security and defence policy (CSDP) operations and missions and its special representatives*.". Attraverso questa definizione viene ampliato il concetto di attacco cibernetico includendo tutte le minacce rivolte contro le istituzioni dell'Unione europea e le sue politiche di sicurezza e difesa comune.

[4] Le minacce ibride si riferiscono a un insieme di azioni, sia militari che non militari, che possono essere adottate da attori statali o non statali per raggiungere obiettivi politici specifici, senza però superare la soglia di un conflitto armato dichiarato. Come stabilito nel NATO *Hybrid Threats and Hybrid Warfare Reference Curriculum 2024*, "*Hybrid threat refers to potential overt and covert military and nonmilitary actions that a state or nonstate actor might take to undermine a targeted society and achieve their political goals. These actions go beyond the normal interaction of states, without necessarily seeking a warlike aim. Not all hybrid threats can be clearly classified as military problems*.". Pur non limitandosi alla sfera strettamente militare, le minacce ibride comprendono anche forme indirette di intervento che, con modalità non convenzionali, destabilizzano e minano la coesione interna di una società. A livello europeo, il concetto di minacce ibride viene delineato nel *Joint Communication to the European Parliament and the Council – Joint Framework on countering hybrid threats, a European Union response*, il quale sottolinea che "*While definitions of hybrid threats vary and need to remain flexible to respond to their evolving nature, the concept aims to capture the mixture of coercive and subversive activity, conventional and unconventional methods (i.e. diplomatic, military, economic, technological), which can be used in a coordinated manner by state or non-state actors to achieve specific objectives while remaining below the threshold of formally declared warfare.*". La Commissione evidenzia come queste minacce utilizzino una combinazione di attività coercitive e sovversive, che spaziano da interventi diplomatici ed economici a operazioni militari non convenzionali, tutte finalizzate a ottenere vantaggi politici senza un'escalation formale verso la guerra. Il documento *A Europe that Protects: Countering Hybrid Threats* (giugno 2018) aggiunge ulteriori dettagli, definendo le minacce ibride come "*Hybrid threats combine conventional and unconventional, military and non-military activities that can be used in a coordinated manner by state or non-state actors to achieve specific political objectives. Hybrid campaigns are multidimensional, combining coercive and subversive measures, using both conventional and unconventional tools and tactics. They are designed to be difficult to detect or attribute. These threats target critical vulnerabilities and seek to create confusion to hinder swift and effective decision-making*.".

[5] M. Willett, *The Cyber Dimension of the Russia–Ukraine War*, in *Adelphi Series* 64 (2024) 105-124.

[6] Il Rapporto Clusit del 2023 evidenzia l'evoluzione delle minacce, ponendo l'accento sull'impiego crescente dell'IA generativa come una potenziale arma per attacchi informatici. Le sue applicazioni potrebbero includere, tra l'altro, l'automazione dell'individuazione di vulnerabilità, il *phishing* più sofisticato e l'ingegneria sociale avanzata, ampliando notevolmente la superficie di attacco e rendendo più difficile la protezione delle infrastrutture critiche dell'UE. Vd. Security Summit, *Clusit - Rapporto 2024 sulla Sicurezza ICT in Italia*, 2024, 145s.

[7] C. Tan, *Ultimate Cyberwarfare for Evasive Cyber Tactics,* 2024.

[8] G. Valenti, *La "Cyberwar". Le sue modalità e gli strumenti giuridici per contrastarla*, in *DPCE Online* 63 (2024) 537-554.

sofisticazione di queste minacce richiede una risposta adeguata da parte delle istituzioni europee, che vanno ben oltre il rafforzamento delle difese passive e abbracciano approcci di deterrenza e resilienza digitale attiva[9]. La *cyberwar* sta evolvendo grazie all'utilizzo di tecnologie come l'IA generativa, che può automatizzare e migliorare le tecniche di attacco, particolarmente pericolose perché consentono di personalizzare gli attacchi, mirandoli a singoli individui o organizzazioni. Questo sviluppo rende gli attacchi informatici più difficili da rilevare e neutralizzare, aumentando i rischi per la sicurezza nazionale e per l'integrità dei sistemi informatici dell'UE.

Tali preoccupazioni sono amplificate dalle dichiarazioni della Presidentessa della Commissione europea, Ursula von der Leyen, la quale ha sottolineato la necessità di regolamentare l'IA, trattandola come una delle principali minacce alla sicurezza, tanto quanto la proliferazione nucleare[10].

La velocità con cui l'IA si sviluppa richiede una risposta immediata, tanto sul piano normativo quanto operativo, al fine di evitarne un utilizzo dannoso, ma al contempo garantirne uno sviluppo sicuro ed etico. In questa prospettiva, è stato adottato il Regolamento (UE) 2024/1689[11], più noto come *AI Act*, ovvero un quadro normativo che cerca non solo di disciplinare l'uso dell'IA all'interno dell'UE, difendendo i suoi cittadini da minacce informatiche, ma anche di promuovere *standard* internazionali per il suo utilizzo sicuro, esportando i suoi valori e norme in materia di IA e *cybersecurity*. A fronte di un'evidente componente geopolitica, il Regolamento sull'IA, pertanto, non è solo una risposta a rischi immediati, ma parte di una strategia di lungo termine per consolidare la sovranità digitale e ridurre le vulnerabilità esterne, bilanciando l'innovazione tecnologica con la necessità di proteggere i propri interessi e valori, mettendo l'Unione in una posizione di *leadership* globale nella regolamentazione tecnologica.

---

[9] V. Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni, del 9 marzo 2021, *Bussola per il digitale 2030: il modello europeo per il decennio digitale*, COM/2021/118 final in EU-lex.

[10] V. *Discorso della Presidente Von Der Leyen sullo stato dell'Unione 2023*, reperibile al sito https://italy.representation.ec.europa.eu/notizie-ed-eventi/notizie/discorso-della-presidente-von-der-leyen-sullo-stato-dellunione-2023-2023-09-13_it.

[11] 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*, in GUUE L, 2024/1689, del 12 luglio 2024. Articolo 1 "Lo scopo del presente regolamento è migliorare il funzionamento del mercato interno e promuovere la diffusione di un'intelligenza artificiale (IA) antropocentrica e affidabile, garantendo nel contempo un livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea, compresi la democrazia, lo Stato di diritto e la protezione dell'ambiente, contro gli effetti nocivi dei sistemi di IA nell'Unione, e promuovendo l'innovazione". Sul punto v. M. Colacurci, *Quale diritto penale dell'IA?,* in *In Jus* 3 (2023) 357-378. C. Novelli, *L'Artificial Intelligence Act Europeo: alcune questioni di implementazione*, in *federalismi.it* (2024) 100s.

"L'AIA ambisce a stabilire standard chiari e affidabili per la progettazione, l'implementazione e l'utilizzo dei sistemi IA, mirando a creare un mercato interno efficace. Un mercato interno ben funzionante dovrebbe essere sicuro per i consumatori, attrattivo per gli investitori e in grado di incentivare l'innovazione tecnologica. L'Unione europea, con questa regolamentazione, potrebbe diventare un pioniere influente, avendo già stimolato paesi come Stati Uniti e Regno Unito a sviluppare proprie normative sull'IA. Tuttavia, esiste il rischio che, in caso di mancata approvazione del testo normativo entro le elezioni del 2024 o comunque di inefficacia dell'infrastruttura regolamentare, l'UE possa perdere il vantaggio da first-muover che aveva acquisito muovendosi con anticipo rispetto ad altri sistemi giuridici. In tal caso, l'Unione Europea rischierebbe di perdere l'occasione di diventare un mercato attrattivo per la progettazione e lo sviluppo dell'intelligenza artificiale. Per scongiurare questa ipotesi, è fondamentale che la regolamentazione stabilisca regole che mantengano elevati standard di sicurezza senza ostacolare la produzione, l'utilizzo e l'innovazione nell'IA. L'approccio basato sul rischio adottato dall'UE ha diversi punti di forza. Per esempio, consente di stabilire priorità e obiettivi in modo chiaro e trasparente, proteggendo i valori fondamentali dell'UE e favorendo lo sviluppo del mercato interno dell'IA…La maggior parte dei regolamenti basati sul rischio trova ispirazione nel principio di precauzione per il quale se un'attività può avere conseguenze gravi e incerte, allora bisogna adottare le precauzioni necessarie o evitarla del tutto. Pertanto, le misure di contenimento e prevenzione del rischio, anche quando introdotte per via legislativa, possono essere avviate anche in assenza di evidenze conclusive sui potenziali effetti dannosi. I regolamenti basati sul rischio consistono generalmente di tre parti: la valutazione, la gestione e la comunicazione del rischio. Anche l'AIA mostra questa impostazione, dedicando maggiore spazio alle misure di gestione del rischio, sottoforma di oneri normativi a carico dei soggetti che fanno parte della catena del valore dei sistemi di IA (principalmente deployers e providers).".

L'adozione di tale regolamento non può prescindere da un inquadramento più ampio che abbraccia il contesto normativo in cui si inserisce, caratterizzato da un'evoluzione progressiva delle strategie europee in materia di cybersicurezza e regolazione digitale. Il rafforzamento delle capacità dell'UE nel contrastare le minacce informatiche e garantire un mercato digitale sicuro ha trovato espressione in diversi atti legislativi e programmatici che si pongono in stretta correlazione con *l'AI Act*, delineando un quadro di intervento organico e strutturato.

L'iniziativa legislativa in materia di Intelligenza artificiale si inserisce nel solco tracciato dalla strategia europea in materia di cybersicurezza, formalizzata nella Comunicazione congiunta del 16 dicembre 2020, con la quale l'Unione ha individuato gli strumenti per la tutela dello spazio digitale nel contesto del decennio digitale. L'adozione di questo quadro strategico ha segnato un punto di svolta nella *governance* della sicurezza informatica, ponendo le basi per un intervento regolatorio più incisivo e orientato alla resilienza tecnologica.

Parallelamente, il rafforzamento della sicurezza digitale si è concretizzato attraverso la regolamentazione del mercato dei servizi digitali e delle piattaforme *online*, con l'approvazione del Regolamento (UE) 2022/2065, che ha riformato il quadro normativo esistente introducendo obblighi specifici per le piattaforme di intermediazione, con l'obiettivo di garantire un ecosistema digitale più trasparente e sicuro. Alla medesima finalità risponde il Regolamento (UE) 2022/1925, il quale disciplina i mercati digitali con particolare attenzione alla prevenzione delle pratiche anticoncorrenziali e alla tutela della libertà di scelta degli utenti, elementi essenziali per la creazione di un ambiente digitale equo e accessibile[12].

L'interconnessione tra tali strumenti normativi e il Regolamento sull'IA evidenzia la volontà dell'Unione di adottare un approccio integrato alla regolazione del digitale, in cui sicurezza, innovazione e tutela dei diritti fondamentali si intrecciano in una strategia coesa. Tale prospettiva trova riflesso non solo nell'obiettivo di disciplinare i rischi connessi allo sviluppo delle nuove tecnologie, ma anche nella definizione di un modello normativo capace di imporsi come *standard* internazionale, rafforzando l'autonomia strategica dell'UE e la sua capacità di incidere sulle dinamiche della *governance* globale dell'Intelligenza artificiale e della cybersicurezza.

2. - Il Contrasto alla disinformazione: L'intervento giuridico e normativo dell'Unione europea

La disinformazione è un fenomeno che affonda le sue radici nella storia, quale elemento costante dell'esperienza umana che si è evoluto insieme agli strumenti di comunicazione e al progresso tecnologico. Dall'antichità, la manipolazione delle informazioni è stata utilizzata come mezzo per consolidare il potere, influenzare l'opinione pubblica e ottenere vantaggi strategici[13]. Lungo il corso della storia, il controllo e la distorsione delle informazioni hanno assunto forme diverse, adattandosi al contesto culturale e tecnologico del tempo

Un esempio emblematico di come la manipolazione dell'informazione abbia avuto implicazioni su vasta scala si può rintracciare nei primi decenni del XIX secolo: false notizie sulla sconfitta di Napoleone nel 1814 furono diffuse intenzionalmente per alterare il mercato azionario britannico, arricchendo gli speculatori responsabili della loro creazione. L'episodio segna uno dei primi casi documentati in cui la disinformazione venne sfruttata per obiettivi economici, dimostrando il suo potenziale di influire non solo sulle percezioni collettive, ma anche sulle dinamiche finanziarie globali[14]. Nel XX secolo, l'evoluzione dei mezzi di comunicazione, dalla stampa ai primi *media* audiovisivi, intensificò il ruolo della disinformazione come arma politica e sociale. Regimi autoritari fecero un uso sistematico della propaganda per manipolare le masse, trasformando l'informazione in

[12] 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)*. 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)*.

[13] T. Sintonen et al., *Leadership Manipulation and Ethics* in *Storytelling. J Bus Ethics* 116 (2013) 415–431.

[14] R. Dale, *Napoleon is dead! The Great Stock Exchange Fraud* in *The History Press* (2017).

uno strumento di controllo totalitario[15]. Con il progredire della modernità, l'utilizzo strategico della disinformazione divenne ancora più evidente in ambito militare, come dimostrato dall'*Operazione Bodyguard* durante la Seconda guerra mondiale, dove la diffusione di informazioni false fu cruciale per depistare il nemico e garantire il successo dello sbarco in Normandia[16].

Tuttavia, con l'avvento della rivoluzione digitale, la disinformazione ha raggiunto una nuova dimensione, diventando più pervasiva e complessa. I *social media* e le tecnologie basate sull'Intelligenza artificiale hanno reso possibile la creazione e la diffusione di contenuti ingannevoli a una velocità senza precedenti, amplificando il loro impatto su scala globale[17]. Fenomeni come le *fake news* e i *deepfake*, manifestazioni emblematiche di questo fenomeno, rappresentano sfide significative per la società contemporanea[18]. La storia della disinformazione, dunque, è un racconto di adattamento e trasformazione, un fenomeno che si è costantemente evoluto per rispondere alle esigenze di chi lo utilizza. Oggi, di fronte a tecnologie sempre più sofisticate, la capacità di affrontare questa sfida richiede non solo una comprensione delle sue radici storiche, ma anche lo sviluppo di strumenti normativi, tecnologici e educativi in grado di salvaguardare l'integrità dell'informazione in una società sempre più connessa[19]. Se in passato il controllo delle informazioni era prerogativa di pochi, oggi la democratizzazione degli strumenti tecnologici ha reso la disinformazione una minaccia diffusa e trasversale, capace di colpire ogni settore della vita pubblica e privata[20].

Le "*fake news*", ovvero notizie o informazioni false, ingannevoli o errate, rappresentano un fenomeno che ha assunto rilevanza globale negli ultimi anni, grazie all'espansione dei mezzi di comunicazione digitali, come i *social media*, i siti *web* e le applicazioni di messaggistica[21]. Le notizie possono essere diffuse in modo deliberato o accidentale e, spesso, sono frutto di manipolazioni o di interpretazioni errate dei fatti. Sebbene il termine "*fake news*" venga spesso utilizzato in modo generico, è importante sottolineare che tale fenomeno non riguarda soltanto notizie inventate, ma anche la distorsione delle informazioni reali, tramite l'omissione di dati importanti o l'interpretazione

---

[15] R. Kuo, A. Marwick, *Critical Disinformation Studies: History, Power, and Politics* in *Misinformation Rewiev* (2021).
[16] T. Rid, *Misure attive. Storia segreta della disinformazione* (2022) 200-245.
[17] F. Sanchez Barillao*, La difesa della democrazia pluralista dinanzi alla paura e alla disinformazione*, in *MediaLaws* 2 (2022).
[18] Nel dicembre del 2020, l'Agenzia dell'Unione europea per la Cybersicurezza (ENISA) ha emesso un monito circa i pericoli legati all'uso illecito delle tecnologie basate sull'Intelligenza artificiale, trattando la tematica nel suo *report* sulle «*AI Cybersecurity Challenges*». Il *report* ha evidenziato come l'adozione e l'utilizzo di sistemi di IA non controllati possano esporre i cittadini, le istituzioni e l'intero sistema democratico a rischi significativi, tra cui l'utilizzo abusivo di tecnologie come i *deepfake*, che potrebbero alterare il corso della giustizia e della comunicazione politica, creando disinformazione e compromettendo la sicurezza dei dati personali. Gli sviluppi tecnologici, infatti, pongono interrogativi cruciali su come bilanciare la protezione dei diritti fondamentali con la promozione di un progresso innovativo e responsabile. Inizialmente, questi strumenti sono stati accolti come strumenti di democratizzazione dell'informazione, capaci di abbattere le barriere tradizionali di accesso all'informazione, rappresentando un potente strumento di partecipazione sociale e politica. La possibilità di connettersi senza intermediari, superando i filtri imposti da istituzioni o media tradizionali, ha spinto molti a considerare i *social media* come veicoli per una democrazia più diretta e partecipativa. La Primavera araba ne è stato un esempio emblematico, dimostrando il potenziale positivo di tali tecnologie in contesti di lotta contro regimi autoritari. Tuttavia, con l'avvento di fenomeni come lo scandalo *Cambridge Analytica* e la crescente consapevolezza dei rischi connessi alla manipolazione dell'opinione pubblica, è emerso il volto oscuro di queste tecnologie. Se inizialmente i *social media* sono stati visti come strumenti di *empowerment* democratico, ben presto sono diventati il terreno fertile per la diffusione di *fake news*, campagne di disinformazione e manipolazione dell'opinione pubblica.
[19] M. G. Coggiola, G. P. Siroli, *Fake news*: *propaganda e disinformazione all'epoca di Internet*, 2018, 32-37.
[20] F. Sanchez Barillao, *La difesa* cit.
[21] Con il termine *fake news* si intende «un'informazione in parte o del tutto non corrispondente al vero, prodotta e divulgata intenzionalmente o inintenzionalmente attraverso il *Web*, i media o le tecnologie digitali di comunicazione, e caratterizzata da un'apparente plausibilità, quest'ultima alimentata da un sistema distorto di aspettative dell'opinione pubblica e da un'amplificazione dei pregiudizi che ne sono alla base, che ne agevola la condivisione e la diffusione pur in assenza di una verifica delle fonti, tale da ledere beni giuridici individuali, come l'onore e la reputazione, ovvero finalizzata a incidere, direttamente o indirettamente, sulla libertà dei cittadini di esercitare il diritto di voto e ad incidere sul corretto funzionamento delle istituzioni democratiche». V. Guerini, *Fake news e diritto penale, la manipolazione digitale del consenso nelle democrazie liberali*, Torino 2020, 28-29.

parziale di eventi[22]. Potremmo quindi definire le *fake news* come un sottotipo di disinformazione, caratterizzate da informazioni false o manipolate create intenzionalmente per ingannare il pubblico[23].

La diffusione di *fake news* può riguardare temi di ogni tipo, dalla politica alla salute, dall'economia alla scienza, influenzando l'opinione pubblica e modificando la percezione di eventi cruciali. Le motivazioni che spingono alla diffusione di *fake news* sono molteplici e variegate. Possono essere scopi politici, volti a manipolare l'opinione pubblica su temi rilevanti per elezioni o decisioni politiche; scopi finanziari, con il fine di generare guadagni attraverso il sensazionalismo o la viralità; o, più semplicemente, il desiderio di attirare l'attenzione, creando allarme o curiosità. Le informazioni, in molti casi, non sono sostenute da fonti verificabili o sono interamente inventate.

I meccanismi di manipolazione possono contribuire a creare un clima di instabilità politica, generando divisioni profonde tra i vari gruppi sociali e politici, e alimentando conflitti[24]. Infatti, la proliferazione di notizie false minaccia seriamente la democrazia stessa. Il fenomeno si pone in diretto contrasto con i valori sanciti dall'art. 2 TUE, in particolare con i principi di democrazia e Stato di diritto, cardini dell'ordinamento giuridico dell'Unione[25]. La proliferazione di notizie false rappresenta una minaccia crescente per il corretto funzionamento delle democrazie liberali, alimentando la disinformazione e contribuendo alla destabilizzazione politica. Questi meccanismi di manipolazione, infatti, non solo generano divisioni profonde tra i vari gruppi sociali e politici, ma compromettono anche la capacità dei cittadini di formarsi un'opinione consapevole e libera, essenziale per il buon funzionamento del sistema democratico.

La libertà di espressione, sancita dall'art. 11 della Carta dei diritti fondamentali dell'Unione europea, rappresenta uno dei principi fondamentali su cui si fonda l'ordinamento giuridico delle democrazie liberali. Tale libertà è essenziale per garantire il pluralismo e il funzionamento di una società democratica, fungendo da garanzia per la protezione di altri diritti fondamentali, tra cui la partecipazione politica dei cittadini. Il diritto di esprimere liberamente il proprio pensiero e di ricevere informazioni verificate sono componenti essenziali di tale protezione, in quanto consentono agli individui di formarsi un'opinione libera e consapevole, necessaria per una piena partecipazione alla vita pubblica. In particolare, l'art. 11 della Carta stabilisce che "ogni individuo ha diritto alla libertà di espressione. Tale diritto include la libertà di opinione e la libertà di ricevere o di comunicare informazioni o idee senza che vi possa essere ingerenza da parte delle autorità pubbliche e senza limiti di frontiera".

Tuttavia, il fenomeno delle *fake news* si pone in contrasto con il principio di libertà di espressione, poiché le notizie false, in quanto manipolazione intenzionale dell'informazione, alterano la percezione pubblica e possono indirizzare le scelte politiche in modo distorto, minando così la rappresentanza democratica. In un sistema democratico, infatti, è imprescindibile che i cittadini possano accedere a informazioni veritiere e verificate, al fine di esercitare un diritto di voto informato e consapevole. La continua esposizione a contenuti non verificati può infatti spingere i cittadini a diventare scettici nei confronti delle fonti di informazione ufficiali e tradizionali, mettendo a rischio

[22] A. Kaplan, *Artificial intelligence, social media, and fake news: is this the end of democracy?,* 2020, 149-161.

[23] Durante eventi cruciali, come le elezioni presidenziali statunitensi del 2016 o il *referendum* sulla *Brexit,* le *fake news* sono state utilizzate per polarizzare l'opinione pubblica e influenzare le scelte dei cittadini. La pandemia di Covid-19 ha ulteriormente evidenziato i pericoli di questo fenomeno, con l'Organizzazione mondiale della sanità (OMS) che ha parlato di una vera e propria "infodemia", ossia una sovrabbondanza di informazioni, spesso false o fuorvianti, che ha ostacolato gli sforzi sanitari e alimentato il panico.

[24] A. Davola , M. Monti, *Disinformazione e Manipolazione del mercato*, 2023, 37-67.

[25] *"In the construction of the area of freedom, security and justice, for example, Art. 67 TFEUexpressly demands respect for both fundamental rights and 'the legal systems and traditions of the Member States'. Some scholars deem this merely highlights the prominence of the principles of subsidiarity and proportionality in EU criminal law,51 yet in a different perspective, reliance on national law in counteracting transnational crimes reflects the legal principle under Union law based on respect for sovereignty and the democratic decision-making of Member States,52 which could 'decentralize' judicial protection in the EU,53 and consequently, impact individual rights.".* Per un approfondimento sul tema vd. A. Oriolo, *The Rule of Law, Transnational Crimes, and the Human Rights-Based Approach in the European Union: The Court of Justice as Ultimate Guardian of the 'Good' Law*s, in T. Russo, A. Oriolo, G. Dalia (eds*) Solidarity and Rule of Law The New Dimension of EU Security*, 2023.

il processo deliberativo su cui si fonda la democrazia. Il contrasto alla disinformazione si configura non solo come una questione di sicurezza, ma come una priorità nell'ambito della protezione dei valori fondamentali dell'Unione.

Le conseguenze economiche delle *fake news* sono altrettanto gravi. La diffusione di informazioni false può danneggiare le reputazioni aziendali, influenzare l'andamento dei mercati finanziari e creare una generale sfiducia nel sistema economico. Non è raro che voci infondate su crisi finanziarie o fallimenti di grandi aziende causino impatti devastanti sui mercati, anche in assenza di qualsiasi fondamento reale[26].

Nonostante gli sforzi di piattaforme tecnologiche e istituzioni giornalistiche per contrastare il fenomeno, come l'etichettatura dei contenuti falsi o lo sviluppo di strumenti di *fact-checking*, il problema persiste, ponendo interrogativi etici e giuridici legati alla libertà di espressione e al ruolo delle piattaforme come arbitri della verità[27]. L'Intelligenza artificiale sta svolgendo un ruolo cruciale nella gestione del problema delle *fake news*, ma allo stesso tempo riveste un ruolo ambivalente sulla disinformazione[28].

Innanzitutto, gli strumenti di IA vengono utilizzati per identificare contenuti manipolati, analizzare dati e smascherare schemi fraudolenti che l'uomo non sarebbe in grado di rilevare altrettanto velocemente[29]. Ovviamente, nonostante i progressi tecnologici, la lotta contro le *fake news* non può essere affidata esclusivamente all'IA. La distinzione tra "notizie false" e "notizie legittime" richiede anche il giudizio critico che solo l'intervento umano può fornire[30]. La disinformazione, infatti, è un fenomeno che sta crescendo con preoccupante regolarità, in particolare a causa della scelta sempre più diffusa di selezionare notizie che si allineano con le proprie convinzioni preesistenti. I *social media* giocano un ruolo fondamentale nel rinforzare questa tendenza, poiché gli utenti tendono a interagire principalmente con chi condivide opinioni simili alle loro. Questo è un problema che richiede soprattutto cultura, educazione e responsabilità individuale. La tecnologia è un aiuto

[26] In tale contesto, le grandi imprese, sempre più consapevoli dell'impatto della disinformazione sul loro *business*, hanno introdotto la figura del "*social manager"*. Una figura professionale che ha il compito di monitorare i social *network*, raccogliere informazioni sulle tendenze, stimolare discussioni *online*, e curare i contenuti aziendali, con l'obiettivo di mantenere una buona reputazione e reagire tempestivamente a eventuali crisi d'immagine. Una parte del lavoro del *social manager* può comprendere anche la manipolazione delle opinioni online, creando contenuti favorevoli all'azienda, anche in modo anonimo, per contrastare notizie negative. In questi casi, l'utilizzo delle piattaforme social assume connotati delicati, giuridicamente rilevanti, poiché il confine tra *marketing* e manipolazione dell'opinione pubblica può essere sottile e controverso.

[27] Il *fact-checking* è il processo di verifica delle informazioni per determinarne la veridicità, la coerenza e l'accuratezza. Viene utilizzato principalmente per analizzare affermazioni pubbliche, notizie, articoli o contenuti sui *social media*, con l'obiettivo di contrastare la diffusione di disinformazione, *fake news* o manipolazioni. Il processo prevede: raccolta delle informazioni attraverso l'identificazione della fonte originale e la contestualizzazione dell'affermazione o della notizia; verifica delle fonti mediante un'analisi delle fonti primarie e confronto con dati affidabili e autorevoli; confronto con fatti consolidati con l'utilizzo di archivi, statistiche o documenti ufficiali per validare o smentire le affermazioni; pubblicazione dei risultati, esponendo l'esito della verifica, spesso accompagnato da spiegazioni dettagliate e riferimenti.

[28] F. Bertoni, *Deepfake*, *ovvero Manipola et impera. Un'analisi sulle cause, gli effetti e gli strumenti per la sicurezza nazionale, nell'ambito dell'utilizzo malevolo dell'Intelligenza artificiale ai fini di disinformazione e propaganda*, in *Ciberspazio e diritto* 1-2 (2019).

[29] Attualmente esistono diversi strumenti che possono contribuire a ridurre la diffusione di notizie false alimentate da IA nelle campagne di disinformazione. Uno degli approcci è l'uso di strumenti di rilevamento per identificare contenuti generati da IA, ma questi non sono infallibili. Ad esempio, strumenti come *GPTZero* e *DetectGPT* sono in grado di rilevare testi generati dall'Intelligenza artificiale, ma non sono sempre precisi e spesso diventano obsoleti rapidamente. Per affrontare questo problema, l'industria ha creato lo *standard C2PA*, che consente di etichettare i contenuti generati da IA, rivelando la loro origine e storia di proprietà. Ad esempio, *Adobe* utilizza questo standard nei suoi strumenti per consentire ai creatori di aggiungere informazioni sul contenuto quando lo esportano. Inoltre, la UE sta finanziando progetti come vera.ai, che utilizzano IA per verificare in tempo reale i fatti e le fonti aggiornate, cercando di combattere la disinformazione online e promuovere la veridicità delle informazioni. Anche se distinguere tra testi generati da IA e quelli scritti da esseri umani è una sfida complessa, il monitoraggio e la verifica tempestiva dei contenuti rimangono cruciali, specialmente per combattere le notizie dell'ultimo minuto. La ricerca e lo sviluppo di soluzioni tecnologiche in questo ambito sono fondamentali per garantire la trasparenza e la veridicità delle informazioni online.

[30] F. Bertoni, *Deepfake* cit.

fondamentale, ma la prevenzione e il contrasto della disinformazione passano anche attraverso l'educazione alla verifica delle fonti, il rafforzamento del senso critico e la promozione di una comunicazione responsabile[31].

Il fenomeno della disinformazione, alimentato in parte dalla diffusione di contenuti manipolati, ha indotto le istituzioni europee a sviluppare un quadro normativo in grado di rispondere in modo efficace[32].

L'approccio dell'Unione europea alla disinformazione *online* è stato sviluppato attraverso una serie di iniziative politiche e normative che mirano a contrastare i danni provocati dalla diffusione di notizie false e ingannevoli su *Internet*. Le raccomandazioni espresse nel 2018 dal Gruppo di esperti ad alto livello sulle *fake news* e la disinformazione *online*, integrate dalla comunicazione della Commissione europea "*Tackling Online Disinformation: A European Approach"*, hanno posto le basi per una serie di misure normative che mirano a garantire un ambiente digitale trasparente, sicuro e rispettoso dei diritti fondamentali[33]. Nel 2018, il Gruppo di esperti ha redatto un rapporto che, dopo aver analizzato il fenomeno della disinformazione, ha suggerito l'introduzione di un codice di principi per le piattaforme *online*, le quali avrebbero dovuto impegnarsi a rispettare una serie di linee guida riguardanti la gestione dei contenuti *online*. Il codice si basava su un approccio di co-regolamentazione ascendente, che promuoveva la trasparenza degli algoritmi, la protezione dei dati personali, la distinzione tra contenuti sponsorizzati e non, e la cooperazione con i media e i verificatori di fatti[34]. Inoltre, il Gruppo ha riconosciuto l'importanza di monitorare i flussi informativi e ha proposto misure di trasparenza particolarmente rilevanti per la gestione delle pubblicità politiche e

[31] La Risoluzione 2255 (2019) dell'Assemblea Parlamentare del Consiglio d'Europa, adottata il 23 gennaio 2019, pone in evidenza l'importante ruolo che i media di servizio pubblico ricoprono nella lotta contro la disinformazione e la propaganda. Il documento sottolinea come i media siano fondamentali per promuovere un dibattito pubblico pluralistico, stimolare la partecipazione democratica e rafforzare la coesione sociale, in un contesto mediatico sempre più frammentato e influenzato dalla diffusione di notizie false, incitamento all'odio e contenuti ingannevoli. In tale scenario, il ruolo dei media di servizio pubblico emerge come una risposta cruciale per contrastare questi fenomeni. La risoluzione riconosce che, affinché i media di servizio pubblico possano svolgere adeguatamente questa funzione, è indispensabile garantirne l'indipendenza editoriale e istituzionale, nonché fornire risorse finanziarie stabili e adeguate. Questi *media* devono essere in grado di produrre giornalismo di alta qualità, basato su fatti verificati e su una pluralità di opinioni, in un contesto dove la disinformazione si diffonde rapidamente, in particolare attraverso le piattaforme *social*. Grazie alla loro indipendenza, i media di servizio pubblico sono in una posizione privilegiata per contrastare la diffusione di contenuti ingannevoli, operando senza pressioni politiche o economiche, a differenza dei *media* commerciali, spesso soggetti a tali influenze. Inoltre, l'Assemblea sottolinea l'importanza di adottare misure preventive, come iniziative educative rivolte al pubblico, per sensibilizzare l'opinione pubblica sull'importanza della verifica delle fonti, del *fact-checking* e sulla necessità di evitare le "bolle informative", fenomeno alimentato dalla disinformazione. In questa ottica, gli Stati membri sono incoraggiati a sostenere e finanziare iniziative di *fact-checking* collaborativo, in cui vengano impiegate risorse umane e strumenti tecnologici avanzati, come l'intelligenza artificiale, per verificare l'affidabilità dei contenuti generati dagli utenti. Questa cooperazione tra media di servizio pubblico, piattaforme digitali e istituzioni statali è cruciale per sviluppare un sistema di monitoraggio efficace e tempestivo della disinformazione.

32 L'adozione di strategie per contrastare la disinformazione online deve necessariamente basarsi sul coinvolgimento attivo delle piattaforme digitali e degli attori del mercato tecnologico. L' approccio si fonda sulla consapevolezza che non è possibile ricorrere al diritto penale per punire la diffusione di notizie false, salvo che queste non causino danni a interessi diversi dalla semplice veridicità dell'informazione, per i quali si ritenga giustificata e necessaria una protezione penale. Per un approfondimento sul tema vd. E. Birritteri*, Contrasto alla disinformazione, Digital Services Act e attività di private enforcement: fondamento, contenuti e limiti degli obblighi di compliance e dei poteri di autonormazione degli operatori* in *Il Digital Services Act e il contrasto alla disinformazione: responsabilità dei provider, obblighi di compliance e modelli di enforcement* (2022). F. Coppola, *Unione europea, Ai Act e responsabilità penale* in A.Oriolo, A.R. Castaldo et al. (cur.) *Criminalità transnazionale e Unione europea* (2024) 663-676.

[33] V. Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al comitato delle regioni *contrastare la disinformazione online: un approccio europeo*, COM/2018/236. *Report of the independent High level Group on fake news and online disinformation*, *a multi-dimensional approach to disinformation,* 2018.

[34] La co-regolamentazione ascendente è un modello di *governance* in cui norme e *standard* vengono definiti attraverso un processo collaborativo che coinvolge vari attori, sia pubblici che privati, seguendo un approccio "dal basso verso l'alto". In questo schema, l'iniziativa parte principalmente dai soggetti privati o dalla società civile, che collaborano con le istituzioni governative per formulare regole condivise, applicarle e monitorarne l'efficacia.

dei contenuti sponsorizzati. Tuttavia, nonostante le proposte iniziali, la Commissione europea ha constatato che l'approccio di autoregolamentazione, basato sull'impegno volontario delle piattaforme *online*, non fosse sufficiente per affrontare in modo efficace il problema della disinformazione[35]. La Commissione ha evidenziato diverse lacune, come la disparità tra gli impegni presi e le azioni concrete, nonché la mancanza di un sistema di controllo indipendente[36]. Questo scenario ha spinto l'Unione europea a adottare un approccio più rigoroso, che combina la co-regolamentazione con misure normative più stringenti. Il rafforzamento della responsabilità delle piattaforme *online* è diventato un tema centrale, tanto che la Commissione ha sollecitato un coinvolgimento maggiore delle piattaforme stesse, affinché non solo rispettassero gli obblighi giuridici previsti dalla normativa europea, ma attuassero politiche più trasparenti e responsabili riguardo alla gestione dei contenuti e alla pubblicità politica. L'adozione di un sistema di monitoraggio delle piattaforme, la creazione di una rete di verificatori di fatti indipendenti e la trasparenza sugli algoritmi che determinano la visibilità dei contenuti sono diventate priorità nelle successive iniziative normative. Di fronte a questi limiti, è stato necessario passare a un approccio di co-regolamentazione discendente[37], promuovendo la creazione di un Codice di condotta più ambizioso e vincolante, che, pur mantenendo un elemento di autoregolamentazione, avrebbe dovuto essere implementato sotto il controllo di autorità indipendenti[38]. L'obiettivo era garantire che le piattaforme *online* adottassero misure concrete per contrastare la disinformazione, in particolare con riferimento alla trasparenza dei contenuti sponsorizzati, al controllo dei posizionamenti pubblicitari, alla chiusura degli *account* falsi, e alla promozione della visibilità dei contenuti affidabili. La necessità di una vigilanza indipendente era vista come un elemento cruciale per garantire che le piattaforme rispettassero gli impegni presi, prevenendo al contempo abusi o manipolazioni.

Una componente fondamentale del nuovo approccio normativo è l'attenzione dedicata alla protezione dei dati personali, in linea con il Regolamento generale sulla protezione dei dati (GDPR) e la Direttiva sui servizi dei media audiovisivi (AVMSD)[39]. Le piattaforme sono chiamate a rispettare elevati *standard* di responsabilità nella gestione dei contenuti, inclusa la rimozione dei contenuti disinformativi e la regolamentazione rigorosa dell'uso dei dati personali per fini pubblicitari, evitando manipolazioni o abusi. Inoltre, la Commissione ha promosso l'introduzione di una piattaforma sicura per il monitoraggio della disinformazione, in grado di raccogliere e analizzare i dati in modo trasparente, senza compromettere la protezione della *privacy*. Il Codice di condotta rafforzato sulla

[35] Nel 2018 i principali attori del settore digitale come *Google*, *Facebook*, e *Twitter* sono stati chiamati a redigere e attuare un Codice di buone pratiche sulla disinformazione, primo esempio a livello globale di autoregolamentazione volontaria dell'industria per combattere la diffusione di notizie false *online*. Nell'ottobre dello stesso anno, le principali piattaforme digitali, aziende tecnologiche e operatori del settore pubblicitario hanno raggiunto un accordo per adottare norme comuni che promuovessero maggiore trasparenza e contrastassero i fenomeni di disinformazione. Il codice, che si ispirava agli obiettivi delineati dalla Commissione europea nella comunicazione dell'aprile 2018, stabiliva 21 impegni specifici. Tra questi figuravano: trasparenza nella pubblicità politica, per rendere chiara l'origine e lo scopo degli annunci politici; demonetizzazione dei promotori di disinformazione, attraverso la limitazione dei guadagni per chi diffonde contenuti falsi o fuorvianti. I firmatari originali includevano piattaforme come *Facebook*, *Google*, *Twitter* e *Mozilla*, insieme a inserzionisti e operatori pubblicitari. Successivamente, altre aziende si sono unite all'iniziativa: *Microsoft* ha aderito nel maggio 2019, seguita da *TikTok* nel giugno 2020.V. https://digital-strategy.ec.europa.eu/it/library/2018-code-practice-disinformation

[36] SWD (2020)180 Final - *Assessment of the Code of Practice on Disinformation.*

[37] La co-regolamentazione discendente è un approccio normativo in cui le istituzioni pubbliche (governi, autorità regolatorie, organismi internazionali) definiscono linee guida, norme o obblighi legali e delegano l'attuazione e il controllo di tali regole agli attori privati, come aziende, associazioni di settore o altri enti non statali. Il modello segue un approccio "dall'alto verso il basso", dove le decisioni principali sono prese dalle autorità e poi implementate dagli operatori.

[38] V. *The Strengthened Code of Practice on Disinformation 2022.*

[39] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, *relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati)* **,** Direttiva 2010/13/UE del Parlamento europeo e del Consiglio, del 10 marzo 2010 , *relativa al coordinamento di determinate disposizioni legislative, regolamentari e amministrative degli Stati membri concernenti la fornitura di servizi di media audiovisivi*.

disinformazione, adottato nel 2022, costituisce un ulteriore aggiornamento sostanziale rispetto alla versione originaria del 2018. Tra gli impegni assunti dalle piattaforme digitali figurano la demonetizzazione della disinformazione, la piena trasparenza in merito alla pubblicità politica e il potenziamento della cooperazione con i *fact-checker*. Il Codice rafforza l'accesso dei ricercatori ai dati necessari per analizzare l'entità e la diffusione della disinformazione, promuovendo un approccio basato su evidenze empiriche nella gestione dei contenuti digitali. In tale ottica, il 13 febbraio 2025 la Commissione europea e il Comitato europeo per i servizi digitali hanno approvato l'integrazione ufficiale del Codice di condotta nel contesto del DSA, stabilendo che il Codice diventerà un parametro di riferimento per determinare la conformità delle *Very Large Online Platforms* (VLOP) e dei *Very Large Online Search Engines* (VLOSE) ai fini della gestione dei rischi legati alla disinformazione[40].

Questo passaggio rappresenta una formalizzazione dei principi previsti dal *Digital Services Act*, che, in vigore dal novembre 2022, impone alle piattaforme online obblighi di trasparenza, controllo e responsabilizzazione[41].

Il *Digital Services Act*, presentato nel 2020 ed entrato in vigore nel novembre 2022, impone alle piattaforme di garantire la trasparenza nelle politiche di moderazione dei contenuti, di adottare misure per ridurre i rischi derivanti dalla diffusione di contenuti illegali o manipolativi e di migliorare la gestione dei sistemi algoritmici utilizzati per la raccomandazione dei contenuti[42]. Tra le

[40] Il *Digital Services Act* stabilisce una serie di regole per le piattaforme online e i motori di ricerca molto grandi, noti rispettivamente come VLOP (*Very Large Online Platforms*) e VLOSE (*Very Large Online Search Engines*). Questi termini si riferiscono a piattaforme o motori di ricerca che hanno oltre 45 milioni di utenti attivi mensili nell'UE. La Commissione europea ha il compito di designare queste piattaforme, che sono obbligate a rispettare le norme più rigide, dato l'impatto che la loro portata e operato possono avere sulla società. V. SEZIONE 5 (DSA) "Obblighi supplementari a carico dei fornitori di piattaforme online di dimensioni molto grandi e di motori di ricerca online di dimensioni molto grandi per la gestione dei rischi sistemici" art 33 e succ.

[41] Nel 2015, la Commissione europea ha presentato la "*Strategia per il mercato unico digitale in Europa*", un documento che, alla luce della crescente diffusione di contenuti dannosi *online* (*fake news*, *hate speech* e altri contenuti illegittimi), ha aperto il dibattito sul ruolo delle piattaforme digitali e sulla necessità di implementare misure per contrastare tali fenomeni. La comunicazione ha messo in evidenza la difficoltà di intervenire efficacemente contro i contenuti illeciti a causa della mancanza di definizioni chiare e uniformi. Nel 2017, la Commissione europea ha risposto a queste preoccupazioni con una seconda comunicazione, che si concentrava specificamente sulla responsabilizzazione delle piattaforme nella gestione dei contenuti *online*. Essa ha evidenziato la lacuna normativa derivante dalla Direttiva 2000/31/CE, che non forniva una definizione chiara di "contenuto illecito", rendendo di fatto difficile applicare un sistema di responsabilità adeguato e uniforme per gli *Internet Service Providers* (ISP). La Commissione, quindi, ha sollecitato l'armonizzazione delle normative a livello europeo, auspicando l'introduzione di un sistema più coerente per regolamentare le attività delle piattaforme digitali. Questo processo culminò nel 2020, quando la Commissione europea ha presentato ufficialmente la proposta per il Digital Services Act (DSA), con l'obiettivo di modernizzare e razionalizzare il quadro normativo riguardante i servizi digitali, introducendo obblighi di maggiore trasparenza, responsabilizzazione e controllo sulle piattaforme *online*. In particolare, il DSA mira a garantire che le piattaforme siano più trasparenti nelle loro politiche di moderazione dei contenuti, rispondano tempestivamente alla diffusione di contenuti dannosi e migliorino la gestione dei sistemi algoritmici per la raccomandazione dei contenuti. Il DSA è stato presentato come una risposta alla necessità di tutelare i diritti fondamentali degli utenti *online*, in particolare la libertà di espressione, pur preservando un ambiente digitale sicuro e responsabile. Il regolamento è stato visto come una risposta concreta all'espansione incontrollata della disinformazione e ad altre minacce digitali, ponendo le basi per una responsabilizzazione adeguata degli attori coinvolti e creando un sistema di monitoraggio più rigoroso per prevenire e contrastare i rischi legati alla gestione dei contenuti *online*. V. Commissione europea, *Strategia per il mercato unico digitale in Europa*, COM(2015) 192, 6 maggio 2015. Parlamento europeo, Risoluzione del 15 giugno 2017 *sulle piattaforme online e il mercato unico digitale*, 2016/2276(INI)

[42] Il 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 (DSA) ha introdotto nell'ordinamento europeo delle norme volte a garantire un ambiente online sicuro, prevedibile e affidabile, in cui i diritti fondamentali degli utenti dei servizi digitali siano efficacemente tutelati e l'innovazione sia agevolata, contrastando la diffusione di contenuti illegali online e i rischi per la società che la diffusione della disinformazione o di altri contenuti illeciti o nocivi può generare.
"I prestatori di servizi intermediari includono nelle loro condizioni generali informazioni sulle restrizioni che impongono in relazione all'uso dei loro servizi per quanto riguarda le informazioni fornite dai destinatari del servizio. Tali informazioni riguardano tra l'altro le politiche, le procedure, le misure e gli strumenti utilizzati ai fini della moderazione dei contenuti, compresi il processo decisionale algoritmico e la verifica umana, nonché le regole procedurali del loro sistema interno di gestione dei reclami. Sono redatte in un linguaggio chiaro, semplice, comprensibile, facilmente

disposizioni principali, spiccano gli obblighi di notifica per la rimozione dei contenuti, l'introduzione di misure di trasparenza pubblicitaria e la possibilità per gli utenti di presentare ricorsi contro decisioni di moderazione[43]. Inoltre, le piattaforme di grandi dimensioni sono obbligate a designare ufficiali per la conformità e a sottoporsi ad *audit* indipendenti, affinché le loro politiche siano effettivamente in linea con la tutela dei diritti fondamentali degli utenti[44].

L'evoluzione da una co-regolamentazione ascendente a una discendente segna un cambiamento fondamentale nella gestione della disinformazione *online*[45]. Sebbene il Codice di buone pratiche inizialmente adottato abbia rappresentato un passo importante verso l'autoregolamentazione del settore privato, è emerso che un approccio completamente volontario non è sufficiente per affrontare i rischi sistemici della disinformazione, soprattutto alla luce delle capacità sempre più sofisticate dei sistemi IA, che amplificano e diffondono contenuti manipolati[46].

Nonostante il DSA introduca strumenti significativi per combattere la disinformazione, vi sono preoccupazioni riguardo alla possibilità che il meccanismo di risposta alle crisi porti alla rimozione indiscriminata di contenuti, creando così un conflitto tra la necessità di fermare la disinformazione e la protezione dei diritti fondamentali[47]. Con il DSA, l'Unione europea ha deciso di stabilire un quadro giuridico più stringente, grazie al quale le piattaforme sono obbligate a garantire misure di trasparenza e responsabilità, mettendo in atto politiche che non solo limitano i contenuti disinformativi, ma

---

fruibile... I prestatori di servizi intermediari agiscono in modo diligente, obiettivo e proporzionato … tenendo debitamente conto dei diritti e degli interessi legittimi di tutte le parti coinvolte, compresi i diritti fondamentali dei destinatari del servizio...” (art.14) “ Qualora venga a conoscenza di informazioni che fanno sospettare che sia stato commesso, si stia commettendo o probabilmente sarà commesso un reato che comporta una minaccia per la vita o la sicurezza di una o più persone, il prestatore di servizi di memorizzazione di informazioni informa senza indugio le autorità giudiziarie o di contrasto dello Stato membro o degli Stati membri interessati in merito ai propri sospetti, fornendo tutte le informazioni pertinenti disponibili. Se non è in grado di individuare con ragionevole certezza lo Stato membro interessato… ne informa Europol o le autorità di contrasto...” (art.15) “I fornitori di servizi di *hosting* devono predisporre meccanismi che consentano a qualsiasi individuo o entità di notificare loro la presenza sul loro servizio di elementi specifici di informazione che l'individuo o l'entità ritiene essere contenuti illegali. Tali meccanismi devono essere di facile accesso e intuitivi e devono consentire l'invio di notifiche esclusivamente tramite mezzi elettronici…” (art.16) “I fornitori di servizi di hosting forniscono una chiara e specifica dichiarazione dei motivi a tutti i destinatari interessati del servizio per una qualsiasi delle seguenti restrizioni imposte sulla base del fatto che le informazioni fornite dal destinatario del servizio sono contenuti illegali o incompatibili con i loro termini e condizioni…” (art.17).

[43] Il Capo III del regolamento disciplina in maniera dettagliata una serie di obblighi di *due diligence*, in particolare relativi alle attività di *private enforcement*, attraverso un sistema normativo strutturato su differenti "livelli" di intensità crescente, in funzione del destinatario. Il regime parte da un livello "base", applicabile a tutti i prestatori di servizi intermediari, per arrivare all'ultimo "gradino", che stabilisce le disposizioni più onerose per le piattaforme *online* e i motori di ricerca di "dimensioni molto grandi". In dottrina, tale struttura è stata correttamente interpretata come un approccio di tipo "*pyramid base*". Il passaggio a ciascun livello superiore implica l'imposizione di obblighi aggiuntivi, che si sovrappongono a quelli previsti nei livelli precedenti, senza sostituirli. V. M.L. Bixio, *Gli obblighi applicabili a tutti i prestatori di servizi intermediari, ai prestatori di servizi di hosting e ai fornitori di piattaforme online (Artt. 11-32 – Capo III, Sezioni, 1, 2, 3 e 4)*, in *Diritto di internet* 1 (2023).

[44] Introdurre procedure dettagliate per la moderazione dei contenuti e la gestione dei reclami, applicabili a tutti i prestatori di servizi intermediari indipendentemente dal mercato di riferimento, dal tipo di attività o dalla dimensione, secondo un modello uniforme per tutti, avrebbe comportato il rischio di imporre oneri eccessivi e non necessari. Inoltre, il richiamo esplicito alla libertà di espressione e al pluralismo dei media risulta particolarmente rilevante, soprattutto nel contesto della lotta contro la disinformazione, poiché evidenzia la necessità per gli operatori di adottare un approccio rispettoso dei diritti fondamentali nell'elaborazione e nell'applicazione delle politiche, che potrebbero avere un impatto significativo su tali diritti e generare effetti indesiderati. D'altro canto, pur senza giustificare un inasprimento burocratico, sarebbe stato utile fornire ulteriori specifiche riguardo ai "diritti di garanzia" minimi dell'utente, in relazione alle misure che la piattaforma può adottare e che potrebbero incidere sui diritti fondamentali, come la libertà di espressione. Per un approfondimento V. E. BIRRITTERI, op. cit.

[45] A. Shaji George, A. S. Hovan George, “*Deepfakes: The Evolution of Hyper realistic Media Manipulation” Manipulation”, in Partners Universal Innovative Research Publication* (PUIRP) 1(2)(2023) 59-73.

[46] E. Chiti, B. Marchetti, *Divergenti? Le strategie di Unione europea e Stati Uniti in materia di Intelligenza artificiale*, in *Rivista della Regolazione dei mercati* 1(2020).

[47] D. Buijs, I. Buri, *The DSA's crisis approach: crisis response mechanism and crisis protocols* in *DSA Observatory* (2023).

promuovono anche l'alfabetizzazione mediatica e la protezione dei diritti fondamentali degli utenti[48]. Approccio che, sebbene rigoroso, non rinuncia a garantire che gli utenti mantengano un controllo sui propri dati, in linea con il principio di autonomia e libertà di espressione, che sono alla base della Costituzione digitale europea[49]. In tal senso, l'UE ha invitato a non delegare la responsabilità di determinare la veridicità o la qualità dei contenuti esclusivamente ai privati, riconoscendo i rischi connessi alla censura, senza però escludere la possibilità di moderare la disinformazione attraverso il supporto di strumenti esterni come i *fact-checkers* e gli esperti indipendenti[50].

3. - *Deepfake* e IA: Le sfide del futuro per la veridicità dell'informazione

La creazione di *fake news* può essere agevolata dall'IA: i *deepfake*, infatti, basati su algoritmi di IA che permettono di creare contenuti audio-visivi altamente realistici ma falsi, rappresentano una vera e propria svolta nella "guerra dell'informazione"[51]. Se le *fake news* si concentrano sulla manipolazione di testi e informazioni, i *deepfake* rappresentano una minaccia ancora più sofisticata[52]. La potenza di questi strumenti risiede nella credibilità immediata che i contenuti video e audio generati hanno sul pubblico. Ciò aumenta notevolmente la portata d'azione delle notizie false, compromettendo tra l'altro la fiducia della società civile verso il momento cognitivo delle informazioni[53]. Tant'è che, inizialmente utilizzati in ambito di intrattenimento e creatività, i *deepfake* stanno rapidamente assumendo una dimensione di minaccia alla sicurezza, con applicazioni in diversi ambiti criminali[54].

L'aspetto criminogeno dei *deepfake* è particolarmente preoccupante, poiché la difficoltà di distinguere tra realtà e finzione può ledere i diritti dei singoli, minare la fiducia pubblica e alimentare conflitti sociali. In questo scenario, l'evoluzione dell'Intelligenza artificiale pone nuove sfide non

---

[48] Il contrasto alla disinformazione non può essere un pretesto per sacrificare i diritti fondamentali dei cittadini. In particolare, la libertà di espressione deve essere preservata, evitando misure di moderazione troppo invasive che possano limitare la pluralità delle voci, ridurre la libertà di pensiero o portare a un controllo eccessivo da parte delle piattaforme. L'alfabetizzazione mediatica è uno strumento fondamentale per sensibilizzare i cittadini sulla necessità di distinguere le informazioni veritiere da quelle manipolate, promuovendo un utilizzo consapevole e critico delle risorse digitali. Inoltre, l'accesso alle informazioni accurate è essenziale per garantire il buon funzionamento della democrazia.
J. W. Seow, M. K. Lim et al., "*A comprehensive overview of Deepfake: Generation, detection, datasets, and opportunities"* in NEUROCOMPUTING 513 (2022).
V. Considerando *Digital Service Act*: (2) "Gli Stati membri stanno sempre più introducendo o stanno valutando di introdurre legislazioni nazionali sulle materie disciplinate dal presente regolamento, imponendo in particolare obblighi di diligenza per i prestatori di servizi intermediari per quanto riguarda il modo in cui dovrebbero contrastare i contenuti illegali, la disinformazione online e altri rischi per la società. Tenendo conto del carattere intrinsecamente transfrontaliero di *internet*…" (3) "Un comportamento responsabile e diligente da parte dei prestatori di servizi intermediari è essenziale per un ambiente *online* sicuro, prevedibile e affidabile e per consentire ai cittadini dell'Unione e ad altre persone di esercitare i loro diritti fondamentali garantiti dalla Carta dei diritti fondamentali dell'Unione europea («Carta»), in particolare la libertà di espressione e di informazione, la libertà di impresa, il diritto alla non discriminazione e il conseguimento di un livello elevato di protezione dei consumatori."

[49] "*Il Digital Services Act (dsa*) è stato definito la nuova costituzione digitale dell'Unione europea, diretto a rafforzare i diritti e creare un ambiente *online* più sano, sicuro e trasparente." cfr. L. Bolognini; E. Pelino, *Digital Services Act e Digital Markets Act: Definizioni e Prime Applicazioni dei Nuovi Regolamenti Europei*, 2023.

[50] Si veda la relazione introduttiva al *Digital Services Act*.

[51] Secondo F. BERTONI, *Deepfake* cit. A differenza di altre forme di manipolazione digitale, come le *fake news*, i *deepfake* si caratterizzano per essere contenuti audiovisivi in grado di creare una rappresentazione falsificata di un soggetto. Non si limitano, quindi, a diffondere informazioni false, ma costruiscono una falsa rappresentazione visiva o sonora che simula eventi o dichiarazioni mai avvenuti. La componente audiovisiva dei *deepfake* li rende particolarmente insidiosi, poiché comporta un livello di autenticità che i testi o le immagini statiche non possono offrire.

[52] F.Marone, *L'evoluzione della disinformazione come minaccia ibrida* in *Osservatorio Strategico* 43 (2022). Per un approfondimento V. A. Gaeta, V. Loia et al., *Information disorder, cognitive weapons e crimini transnazionali: Scenari, Sfide E Possibili Soluzioni* in A. Oriolo, A. R. Castaldo, A. Di Stasi e M. Nino(cur.) *Criminalità transnazionale e Unione europea*, 2024.

[53] M. Filimowicz, *Deep Fakes: Algorithms and Society*, 2022.

[54] S. Fontana, *I deepfake stanno diventando una cosa seria* in *Facta News* (2023).

solo per il progresso tecnologico, ma anche per la protezione dei diritti fondamentali dell'individuo e per la necessità di una regolamentazione giuridica che sappia affrontare i pericoli derivanti dall'abuso di tali tecnologie. Segnatamente, riveste particolare rilievo il tema della tutela del diritto d'autore, che protegge sia i diritti morali sia quelli patrimoniali degli autori su opere dell'ingegno, nonché i diritti connessi, come il diritto all'immagine. Oltre al diritto d'autore, l'utilizzo illecito dei *deepfake* solleva ulteriori questioni giuridiche legate alla protezione dei dati personali, alla diffamazione e alla manipolazione dell'informazione[55]. Tuttavia, nonostante la percezione allarmistica, l'impatto dei *deepfake* sui diritti individuali non sembra avere, almeno per ora, la stessa gravità che caratterizza altri sistemi di IA, come le tecnologie di riconoscimento biometrico[56]. Ciononostante, la difficoltà nel distinguere i *deepfake* dalla realtà pone serie sfide non solo per il sistema giuridico, ma anche per l'affidabilità delle relazioni interpersonali e istituzionali[57].

Nel tentativo di fornire una definizione del fenomeno è possibile considerare che, come noto, il concetto di *deepfake* è di recente elaborazione e rappresenta una delle applicazioni più sofisticate e controverse dell'Intelligenza artificiale[58]. Il termine è composto dai vocaboli anglosassoni *deep*, che richiama il *deep learning*, ovvero un sistema di apprendimento automatico basato su reti neurali progettate per simulare i processi cognitivi umani, e *fake*, che significa falso[59]. I *deepfake*, frutto di tecniche avanzate di Intelligenza artificiale, permettono di creare contenuti falsificati che appaiono incredibilmente reali in modo da ingannare l'osservatore. In ambito tecnico-giuridico, con *deepfake* si intende sia la tecnologia che consente di creare contenuti digitali falsificati, sia il prodotto finale di tale processo: immagini, video o audio che appaiono incredibilmente realistici, ma sono del tutto artefatti[60]. Secondo la definizione fornita dal Garante per la protezione dei dati personali, "*I deepfake*

[55] V. An Observatory Report from the Europol Innovation Lab, *Facing reality? Law enforcement and the challenge of deepfakes,* 2022.

[56] Infatti, il Regolamento sull'Intelligenza artificiale, classifica i *deepfake* come tecnologie ad alto rischio ma non li pone allo stesso livello di pericolosità delle tecnologie di riconoscimento biometrico in tempo reale. Il ché riflette una valutazione più bilanciata del loro impatto attuale: mentre i *deepfake* hanno un potenziale significativo per causare danni, l'uso improprio di sistemi biometrici ha implicazioni più gravi e immediate per i diritti fondamentali, come la *privacy*, la libertà di movimento e la protezione contro la discriminazione.

[57] R. Chesney, D.K. Citron, *Deep Fakes*: *A Looming Challenge for Privacy, Democracy, and National Security* in *California Law Review* (2019).

[58] V. la Risoluzione del Parlamento europeo, del 20 gennaio 2021, sull'Intelligenza artificiale: *questioni relative all'interpretazione e applicazione del diritto internazionale nella misura in cui l'UE è interessata relativamente agli impieghi civili e militari e all'autorità dello Stato al di fuori dell'ambito della giustizia penale* (2020/2013(INI)). Al punto 76, il Parlamento "esprime profonda preoccupazione per le tecnologie di *deepfake*, che consentono di produrre foto, audio e video falsificati sempre più realistici che potrebbero essere utilizzati per compiere ricatti, creare notizie false o minare la fiducia dei cittadini e influenzare il dibattito pubblico; ritiene che tali pratiche siano in grado di destabilizzare paesi, diffondere la disinformazione e influenzare le consultazioni elettorali; chiede pertanto l'introduzione di un obbligo in base al quale tutti i materiali deepfake o altri video artificiali realizzati in modo realistico debbano essere etichettati come "non originali" dal loro creatore, con severi limiti al loro utilizzo a fini elettorali, e che tale obbligo sia applicato rigorosamente; chiede che siano svolte adeguate attività di ricerca in questo campo per garantire che le tecnologie di contrasto dei suddetti fenomeni siano al passo con gli utilizzi dolosi dell'IA".

[59] In dottrina, i *deepfake* vengono descritti come una tecnica che combina o sovrappone immagini e video esistenti con altre rappresentazioni, oppure come uno strumento in grado di generare contenuti totalmente nuovi, privi di un riferimento reale, ma difficilmente distinguibili dal vero.

[60] E. Chizzola, *Deepfake: tra furto d'identità, danno d'immagine e rischio privacy* in *Norme&Tributi Plus Diritto*, *Il Sole 24Ore* (2021).

Secondo l'art 3(60) del Regolamento sull'Intelligenza artificiale " '*deep fake' means AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthfu*.". Al fine di garantire la corretta informazione, il Regolamento stabilisce obblighi di trasparenza per i destinatari di sistemi di IA che generano o manipolano contenuti, come *deepfake* (immagini, video o audio manipolati), e testi, obbligandoli a dichiarare chiaramente che il contenuto è artificiale. Art.50 (4) ". *Deployers of an AI system that generates or manipulates image, audio or video content constituting a deep fake, shall disclose that the content has been artificially generated or manipulated. This obligation shall not apply where the use is authorised by law to detect, prevent, investigate or prosecute criminal offence. Where the content forms part of an evidently artistic, creative, satirical, fictional or analogous work or programme, the transparency obligations set out in this paragraph are limited to disclosure of the existence of such generated or manipulated content in an appropriate manner that does not hamper*

sono foto, video e audio creati grazie a software di Intelligenza artificiale ("AI") che, partendo da contenuti reali (immagini e audio), riescono a modificare o ricreare, in modo estremamente realistico, le caratteristiche e i movimenti di un volto o di un corpo e a imitare fedelmente una determinata voce"[61]. Questa definizione fornisce un riferimento chiaro per comprendere sia gli aspetti tecnici sia le implicazioni giuridiche dei *deepfake*, mettendo in luce gli elementi fondamentali che ne caratterizzano la natura e le criticità. Un'analisi approfondita di tali elementi, unitamente al ruolo giocato dall'IA nel processo di formazione dei medesimi, consente di esaminare in modo critico le conseguenze di questa tecnologia, soprattutto in relazione alla regolamentazione normativa e alla salvaguardia dei diritti fondamentali[62].

La produzione di *deepfake* è strettamente connessa all'utilizzo di *software* di Intelligenza artificiale che hanno come finalità principale quella di modificare o ricreare contenuti audiovisivi in modo da alterare significativamente la realtà rappresentata[63]. I *software* utilizzati si distinguono per la loro capacità di elaborare enormi quantità di dati (foto, video o audio) per generare contenuti manipolati altamente realistici. Un processo non comparabile a semplici tecniche di fotoritocco o montaggio audiovisivo, poiché l'IA consente una manipolazione biometrica avanzata, che coinvolge caratteristiche facciali, movimenti del corpo e modulazioni vocali in modo dettagliato e credibile[64]. La natura tecnica dei *deepfake* rende complessa la loro individuazione e crea una zona grigia normativa. Non sempre è chiaro quando una manipolazione audiovisiva debba essere considerata una violazione del diritto all'immagine o un contenuto legittimamente creato per scopi artistici, satirici o educativi[65].

La creazione dei *deepfake* presuppone l'utilizzo di contenuti reali come base di partenza. Foto, video o registrazioni audio autentiche vengono elaborate dai *software* di IA per generare *output* falsificati. La necessità di un *input* reale è una caratteristica distintiva dei *deepfake*, poiché l'obiettivo principale di questa tecnologia è creare contenuti falsi che mantengano tratti biometrici riconoscibili, aumentando così il loro effetto ingannevole[66]. Pertanto, da un punto di vista giuridico, l'utilizzo di contenuti reali solleva diverse problematiche, tra cui, a monte, la violazione del diritto alla *privacy*, poiché il trattamento non autorizzato di immagini o registrazioni personali costituisce una lesione della riservatezza dell'individuo[67]. In secondo luogo, sebbene la mera creazione di contenuti *deepfake* non integri di per sé un reato, è la finalità del loro utilizzo a determinare l'eventuale rilevanza penale o civile della condotta. Laddove essi siano utilizzati per finalità fraudolente, diffamatorie o estorsive, si rende necessario un rafforzamento delle normative esistenti, accompagnato dallo sviluppo di nuove

the display or enjoyment of the work. Deployers of an AI system that generates or manipulates text which is published with the purpose of informing the public on matters of public interest shall disclose that the text has been artificially generated or manipulated. This obligation shall not apply where the use is authorised by law to detect, prevent, investigate or prosecute criminal offences or where the AI-generated content has undergone a process of human review or editorial control and where a natural or legal person holds editorial responsibility for the publication of the content.".

[61] V *Deepfake: dal Garante una scheda informativa sui rischi dell'uso malevolo di questa nuova tecnologia, 2020.*

[62] E. Meskys et al, *Regulating Deep Fakes: Legal and Ethical Considerations* in *Journal of Intellectual Property Law & Practice* 15(1)(2020) 24–31.

[63] Y. Mirsky, *The Creation and Detection of Deepfakes: A Survey* in ACM Computing Surveys (CSUR) 54(1) (2021) 1-41.

[64] T. T. Nguyen, *Deep learning for deepfakes creation and detection: A survey Author links open overlay panel* in *Computer Vision and Image Understanding* 223 (2022).

[65] Tra le tipologie più diffuse di *deepfake* troviamo: video falsificati di tipo satirico o umoristico, che spesso utilizzano volti di personaggi famosi per scopi di intrattenimento; interviste o dichiarazioni fake, sfruttate come strumenti di disinformazione o propaganda politica; *deep porn*, contenuti pornografici falsificati che rappresentano una grave violazione della dignità personale, specialmente quando realizzati senza il consenso del soggetto ritratto. *Il deep porn* è particolarmente lesivo sotto il profilo dei diritti della persona, infatti, è considerato una forma di violenza digitale che può costituire un reato, come quello di *revenge porn* o di diffusione illecita di materiale sensibile, punito da normative nazionali e internazionali.

[66] D.L. Suying, *Deep fakes and Disinformation in Asia,* in Deep Fakes. Algorithms and Society, New York (2022).

[67] E. Chizzola, *Deepfake* cit.

strategie legislative e regolatorie idonee a contrastare tali fenomeni in maniera efficace e proporzionata[68].

La natura intrinsecamente transfrontaliera dei *deepfake* e l'interconnessione globale dei *media* digitali sottolineano l'opportunità di adottare un approccio integrato e olistico, piuttosto che affidarsi unicamente a iniziative legislative nazionali.È quindi indispensabile promuovere una collaborazione internazionale che armonizzi gli sforzi regolatori, favorisca lo scambio di competenze tecniche e giuridiche e crei *standard* condivisi per la gestione di questa tecnologia. La collaborazione dovrebbe prevedere anche meccanismi di *enforcement* che permettano un'azione coordinata contro gli illeciti, evitando vuoti normativi che potrebbero essere sfruttati per eludere la giustizia. Sotto la lente del giurista, si evidenzia la necessità di distinguere i *deepfake* da altre forme di manipolazione digitale, formulando definizioni precise e puntuali che consentano di colmare eventuali lacune normative[69]. Inoltre, il quadro giuridico dovrebbe prevedere responsabilità non solo per gli utilizzatori illeciti di *deepfake*, ma anche per gli sviluppatori e i distributori di strumenti tecnologici che, se non adeguatamente regolamentati, possono essere impiegati per finalità contrarie al diritto. In tal senso, l'imposizione di obblighi di trasparenza, sicurezza e conformità ai fornitori di tecnologia rappresenta una misura fondamentale per prevenire abusi[70].

A tal riguardo, nonostante l'Unione europea stia compiendo passi significativi con iniziative come *l'Artificial Intelligence Act*, è evidente che il contesto globale richiede un approccio armonizzato e interdisciplinare. La creazione di *standard* comuni, il rafforzamento della cooperazione internazionale e l'investimento in tecnologie di rilevamento avanzate sono essenziali per affrontare i rischi legati ai *deepfake*, proteggendo al contempo i principi democratici e la dignità umana. Parallelamente, una maggiore consapevolezza pubblica e la formazione di professionisti qualificati nel campo giuridico e tecnologico saranno fondamentali per contrastare in modo efficace le potenziali minacce. Tuttavia, il successo di queste iniziative, di seguito analizzate, dipende dalla capacità di integrare la regolamentazione con soluzioni tecnologiche, promuovendo al contempo un uso responsabile dell'Intelligenza artificiale. Solo attraverso un impegno globale e collaborativo sarà

[68] Si pensi alla Direttiva 2011/93/UE del Parlamento e del Consiglio dell'Unione europea che ha come obiettivo primario la protezione dei minori da abusi sessuali e sfruttamento, con particolare attenzione alla pornografia minorile e alle relative attività criminali. La direttiva stabilisce un quadro giuridico che impone agli Stati membri dell'UE di adottare misure efficaci per combattere la produzione, distribuzione e il possesso di contenuti pornografici che coinvolgono minori, sia in contesti fisici che *online*. Nel contesto digitale, la direttiva stabilisce chiaramente che i fornitori di servizi *online*, come piattaforme di social media, motori di ricerca, servizi di *hosting* e di condivisione di contenuti, devono essere obbligati a rimuovere tempestivamente i contenuti che presentano pornografia infantile. Questi fornitori sono inoltre tenuti a implementare meccanismi di monitoraggio per rilevare e rimuovere i contenuti illeciti, nonché ad adottare procedure per identificare chi carica il materiale dannoso. Con riguardo alle sanzioni penali, essa stabilisce pene severe per chiunque produca, distribuisca, possieda o faciliti la diffusione di materiale pornografico minorile. Le sanzioni devono essere efficaci, proporzionate e dissuasive, con pene detentive che variano in base alla gravità del crimine. Inoltre, anche coloro che gestiscono piattaforme online o servizi digitali, e che non adempiono agli obblighi di rimozione dei contenuti illeciti o di protezione dei minori, sono soggetti a sanzioni legali. La Direttiva stabilisce, pertanto, una base solida per l'azione legale nei confronti dei trasgressori, in un'ottica di protezione dei diritti e della sicurezza dei minori. Tuttavia, la Direttiva 2011/93/UE non affronta esplicitamente il fenomeno emergente dei *deepfake*, il cui uso per creare pornografia minorile virtuale si sta diffondendo rapidamente, violando non solo i diritti dei minori coinvolti, ma anche danneggiando irreparabilmente la loro dignità e integrità psicologica. Sebbene non espressamente regolamentati dalla Direttiva 2011/93/UE, è possibile che questi crimini, possano essere inclusi nell'ambito di applicazione della direttiva. Le azioni legali contro la produzione e la distribuzione di pornografia minorile, inclusa quella generata tramite tecnologie come il deep fake, potrebbero essere perseguite in virtù delle leggi già in vigore. Pertanto, si potrebbe auspicare un ampliamento della direttiva per includere esplicitamente l'uso delle tecnologie di manipolazione digitale come il *deep fake*, con l'introduzione di pene specifiche per chi creasse o diffondesse contenuti pornografici minorili generati da queste tecniche.

[69] Id.

[70] D. Della Rosa, F. Criscuoli, *AI Act, pratiche vietate e regole per i sistemi ad alto rischio nel segno della trasparenza e sicurezza* in NT Diritto (2023).

Nell'antico Egitto, la *damnatio memoriae* rappresentava una pratica deliberata per riscrivere la storia, eliminando il ricordo di individui ritenuti scomodi dal potere dominante. La manipolazione della memoria storica non era solo un tentativo di esercitare controllo sul passato, ma un mezzo per consolidare l'autorità nel presente, dimostrando il legame intrinseco tra controllo dell'informazione e potere politico.

possibile garantire che i progressi tecnologici non compromettano la fiducia e la stabilità delle società democratiche, ma anzi contribuiscano a un futuro digitale sicuro ed equo.

4. - Il ruolo dell'Unione europea nella gestione dei contenuti manipolati

A livello legislativo, l'Unione europea ha avviato azioni mirate per mitigare i rischi legati all'Intelligenza artificiale. Nonostante non disponga ancora di una normativa specifica sui *deepfake*, sta riconoscendo progressivamente le minacce associate a tali tecnologie[71]. Il Parlamento europeo ha mostrato grande interesse nell'affrontare le problematiche connesse ai *deepfake*, riconoscendone tanto i potenziali benefici quanto i pericoli, soprattutto nel contesto della disinformazione e della propaganda politica[72]. In diverse raccomandazioni e risoluzioni, il Parlamento ha sottolineato la necessità di un approccio normativo che protegga i cittadini dai danni derivanti dall'uso improprio di queste tecnologie. Nel 2022, la Commissione europea ha elaborato lo *Strengthened Code of Practice on Disinformation*, in particolare per affrontare le sfide poste dalla pandemia di COVID-19[73]. Nel documento si pone l'accento sulle minacce emergenti e si raccomandano misure efficaci per combattere la manipolazione delle informazioni *online*, con l'obiettivo di limitare l'impatto negativo della disinformazione nel contesto digitale[74].

In particolare, il Codice menziona esplicitamente i *deepfake* come uno degli strumenti di manipolazione da contrastare, con l'impegno delle piattaforme a sviluppare politiche per limitare la diffusione di contenuti falsificati. Nonostante questi sforzi, dai *report* pubblicati nel 2023 emerge che le piattaforme principali non hanno adottato modifiche sostanziali alle loro politiche, continuando a fare riferimento alle normative già esistenti[75]. Sebbene questi interventi siano un primo passo, resta molto da fare per garantire una gestione adeguata e uniforme del fenomeno a livello europeo[76].

A differenza di altre forme di manipolazione informativa, i *deepfake* si distinguono per la loro capacità di ingannare e influenzare in maniera diretta l'opinione pubblica, grazie alla loro apparente autenticità[77]. La persuasività di un contenuto video manipolato è infatti notevolmente maggiore rispetto ad altri tipi di notizie false, poiché si associa alla familiarità visiva e alla riproduzione di situazioni che sembrano autentiche, ma che in realtà sono artificiali. Questa caratteristica rende giustificabile, almeno in via teorica, un intervento normativo più severo rispetto ad altre pratiche di disinformazione[78]. Tuttavia, il rischio di introdurre leggi troppo rigide o penalizzanti, che interferiscano con la libertà di espressione o con il diritto alla creatività, resta sempre presente.

---

[71] In Italia manca ancora una legislazione specifica che regolamenti il fenomeno dei *deepfake*, e non esiste una normativa complessiva sulla disinformazione, probabilmente anche a causa dell'assenza di casi eclatanti in cui i *deepfake* abbiano coinvolto figure pubbliche italiane. Al contrario, negli Stati Uniti, tra il 2018 e il 2021, alcuni Stati hanno adottato leggi mirate per regolare questo fenomeno, portando l'attenzione su tale problema a livello nazionale.

[72] V. EPRS | EUROPEAN PARLIAMENTARY RESEARCH SERVICE, *Tackling deepfakes in European policy*,2021.

[73] Il Codice di buone pratiche sulla disinformazione 2022 è un aggiornamento delle linee guida introdotte dall'UE nel 2018 per contrastare la diffusione di informazioni false *online*. Il codice, sviluppato con la partecipazione di piattaforme digitali, aziende tecnologiche e operatori pubblicitari, mira a migliorare la trasparenza e la responsabilità nella gestione della disinformazione, con particolare attenzione alla pubblicità politica e alla manipolazione delle opinioni pubbliche.Rispetto alla versione precedente, il Codice 2022 introduce obblighi più stringenti, come la dichiarazione chiara dei finanziamenti per gli annunci e la cooperazione con ricercatori e *fact-checkers* indipendenti. Si integra inoltre con il *Digital Services Act*, garantendo un approccio più strutturato e misurabile nella lotta contro la disinformazione, senza sacrificare la libertà di espressione.

[74] F. Ramos-Zaga, *Deepfake: Analisi delle sue implicazioni tecnologiche e giuridiche nell'era dell'Intelligenza artificiale*, 2022.

[75] *Meta*, ad esempio, ha avviato uno studio con l'Università del Michigan per sviluppare strumenti di *reverse engineering* volti a rilevare contenuti manipolati, ma non ha annunciato cambiamenti significativi nelle proprie politiche. Situazione simile per *Twitter* e *Google*, che non hanno apportato modifiche sostanziali alle loro linee guida.

[76] C. Schepisi*, Le "dimensioni" della regolazione dell'Intelligenza artificiale nella proposta di regolamento della Commissione* 2022*,* 330-356.

[77] F. Bertoni, *Deepfake* cit.

[78] L'introduzione di nuove fattispecie penali per sanzionare la creazione e la diffusione di deepfake con finalità disinformative è una proposta che ha suscitato ampio dibattito giuridico. Alcuni Stati, come il Texas, hanno cercato di

5. - L'*AI Act* e la nuova *governance* europea dell'Intelligenza artificiale

L'avvento dell'Intelligenza artificiale generativa ha sollevato interrogativi rilevanti a livello europeo, portando l'Unione europea a intervenire in modo normativo, con l'adozione dell'*Artificial Intelligence Act*, una legislazione unica nel panorama mondiale. Come anticipato, il dibattito giuridico degli ultimi anni ha messo in luce la necessità di una regolamentazione dell'IA non solo per tutelare i diritti individuali, ma anche per garantire la sovranità digitale europea e favorire l'espansione del mercato unico, principio fondante della stessa Unione europea[79]. L'*AI Act*, nella sua articolazione, ha proprio come obiettivo quello di migliorare il funzionamento del mercato interno, stabilendo un quadro giuridico uniforme per lo sviluppo, la commercializzazione, l'uso e l'immissione sul mercato dei sistemi di IA nell'Unione europea[80]. A tal fine, si impegna a promuovere un'IA "antropocentrica" e affidabile, capace di rispettare i valori dell'Unione, in particolare quelli legati alla salute, sicurezza, e ai diritti fondamentali come sancito dalla Carta dei diritti fondamentali dell'UE[81]. Il regolamento si pone quindi in funzione di tutela contro gli effetti nocivi dei sistemi di IA, pur sostenendo l'innovazione tecnologica come motore di sviluppo economico.

Approvato con un ampio consenso – 499 voti favorevoli, 28 contrari e 93 astensioni – il Regolamento rappresenta un passo decisivo verso l'adozione di una normativa armonizzata per la regolamentazione dell'Intelligenza artificiale nell'Unione europea[82]. Tra i diversi obiettivi stabiliti dall'atto, i principali includono: 1) garantire la sicurezza dei sistemi di IA immessi nel mercato europeo e il rispetto delle normative in materia di diritti fondamentali; 2) offrire certezza giuridica per favorire gli investimenti e stimolare la ricerca e l'innovazione; 3) rafforzare la *governance* e

affrontare direttamente il problema con leggi che criminalizzano i deepfake mirati a danneggiare l'immagine di candidati politici o a influenzare in maniera indebita i risultati elettorali. Tuttavia, esperti di diritto come Orin Samuel Kerr evidenziano che l'introduzione di norme penali in questo ambito potrebbe essere problematica, poiché rischia di sfociare in un'applicazione arbitraria della legge, soprattutto in un contesto tanto dinamico e complesso come quello della manipolazione digitale.

[79] In vari interventi, tra cui il discorso sullo *State of the Union* del 2020, Ursula von der Leyen, Presidente della Commissione europea, ha evidenziato l'importanza di una "sovranità digitale" per proteggere i dati europei e definire standard globali sull'IA, in contrapposizione alla crescente influenza normativa di paesi come Stati Uniti e Cina. Ha ribadito che la regolamentazione dell'IA non è solo una questione di diritti fondamentali, ma anche di competitività economica e indipendenza tecnologica. V. https://state-of-the-union.ec.europa.eu/document/download/e84f2747-e1b9-4bd8-9a411f5232de6812_it?filename=SOTEU%202023_Brochure_IT.pdf

[80] A. ADINOLFI, *L'Unione europea dinanzi allo sviluppo dell'Intelligenza artificiale: la costruzione di uno schema di regolamentazione europeo tra mercato unico digitale e tutela dei diritti fondamentali* in IL RAGIONAMENTO GIURIDICO NELL'ERA DELL'INTELLIGENZA ARTIFICIALE, S. DORIGO (a cura di), 2020, pp. 321-343.

[81] A. DI STASI, A. IERMANO, A. ORIOLO ET AL., *Spazio europeo di giustizia e applicazione giurisprudenziale del titolo VI della Carta dei diritti fondamentali dell'Unione europea*, 2024, pp. 1-356.

La Carta dei diritti fondamentali dell'Unione europea (Carta di Nizza) rappresenta un pilastro fondamentale per la protezione dei diritti e delle libertà dei cittadini europei, sancendo diritti che devono essere rispettati in ogni ambito normativo dell'Unione, incluso il diritto penale. In tale contesto, l'armonizzazione delle legislazioni penali degli Stati membri è cruciale per garantire un'efficace protezione dei diritti fondamentali, evitando fenomeni di *ne bis in idem*, ossia la possibilità di essere giudicati due volte per lo stesso reato. L'armonizzazione del diritto penale nell'UE ha come obiettivo il rafforzamento della cooperazione giuridica tra gli Stati membri, assicurando che la protezione dei diritti dei cittadini sia uniforme e che le decisioni giuridiche siano reciprocamente riconosciute e rispettate. In tale ambito, il principio di *ne bis in idem* assume un'importanza fondamentale, garantendo che nessuna persona possa essere perseguita o punita più di una volta per lo stesso crimine, evitando situazioni di doppia incriminazione che ledano il diritto alla giustizia. A tal fine, l'Unione europea si impegna a promuovere lo sviluppo di Intelligenza artificiale (IA) in un'ottica antropocentrica, ossia centrata sull'essere umano e sul rispetto dei suoi diritti fondamentali.

[82] Il 21 aprile 2021 la Commissione europea ha presentato una proposta di regolamento, con l'obiettivo di creare un quadro normativo unificato per la gestione dell'IA. Successivamente, il 9 dicembre 2023, è stato raggiunto un accordo politico tra il Parlamento e il Consiglio dell'Unione europea, con l'approvazione definitiva del testo del regolamento avvenuta il 2 febbraio 2024.

assicurare un'effettiva applicazione delle normative sui sistemi di IA; 4) creare un mercato unico per applicazioni di IA sicure, affidabili e lecite, evitando la frammentazione normativa tra Stati membri.

Si segnala che il legislatore europeo ha scelto come base giuridica l'art. 114 del Trattato sul funzionamento dell'Unione europea (TFUE), che consente di adottare misure per l'armonizzazione delle legislazioni degli Stati membri e consente l'adozione di misure volte a garantire il funzionamento del mercato interno. Vieppiù, la scelta di adottare un regolamento anziché una direttiva si fonda sulla necessità di creare un quadro normativo uniforme in tutta l'Unione, evitando che diverse normative nazionali possano ostacolare l'innovazione tecnologica e lo sviluppo di tale settore[83]. Allo stesso tempo, il regolamento lascia margini di flessibilità agli Stati membri, soprattutto per quanto concerne l'organizzazione interna e il regime sanzionatorio, favorendo l'adattamento a specificità nazionali[84]. Ciò è particolarmente rilevante in un contesto caratterizzato dall'evoluzione rapida della tecnologia IA, alimentata dai veloci sviluppi nel campo del *machine learning* e del *deep learning*, che conferiscono ai sistemi un'autonomia crescente.

Il regolamento incarna una visione profondamente europea dello sviluppo tecnologico, basata sulla priorità della dignità umana e sul rispetto dei diritti fondamentali, in contrapposizione a modelli alternativi che privilegiano la crescita economica a controllo statale o la libertà di impresa senza vincoli etici[85]. L'IA viene dunque inquadrata come uno strumento da sviluppare e utilizzare nel rispetto dei principi di sicurezza, affidabilità e trasparenza, al fine di promuovere un'innovazione che non comprometta la protezione dei cittadini.

La struttura dell'*AI act* si basa su un modello regolatorio che si concentra sulla gestione dei rischi, ponendo al centro la protezione dei diritti fondamentali e la tutela del processo democratico.

Il Titolo I dell'*AI act* stabilisce i principi fondamentali del Regolamento, delineando il suo oggetto e il campo di applicazione. Le norme introdotte non si limitano a disciplinare i fornitori e gli utenti di Intelligenza artificiale con sede nell'Unione europea, ma si estendono anche agli operatori di paesi terzi, qualora il loro *output* venga utilizzato nel territorio dell'Unione. L'utilizzo di un approccio globale rispecchia l'intento del legislatore europeo di garantire un'efficace tutela dei diritti fondamentali e di promuovere un mercato unico digitale in grado di affrontare le sfide poste dall'IA, a prescindere dalla localizzazione geografica del sistema sviluppato[86]. Nel contesto di questa regolamentazione, l'Unione europea si propone come pioniere nella legislazione sull'IA a livello globale. Con l'adozione dell'AI *act*, l'Unione europea non solo stabilisce uno *standard* interno di riferimento, ma diventa anche un punto di riferimento per altre giurisdizioni che intendono affrontare

[83] Un regolamento ha la caratteristica di essere direttamente applicabile in tutti gli Stati membri senza necessità di recepimento, il che contribuisce a ridurre la frammentazione giuridica e a promuovere un mercato unico, eliminando ostacoli legali che potrebbero altrimenti limitare la libera circolazione di queste tecnologie. Questa base giuridica offre una cornice uniforme e vincolante per gli Stati membri, assicurando che la regolamentazione sull'IA venga applicata in modo omogeneo in tutta l'Unione. Per un commento, cfr. C. Schepisi, *Le "dimensioni*" cit.; A. Alaimo*, Il Regolamento sull'Intelligenza artificiale: dalla proposta della Commissione al testo approvato dal Parlamento. Ha ancora senso il pensiero pessimistico?* in *federalismi.it* (2023) 133-149.

[84] Tra i principali strumenti di adattamento troviamo gli allegati tecnici modificabili dalla Commissione europea, come quelli che definiscono i dispositivi ad alto rischio e le relative procedure di verifica di conformità, i quali possono essere aggiornati attraverso un processo più rapido e meno burocratico rispetto alla revisione legislativa ordinaria. La Commissione, assistita da comitati *ad hoc*, può apportare modifiche per rispondere a esigenze specifiche o risolvere problematiche emergenti, garantendo così una maggiore dinamicità normativa. Inoltre, sono identificati meccanismi di valutazione periodica. L'articolo 84 della proposta prevede che, entro tre anni dall'entrata in vigore del regolamento, la Commissione debba presentare al Parlamento europeo e al Consiglio una relazione di valutazione e riesame. Successivamente, tali valutazioni dovranno essere effettuate ogni quattro anni e pubblicate, garantendo trasparenza e consentendo un monitoraggio continuo dell'efficacia e dell'adeguatezza del regolamento. I meccanismi di flessibilità rendono il quadro normativo *future-proof,* ovvero capace di adattarsi ai futuri sviluppi tecnologici senza dover ricorrere necessariamente a complesse revisioni legislative a detta di G. Finocchiaro, *La regolazione dell'Intelligenza artificiale* in *Rivista Trimestrale di diritto pubblico* 4(2022) 1085-1099.

[85] A. Alaimo, *Il regolamento* cit.

[86] A. Lavorgna, G. Suffia, *La nuova proposta europea per regolamentare i Sistemi di Intelligenza artificiale e la sua rilevanza nell'ambito della giustizia penale: un passo necessario, ma non sufficiente, nella giusta direzione* in *Diritto Penale Contemporaneo* (2021).

le sfide legate all'Intelligenza artificiale[87]. Gli altri ordinamenti, infatti, potrebbero seguire l'esempio dell'UE nella creazione di normative che equilibrano l'innovazione tecnologica con la protezione dei diritti fondamentali.

In questa prospettiva si inserisce la Convenzione quadro del Consiglio d'Europa sull'Intelligenza artificiale, aperta alla firma il 5 settembre 2024. rappresenta il primo trattato giuridicamente vincolante in ambito IA e si integra perfettamente con l'*AI Act* dell'Unione europea, ampliando e rafforzando gli sforzi per la protezione dei diritti umani, della democrazia e dello stato di diritto a livello globale[88]. Mentre l'*AI Act* stabilisce un quadro normativo per i Paesi membri dell'UE, la Convenzione estende tali principi a livello internazionale, mirando a colmare le lacune giuridiche esistenti in altri ordinamenti e a promuovere una regolamentazione che risponda alle sfide etiche, legali e sociali derivanti dall'uso dell'IA.

Questi strumenti giuridici, pur operando in ambiti giuridici diversi, perseguono una visione comune di equilibrio tra innovazione tecnologica e protezione dei diritti fondamentali. La Convenzione Quadro, infatti, richiede che tutte le attività legate ai sistemi di IA siano allineate ai principi universali di diritti umani, democrazia e stato di diritto, senza compromettere lo sviluppo tecnologico. L'approccio di neutralità tecnologica, che caratterizza la Convenzione, si affianca e supporta l'*AI Act*, poiché entrambi mirano a garantire che le tecnologie emergenti siano progettate, implementate e monitorate in modo tale da tutelare i diritti individuali e collettivi.

Inoltre, entrambe le iniziative sottolineano l'importanza della trasparenza e della responsabilità nelle decisioni automatizzate, e promuovono un sistema di monitoraggio e valutazione continua dei rischi derivanti dall'uso dell'IA[89]. La Convenzione introduce una gestione dei rischi che impone agli Stati di effettuare valutazioni periodiche sugli impatti dei sistemi di IA sui diritti umani, sulla democrazia e sullo stato di diritto. Sulla base di queste valutazioni, le autorità competenti devono adottare misure per prevenire o mitigare i rischi, inclusa la possibilità di introdurre divieti o moratorie su applicazioni specifiche di IA che possano risultare incompatibili con i diritti fondamentali.

La Convenzione si applica a tutte le attività legate all'IA, comprese quelle condotte da soggetti privati, ma esclude settori specifici come la difesa nazionale e la ricerca, pur stabilendo che anche in questi casi si deve rispettare i principi internazionali sui diritti umani e la democrazia.

Il trattato ha visto il coinvolgimento di una vasta gamma di attori, tra cui Stati membri, Stati osservatori, società civile, mondo accademico e organizzazioni internazionali, garantendo così che la normativa risponda a un ampio spettro di esigenze e preoccupazioni globali. L'inclusione di Paesi non membri del Consiglio d'Europa, come Canada, Giappone, Stati Uniti, Israele e l'Unione europea, ha arricchito il processo con diverse prospettive giuridiche, ma unite dalla necessità di proteggere i diritti umani in un contesto di rapide trasformazioni tecnologiche.

6. Sistemi ad alto rischio: *Deepfake* e regolamentazione preventiva in Europa

A causa della rapida evoluzione della tecnologia, della complessità del fenomeno e del naturale ritardo dell'intervento normativo, per anni è mancata una definizione chiara del concetto di "*deepfake*". L'A*I Act*, all'art. 3, par. 60, interviene a colmare questa lacuna, definendo un "*deepfake*" come: "contenuto di immagini, audio o video generato o manipolato dall'Intelligenza artificiale che ricorda persone, oggetti, luoghi, entità o eventi esistenti e che apparirebbe falsamente autentico o veritiero a una persona". Allo stesso modo, il considerando 134 – l'unico considerando del Regolamento che menziona esplicitamente i *deepfake* – riprende questa formulazione, stabilendo che: "Oltre alle soluzioni tecniche impiegate dai fornitori del sistema di intelligenza artificiale, gli operatori che utilizzano un sistema di intelligenza artificiale per generare o manipolare contenuti di

[87] A. Rociola , *Via libera dei Paesi europei alle norme sull'Intelligenza artificiale* in *La Repubblica* (2024).
[88] CM(2024)52-final, *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law,* 2024.
[89] A. Stiano, *Il c.d. risk-based approach nell'AI Act dell'Unione Europea e nella Convenzione-quadro sull'intelligenza artificiale del Consiglio d'Europa* in *Rivista di diritto internazionale* (2024) 1130-1136.

immagini, audio o video che presentano una somiglianza marcata a persone, oggetti, luoghi, entità o eventi esistenti e che potrebbero apparire falsamente a una persona come autentici o veritieri (*deepfake*), dovrebbero anche dichiarare in modo chiaro e distinto che il contenuto è stato creato o manipolato artificialmente, etichettando di conseguenza l'*output* dell'Intelligenza artificiale e rivelandone l'origine artificiale". Entrambe le formulazioni utilizzano l'espressione "assomiglia a [...] esistenti", suggerendo che, per essere qualificato come *deepfake*, il contenuto deve richiamare elementi realmente esistenti, come una persona specifica, un luogo riconoscibile o un evento concreto. Questa interpretazione riflette la concezione comunemente accettata del *deepfake*, inteso come manipolazione di elementi reali finalizzata a creare una rappresentazione ingannevolmente verosimile. Restano, tuttavia, ampi margini di incertezza, in quanto l'effettiva somiglianza tra contenuti generati artificialmente e soggetti reali può risultare difficile da delimitare, considerata la vastità e la diversità delle entità a cui un *deepfake* potrebbe richiamare[90].

Un'interpretazione letterale della norma, infatti, restringerebbe la portata dei contenuti considerati *deepfake*, aprendo la porta a possibili tentativi di aggirare le regole, soprattutto quelle relative agli obblighi di trasparenza previsti per chi utilizza determinati sistemi di Intelligenza artificiale, come indicato all'articolo 50(4) dell'*AI Act*. In questo scenario, i contenuti generati interamente dall'IA, privi di un corrispettivo reale, rischierebbero di restare esclusi dall'etichettatura come *deepfake*, aumentando il rischio di inganno e contribuendo alla diffusione della disinformazione.

Analizzare la regolamentazione dei *deepfake* prevista dall'*AI Act* significa considerare l'obiettivo primario della norma e i valori che hanno ispirato l'introduzione degli obblighi di trasparenza. Alla base di questo approccio c'è la volontà di tutelare i cittadini: la protezione delle persone è infatti uno dei principi cardine dell'*AI Act*, come sancito dall'articolo 1 e ribadito nel Considerando 1, dove si esplicitano gli obiettivi del Regolamento e i valori fondamentali su cui si fonda. L'UE mira a promuovere l'innovazione, lo sviluppo del mercato interno e l'armonizzazione normativa, sempre nel rispetto di diritti fondamentali e principi democratici.

L'interpretazione dei *deepfake* va dunque esaminata alla luce degli scopi normativi e dei valori fondamentali dell'UE. Risulta necessario adottare un approccio teleologico che non si limiti alla mera formulazione testuale, ma tenga conto anche dell'obiettivo ultimo di proteggere la società dall'inganno e dalla manipolazione, garantendo al contempo la libertà degli individui di compiere scelte informate o di allontanarsi da situazioni potenzialmente ingannevoli[91]. La regolamentazione dell'IA attraverso l'*AI act* si basa su un principio di analisi del rischio, che suddivide i sistemi di IA in base ai potenziali danni che possono causare. Il modello tiene conto sia della natura dei sistemi di IA, sia del contesto in cui vengono utilizzati, stabilendo regole proporzionali alla gravità del rischio[92]. Il bilanciamento tra i diritti fondamentali e la protezione dei soggetti vulnerabili rappresenta un

[90] Se il legislatore avesse voluto includere in modo più ampio tutti quei contenuti artificiali che risultano difficili da distinguere da quelli reali, sarebbe stato più efficace adottare una definizione basata sulla semplice somiglianza con elementi –come oggetti, luoghi, entità o eventi – senza necessariamente fare riferimento a soggetti specifici ed esistenti. Il fatto che questa strada non sia stata esplicitamente seguita suggerisce che, secondo l'*AI Act*, un *deepfake* debba riferirsi a soggetti o elementi reali e identificabili per rientrare nella definizione legale.

[91] V. M. Łabuz, *Un'interpretazione teleologica della definizione di DeepFake nell'Artificial Intelligence Act dell'UE: un approccio basato sullo scopo ai potenziali problemi con la parola "esistente",* 2025.

[92] I tre livelli di rischio individuati dal legislatore europeo sono: 1) rischio inaccettabile: pratiche e sistemi che comportano un rischio elevato per la sicurezza e i diritti fondamentali delle persone. In tali casi, la proposta di regolamento prevede il divieto assoluto di utilizzo di tali tecnologie, come nel caso del *social scoring* o dei sistemi di identificazione biometrica remota in tempo reale, con alcune eccezioni molto restrittive per specifiche applicazioni di ordine pubblico e sicurezza; 2) rischio alto: riguarda i sistemi di IA che, pur non comportando rischi inaccettabili, rappresentano una minaccia significativa per la salute, la sicurezza e i diritti delle persone. In questa categoria rientrano i sistemi che devono essere sottoposti a valutazione di conformità prima di essere immessi sul mercato, sono infatti soggetti a rigorosi controlli di documentazione, trasparenza e monitoraggio. I sistemi ad alto rischio sono anche monitorati dalle autorità nazionali competenti, che verificano periodicamente la conformità dei prodotti sul mercato; 3) rischio limitato e minimo: per i sistemi di IA che comportano rischi minori, la proposta prevede un approccio più flessibile, fondato principalmente su obblighi di trasparenza. Per un commento, v. C. Schepisi, *Le "dimensioni"* cit.

aspetto cruciale e delicato: in questo senso, il regolamento si distingue per l'approccio preventivo e la necessità di ponderare con attenzione il rischio, prevedendo che il legislatore non si limiti a reagire a danni già occorsi, ma agisca a monte, prima che i danni possano concretizzarsi, al fine di evitare danni irreparabili[93].

In linea con la giurisprudenza costituzionale e gli orientamenti internazionali, esso si fonda sul principio di prevenzione del danno[94]. La categoria di rischio "inaccettabile", introdotta dal regolamento, si colloca proprio su questa linea di principio: essa riguarda tecnologie che, se non controllate, hanno il potenziale di causare danni irreversibili, come nel caso dell'uso di sistemi di riconoscimento biometrico in tempo reale da parte delle autorità pubbliche, che sarebbe incompatibile con il diritto alla *privacy* e alla protezione dei dati personali.

Tuttavia, il regolamento non si limita a impedire l'utilizzo di queste tecnologie in tutti i casi, ma prevede delle eccezioni, laddove l'uso di tali strumenti sia giustificato dalla necessità di proteggere la vita e l'integrità fisica delle persone. È proprio in questa prospettiva che l'analisi giuridica dell'uso della tecnologia IA si trasforma in un'operazione complessa di bilanciamento tra diversi diritti[95.] Il punto nodale del bilanciamento dei diritti è la necessità di un approccio che non solo comprenda la gravità del rischio per ciascun diritto coinvolto, ma che tenga anche conto delle dinamiche intersecate tra questi diritti. Il Regolamento, in effetti, sembra optare per una concezione dinamica, dove la priorità di un diritto rispetto a un altro dipende dal contesto di applicazione del sistema di IA. Ciò nonostante, le scelte normative compiute devono essere giustificate con maggiore chiarezza, per evitare che principi come la libertà di espressione non vengano strumentalizzati per eludere l'obbligo di tutela dei diritti più vulnerabili[96] .

Sebbene il modello del bilanciamento *ex ante* sia sicuramente utile per evitare danni irreparabili, è altrettanto importante che la valutazione dei rischi non diventi un'arma a doppio taglio, lasciando che gli interessi economici prevalgano su quelli dei singoli individui. Il legislatore europeo è chiamato a fare delle scelte difficili, ma necessarie, per evitare che l'innovazione tecnologica, pur giustificata da intenti legittimi, non venga impiegata a danno della protezione dei diritti umani[97]. L'articolo 50, intitolato "Obblighi di trasparenza per i fornitori e gli utilizzatori di determinati sistemi di intelligenza artificiale", si concentra sugli obblighi di trasparenza, al fine di garantire che gli utenti siano adeguatamente informati sulla natura e sul funzionamento delle tecnologie con cui interagiscono. Nel contesto della gerarchia dei rischi delineata dalla normativa, l'articolo 50 si applica a tutte le categorie di sistemi che interagiscono direttamente con le persone fisiche. Il riferimento specifico ai *deepfake* è individuabile nel paragrafo 4 dell'articolo 50, il quale stabilisce che: "Gli utilizzatori di un sistema di intelligenza artificiale che genera o manipola contenuti di immagini,

---

[93] C. SCHEPISI, *Le "dimensioni"* cit.

[94] Il *Report* CAHAI (2020)23 del Comitato *ad hoc* sull'Intelligenza artificiale afferma: "La prevenzione del danno è un principio fondamentale che deve essere rispettato, sia nella dimensione individuale che in quella collettiva, soprattutto quando tale danno riguarda l'impatto negativo sui diritti umani, la democrazia e lo stato di diritto. L'integrità fisica e mentale degli esseri umani deve essere adeguatamente tutelata, con ulteriori garanzie per le persone e i gruppi più vulnerabili. Particolare attenzione deve essere prestata anche alle situazioni in cui l'uso di sistemi di Intelligenza artificiale può causare o aggravare impatti negativi a causa di asimmetrie di potere o di informazione, come nel caso di relazioni tra datori di lavoro e dipendenti, imprese e consumatori, o governi e cittadini". [trad. propria].

[95] A. Adinolfi, *L'Unione* cit.

Questa dinamica diventa ancora più evidente quando si considera che, in alcuni casi, l'adozione di tecnologie avanzate, pur comportando dei rischi per certi diritti, può allo stesso tempo contribuire in modo significativo alla protezione di altri diritti. Si pensi, ad esempio, ai sistemi di IA impiegati per la diagnosi di malattie o nella ricerca di persone scomparse. Qui, l'interesse pubblico e il diritto alla salute o alla sicurezza pubblica prevalgono, in alcuni contesti, sul diritto alla *privacy* o alla protezione dei dati personali. In questi casi, l'uso delle tecnologie diventa giustificato, ma solo se limitato a scopi legittimi, proporzionato al rischio e sottoposto a rigorosi controlli di trasparenza e sicurezza. Tale approccio evidenzia la necessità di una regolazione multilivello che non solo contempli la protezione dei diritti individuali, ma che bilanci anche le esigenze collettive e sociali, senza sacrificare i principi di equità e dignità umana.

[96] Il caso dei *deepfake* evidenzia proprio questa difficoltà di sintesi tra un approccio basato sulla protezione di diritti fondamentali e uno che consenta l'innovazione tecnologica senza compromettere la coesione sociale.

[97] C. Schepisi, *Le "dimensioni"* cit.

audio o video che costituiscono un *deepfake*, devono dichiarare che il contenuto è stato generato o manipolato artificialmente". La disposizione definisce in modo chiaro l'obbligo per gli utilizzatori: qualora un sistema AI venga impiegato per generare o manipolare contenuti visivi o sonori rientranti nella definizione di *deepfake*, è necessario rendere evidente l'origine artificiale dei contenuti.

La regolamentazione europea in merito ai *deepfake*, come delineato nel Regolamento, segue un approccio che privilegia la trasparenza piuttosto che un divieto assoluto sulla creazione e diffusione di questi contenuti[98]. La scelta di limitarsi a un obbligo di etichettatura, pur apparendo meno restrittiva, ha sollevato preoccupazioni legittime, in particolare per quanto riguarda l'efficacia del meccanismo. Chi intenda realizzare contenuti con finalità fraudolente o malevole, infatti, potrebbe facilmente eludere l'obbligo di trasparenza adottando tecniche per nascondere la propria identità, come l'utilizzo di *account* anonimi o *bot*. In tal modo, l'obbligo di informare l'utente che il contenuto è stato manipolato potrebbe risultare inefficace come deterrente, non riuscendo a fermare la diffusione di contenuti ingannevoli[99]. Il rischio è che, senza un quadro operativo chiaro, l'adempimento dell'obbligo di trasparenza resti frammentato e poco efficace.

L' obbligo di divulgazione appare relativamente blando se si confronta il suo impatto con il potenziale dannoso dei *deepfake*. È importante osservare che altre disposizioni dell'*AI Act* possono prevedere un regime più severo nei confronti dei sistemi di Intelligenza artificiale che generano *deepfake* particolarmente lesivi. In tali casi, la normativa prevede la possibilità di modificare la valutazione del rischio del sistema – ad esempio, facendolo passare da una categoria a rischio limitato a una categoria ad alto rischio – o, in casi estremi, di imporre divieti. In particolare, l'articolo 5(1)(b) stabilisce che i sistemi di Intelligenza artificiale ad alto rischio devono rispettare requisiti stringenti in termini di affidabilità, trasparenza e sicurezza, e che il loro impiego è subordinato a una rigorosa valutazione del rischio. Analogamente, l'articolo 6(1) prevede che i sistemi considerati particolarmente pericolosi, ossia quelli che presentano rischi inaccettabili per i diritti fondamentali, la sicurezza o altri valori fondamentali, possano essere soggetti a un divieto operativo[i]. L'articolo 7 garantisce un'ulteriore flessibilità regolatoria conferendo alla Commissione europea il potere di aggiornare l'elenco dei sistemi ad alto rischio. In base al paragrafo 1, la Commissione può includere nuove applicazioni qualora presentino un livello di rischio pari o superiore a quello delle applicazioni già elencate nell'Allegato III. Il paragrafo 2 consente inoltre di rivedere la classificazione esistente tenendo conto di elementi come la frequenza e la gravità dei danni rilevati, il numero di persone coinvolte e il quadro normativo vigente.

Qualsiasi tentativo di classificare i *deepfake* all'interno delle categorie di rischio previste dall'*AI Act* dovrebbe partire da una considerazione fondamentale: il rischio non è insito nella tecnologia in sé, ma dipende dal contesto e dagli usi concreti che se ne fanno. I *deepfake*, dunque, non andrebbero considerati a priori come strumenti pericolosi o moralmente discutibili. Al contrario, è importante riconoscere il potenziale positivo di questa tecnologia, che può trovare applicazione in ambiti come l'educazione, l'arte o la medicina. In quest'ottica, i *deepfake* dovrebbero essere considerati come strumenti neutri, il cui impatto dipende dall'intenzione con cui vengono creati e diffusi[100]. L'attuale definizione normativa dei *deepfake*, tuttavia, trascura la dimensione contestuale, ovvero quegli elementi – come la finalità, il pubblico destinatario e le modalità di diffusione – che spesso sono determinanti nel valutarne il potenziale dannoso. Un'eccessiva enfasi sui possibili usi negativi ha contribuito a costruire un'immagine distorta e stigmatizzante della tecnologia, con il rischio di portare a regolamentazioni sproporzionate che potrebbero frenare l'innovazione e compromettere la competitività europea.

[98] S. Fontana, *I deepfake* cit.
[99] R. Chesney, D.K. Citron, *Deep Fake: una sfida imminente per la privacy, la democrazia e la sicurezza* nazionale in *California Law Reviw* 107 (2019) 1755-1819.
[100] M. Łabuz, *Un'interpretazione* cit.

La classificazione proposta dall'*AI Act* è stata da alcuni considerata poco solida e arbitraria, in quanto non tiene conto dell'intero ciclo di vita dei sistemi di intelligenza artificiale, escludendo scenari d'uso rilevanti e talvolta rischiosi[101].

Nel complesso, l'approccio dell'Unione europea ai *deepfake* si caratterizza per una certa prudenza, evitando divieti rigidi e cercando di tutelare al contempo i diritti fondamentali e l'innovazione tecnologica. Questo approccio cauto contribuisce, come si diceva, ad accrescere gli interrogativi sull'efficacia della normativa nel contrastare gli abusi della tecnologia.

La sfida principale rimane quella di trovare una regolamentazione che, senza frenare l'innovazione, possa davvero prevenire gli abusi e tutelare i diritti dei cittadini. Il quadro normativo europeo è ancora in fase di evoluzione e, con l'avanzare della tecnologia, sarà necessario monitorare attentamente le future modifiche e aggiornamenti delle istituzioni europee. La regolamentazione dei *deepfake* è una questione delicata, che si inserisce in un contesto normativo più ampio, simile a quello della disinformazione in generale.

7. - Conclusioni

L'evoluzione tecnologica e la rapida diffusione degli strumenti di Intelligenza artificiale, come il *deepfake*, hanno evidenziato questioni complesse e variegate, che vanno dalla sicurezza informatica alla protezione dei diritti fondamentali, dalla tutela della *privacy* al mantenimento dell'integrità dei processi democratici. Questo fenomeno, definito dalla capacità di creare contenuti audiovisivi manipolati con un realismo tale da rendere difficile distinguere tra verità e finzione, rappresenta una minaccia senza precedenti e particolarmente subdola per la stabilità dello Stato di diritto e il corretto funzionamento della sfera pubblica. In ambito giuridico, si è resa necessaria una regolamentazione che non solo gestisca l'uso delle tecnologie di Intelligenza artificiale, ma che adotti anche una visione preventiva, capace di evitare danni gravi, attraverso un'attenta valutazione dei rischi legati a questi strumenti. Il quadro normativo tracciato dall'Unione europea, ben rappresentato dall'adozione dell'*Artificial Intelligence Act*, segna un primo passo importante verso un'armonizzazione legislativa che mira a bilanciare l'innovazione tecnologica con la protezione dei diritti fondamentali. Il modello proposto, fondato su un'analisi preventiva dei rischi e su una classificazione delle tecnologie in base al loro potenziale impatto si presenta come una risposta strutturata e flessibile alle sfide poste dalla rapida evoluzione dell'IA.

Nel caso specifico dei *deepfake*, il legislatore ha scelto un approccio improntato alla trasparenza e all'obbligo di segnalare chiaramente quando un contenuto è stato manipolato, così da permettere agli utenti di riconoscerne subito la natura artificiale. Questa linea normativa, pur senza vietare in modo assoluto la creazione o la condivisione di tali contenuti, punta a ridurre i rischi legati alla disinformazione e all'uso di video o audio falsificati in contesti come le campagne di manipolazione politica o le operazioni di propaganda. L'obbligo di rendere riconoscibili i contenuti alterati non solo rafforza il diritto a un'informazione chiara e a un'opinione consapevole, ma rappresenta anche uno strumento essenziale per contrastare minacce informatiche e strategie di disinformazione, fenomeni che, soprattutto se combinati con forme di guerra ibrida, mettono a rischio la tenuta sociale e la stabilità democratica. L'analisi complessiva del fenomeno mostra chiaramente come l'incontro tra tecnologie avanzate e strategie di manipolazione dell'informazione richieda un intervento giuridico ampio e strutturato, che vada oltre le sole regole preventive per includere anche sistemi di controllo e monitoraggio continui. In questa prospettiva, l'Unione europea si posiziona come un attore centrale nella definizione di *standard* condivisi a livello globale, promuovendo un modello regolatorio che potrebbe fare da guida anche per altri Paesi. L'integrazione tra l'*AI Act* e strumenti affini, come il *Digital Services Act* e il Codice rafforzato di buone pratiche sulla disinformazione, dimostra la

[101] V. F. R. Moreno, *Intelligenza artificiale generativa e deepfake: un approccio basato sui diritti umani per contrastare i contenuti dannosi*, 2023.

volontà di costruire un approccio coerente e coordinato, capace di rispondere in modo concreto ai rischi legati all'uso scorretto e ingannevole delle tecnologie digitali.

Dal punto di vista giuridico, il tema dei *deepfake* e della disinformazione si inserisce in un quadro complesso che impone un equilibrio delicato tra la libertà di espressione e la necessità di proteggere i diritti, sia individuali che collettivi, come la *privacy* e l'accesso a informazioni affidabili. In questo scenario, la regolamentazione non può ignorare i rischi legati all'uso improprio delle tecnologie, ma deve anche includere strumenti concreti per evitare che le piattaforme digitali vengano usate in modo distorto da attori statali o privati, in grado di approfittare delle debolezze del cyberspazio per fare propaganda o creare instabilità.

Il percorso normativo avviato dall'Unione europea, pur non privo di critiche e ancora bisognoso di ulteriori chiarimenti sul piano interpretativo, rappresenta comunque un passo concreto verso l'equilibrio tra innovazione tecnologica e tutela dei diritti.

Occorre tuttavia tenere presente che il problema dei *deepfake* supera la semplice produzione, estendendosi e amplificandosi attraverso la loro diffusione. Le modalità di condivisione sono fortemente influenzate dalle politiche adottate dalle piattaforme *online* e dai *media*, che ne determinano sia la diffusione che l'impatto. È pertanto urgente sviluppare una regolamentazione che non si limiti a valutare la tecnologia, ma tenga conto anche del contesto sociale in cui questi contenuti vengono creati e propagati. I responsabili delle politiche devono agire su diversi livelli, esaminando il "ciclo di vita" dei *deepfake* – dalla fase di creazione, alla circolazione, fino al *targeting* e alla ricezione da parte del pubblico – per affrontare in modo sistemico i rischi correlati. In questo modo, si potrà tutelare la *privacy*, la sicurezza digitale, i diritti d'autore e il diritto all'immagine, contrastando al contempo la diffusione della disinformazione che mette a repentaglio i processi democratici.

Considerando l'attuale quadro normativo europeo, le soluzioni future dovranno inserirsi in un sistema dinamico che unisca strumenti tecnologici avanzati, regolamentazioni più trasparenti e una cooperazione internazionale rafforzata. Sarà, inoltre, essenziale perfezionare continuamente gli algoritmi per l'identificazione dei contenuti manipolati, in modo da mantenere il passo con tecniche di falsificazione sempre più sofisticate. La proposta dell'*Artificial Intelligence Act*, unitamente al GDPR e al *Digital Services Act*, fornisce una solida base normativa da cui sviluppare ulteriori misure. Inoltre, l'attenzione non può limitarsi alla produzione dei *deepfake*, ma deve estendersi alla loro diffusione, investire nella formazione degli utenti e migliorare la *digital literacy* è fondamentale per rafforzare la consapevolezza critica e contenere l'impatto della disinformazione.

In definitiva, è necessario un impegno coordinato che integri innovazione tecnologica e un'interpretazione pragmatica delle norme, garantendo così che l'evoluzione dei *deepfake* non comprometta i diritti fondamentali né la stabilità del sistema informativo europeo. I *deepfake* costituiscono oggi una delle manifestazioni più insidiose del potenziale manipolativo delle tecnologie digitali. La regolazione giuridica, pertanto, non può fermarsi all'analisi del rischio insito negli strumenti di generazione automatica, ma deve estendersi anche alle modalità di diffusione e fruizione dei contenuti alterati. Il vero impatto, infatti, si registra nella fase di circolazione, quando video o audio falsificati si propagano attraverso le piattaforme digitali senza filtri adeguati, spesso in assenza di *policy* efficaci da parte degli attori del *web*. In tale dinamica, soggetti pubblici e privati possono facilmente sfruttare le vulnerabilità del cyberspazio per proseguire i propri interessi. Da qui la necessità di un assetto normativo che sia al contempo coerente e dinamico, capace di affrontare l'intero fenomeno dei *deepfake* in modo sistemico, anche alla luce dell'evoluzione rapida delle tecnologie implicate.

Il percorso avviato dall'Unione europea, pur tra critiche e la necessità di chiarimenti interpretativi, rappresenta un passo concreto verso un equilibrio tra progresso tecnologico e tutela dei diritti fondamentali. L'orientamento normativo non si limita a disciplinare gli strumenti, ma cerca di leggere il fenomeno *deepfake* nel suo contesto d'uso, considerando l'intero processo: dalla creazione, alla condivisione, fino alla ricezione da parte del pubblico. Questo approccio multilivello punta a intervenire su tutte le fasi del ciclo di vita dei contenuti manipolati, dagli aspetti tecnologici alla distribuzione, dalla protezione delle persone ritratte fino alla consapevolezza degli utenti. Per essere

realmente efficace, la risposta regolatoria deve muoversi in modo coordinato, combinando strumenti di tutela della *privacy*, salvaguardia del diritto all'immagine e contrasto attivo alla disinformazione. L'attuazione concreta di strumenti come l'etichettatura obbligatoria dei contenuti manipolati continua a presentare nodi critici. La mancanza di linee guida operative chiare e di un sistema sanzionatorio efficace rischia di svuotare di contenuto la normativa, soprattutto nei casi, sempre più frequenti, in cui i *deepfake* circolano in modo anonimo o su piattaforme difficili da monitorare.

A ciò si aggiunge un altro aspetto essenziale: la tutela dei dati personali. Quando le tecnologie generative utilizzano, senza consenso, immagini, voci o dati sensibili, si entra in una zona grigia che solleva questioni delicate sotto il profilo del diritto alla *privacy* e alla protezione dell'identità. Sebbene l'*AI Act,* il GDPR e il *Digital Services Act* offrano già un impianto normativo rilevante, non basta: serve un'integrazione più incisiva, che colmi le lacune interpretative e renda davvero effettive le tutele previste. Perché queste norme abbiano un impatto concreto, è fondamentale affiancarle a linee guida operative ben definite, strumenti di enforcement realmente efficaci e un coordinamento più stretto tra istituzioni pubbliche, piattaforme digitali e autorità di vigilanza. Parallelamente, diventa imprescindibile investire sull'educazione digitale: rafforzare la "*digital literacy*" degli utenti è il primo passo per sviluppare un approccio critico all'informazione, contrastare le dinamiche manipolative e aumentare la resilienza collettiva contro la disinformazione. Un modello di governance multilivello, che coniughi innovazione tecnologica e tutela sostanziale dei diritti, è l'unico modo per affrontare le sfide dei *deepfake* senza mettere a rischio la solidità del sistema democratico o frenare il potenziale dello sviluppo digitale.

INDICE

INTRODUZIONE

1. Il contrasto alla criminalità transnazionale in ambito internazionale ed europeo

1.1 Le tappe del processo di regolamentazione del contrasto al crimine organizzato transnazionale

1.1.1 La criminalità transnazionale e il cambiamento geopolitico

1.1.2 Tra dibattito e azione: gli atti preliminari alla Convenzione delle Nazioni Unite contro la criminalità transnazionale

1.1.3 La Conferenza di Napoli: un passo decisivo verso la Convenzione contro il crimine organizzato transnazionale

1.1.4 La Convenzione di Palermo e i protocolli addizionali: un modello "globale" per combattere la criminalità transnazionale

1.2 Contenuti e limiti della Convenzione delle Nazioni Unite sul contrasto alla criminalità organizzata transazionale del 2000

1.2.1 La natura transnazionale degli illeciti di interesse globale

1.2.2 Gli strumenti di contrasto alla criminalità organizzata transnazionale

1.2.3 Il ruolo della prevenzione nella cooperazione internazionale contro la criminalità

1.2.4 Le sfide e le prospettive della Convenzione di Palermo: il meccanismo di revisione

1.3 Criminalità transnazionale e Unione europea

1.3.1 Il regime giuridico europeo nella lotta alla criminalità transnazionale: differenze e sinergie con il sistema internazionale

1.3.2 Libertà, sicurezza e giustizia: il contributo dell'Atto Unico europeo e degli accordi di Schengen nel contrasto ai reati transnazionali

1.3.3 Maastricht e Amsterdam passaggi decisivi per l'integrazione europea nel settore giudiziario e di polizia

1.3.4 Il Trattato di Lisbona e il rinnovamento del sistema penale europeo per una giustizia condivisa

1.4 Le strutture e meccanismi operativi della Cooperazione giudiziaria in materia penale nell'Unione europea

1.4.1 La rete giudiziaria europea e i magistrati di collegamento: esempi di cooperazione orizzontale per la giustizia penale

1.4.2 Il contributo fondamentale di Europol nella lotta al crimine transnazionale

1.4.3 Il ruolo di Eurojust nel coordinamento delle indagini giudiziarie transnazionali in Europa

1.4.4 La Procura europea nel rafforzamento della giustizia eurounitaria

1.4.5 La nuova visione dell'Unione europea nella lotta alla criminalità organizzata (2021 – 2025)

2. *Cybercrime e Cibersecurity*

2.5 *Cybercrime*: la minaccia del crimine digitale

2.5.1 Tecnologia, crimine e regolamentazione: la nuova era della connettività digitale

2.5.2 Il crimine digitale e la disparità nelle definizioni: riflessioni su un quadro normativo incompleto

2.5.3 Esplorando il *Cybercrime*: tipologie, tecniche e impatti dei reati digitali

2.6 La Convenzione di Budapest: fondamenti giuridici e prospettive per la lotta al *cybercrime*

2.6.1 L'evoluzione della giustizia penale nell'era digitale

2.6.2 Armonizzazione e definizioni: il linguaggio giuridico comune nella classificazione dei reati informatici

2.6.3 Strumenti e procedure per contrastare i reati informatici

2.6.4 Nuovi sviluppi nella lotta al *cybercrime*: l'aggiornamento convenzionale del 2024

2.7 C*ybercrime* e Unione europea: iniziative e legislazione

2.7.1 Le iniziative dell'Unione europea nel contrasto al *cybercrime*: la Decisione 222/2005161

2.7.2 La Direttiva 2013/40/UE e il futuro dell'Europa digitale

2.7.3 Costruire un *cyberspazio* sicuro: la resilienza come pilastro della sicurezza europea

2.7.4 ENISA e la gestione degli incidenti cibernetici: promuovere una cultura della sicurezza in Europa
2.8 *Cybersecurity* e Unione europea: proteggere il *cyberspazio*
2.8.1 Proteggere il *cyberspazio*: dalla minaccia alla difesa, le politiche per un futuro sicuro
2.8.2 La strategia UE 2013 per una difesa cibernetica sostenibile
2.8.3 Da NIS1 a NIS2: rafforzare la resilienza cibernetica nell'Unione europea
2.8.4 Il Regolamento UE 2023/2841, un nuovo capitolo nella sicurezza cibernetica dell'Unione europea
3.Intelligenza artificiale, *Deepfake* e Unione europea
3.1 Origine ed evoluzione dell'Intelligenza artificiale
3.1.1 Opportunità e rischi delle nuove tecnologie
3.1.2 Le origini e le definizioni dell'Intelligenza artificiale
3.1.3 L'evoluzione della definizione di IA nell'Unione europea
3.2 Intelligenza artificiale e sicurezza
3.2.1 La dualità della tecnologia nella lotta e nel rafforzamento dei crimini informatici
3.2.2 *Predictive policing*: tra prevenzione del crimine e protezione dei diritti fondamentali
3.2.3 Limiti e prospettive del diritto penale del futuro
3.2.4 GDPR e giustizia predittiva: garanzie per la difesa e la trasparenza
3.3 Regolamentazione e futuro dell'Intelligenza artificiale in Europa
3.3.1 L'*AI act* e la creazione di un mercato unico sicuro e affidabile
3.3.2 La valutazione del rischio come approccio equilibrato tra innovazione e sicurezza
3.3.3 Il divieto delle tecnologie invasive: proteggere i diritti fondamentali dell'era digitale
3.3.4 La funzione dello Stato nell'attuazione dell'*AI act*
3.4 *Deep fake* e implicazione nella società contemporanea
3.4.1 La rivoluzione dei *Deep fake*: impatti giuridici ed etici nella società contemporanea
3.4.2 La trasformazione della disinformazione e le nuove minacce alla società e alla democrazia
3.4.3 Il *Digital Service Act* contro la disinformazione: trasparenza nella moderazione dei contenuti *on line*
3.4.4 Il ruolo dell'Unione europea nella gestione dei contenuti manipolati
3.5 Sovranità digitale e minacce ibride
3.5.1 Sovranità digitale e geopolitica: l'UE di fronte alle minacce esterne
3.5.2 Minacce ibride e *cyberwarfare*: il nuovo volto dei conflitti globali
3.5.3 La dinamica delle misure attive nella guerra ibrida
3.5.4 L'uso dei *Deep fake* nel conflitto russo – ucraino

Conclusioni

Il contributo riprende, ampliandole ed aggiornandole attività di ricerca condotte nel quadro del Modulo Jean Monnet EU-GLOBACT "Progetto finanziato dal Programma Erasmus+ della Commissione Europea – Progetto d'Azione Jean Monnet numero 101126599. Finanziato dall'Unione Europea. I punti di vista e le opinioni espressi sono tuttavia solo quelli dell'autore o degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili per questo".

Abstract:

L'articolo esamina il tema dei *deepfake* e della disinformazione soffermandosi sulla necessità di un equilibrio tra la libertà di espressione e la necessità di proteggere i diritti, sia individuali che collettivi, come la *privacy* e l'accesso a informazioni affidabili. In questo scenario, la regolamentazione non può ignorare i rischi legati all'uso improprio delle tecnologie, ma deve anche includere strumenti concreti per evitare che le piattaforme digitali vengano usate in modo distorto da attori statali o privati, in grado di approfittare delle debolezze del cyberspazio per fare propaganda o creare instabilità.

Come l'indagine mostrerà , il percorso normativo avviato dall'Unione europea, pur non privo di critiche e ancora bisognoso di ulteriori chiarimenti sul piano interpretativo, rappresenta comunque un passo concreto verso l'equilibrio tra innovazione tecnologica e tutela dei diritti.

Un modello di governance multilivello, che coniughi innovazione tecnologica e tutela sostanziale dei diritti, è l'unico modo per affrontare le sfide dei *deepfake* senza mettere a rischio la solidità del sistema democratico o frenare il potenziale dello sviluppo digitale.

The article examines the topic of *deepfakes* and disinformation by focusing on the need for a balance between freedom of expression and the need to protect rights, both individual and collective, such as *privacy* and access to reliable information. In this *scenario*, regulation cannot ignore the risks associated with the misuse of technologies, but must also include concrete tools to prevent digital platforms from being used in a distorted way by state or private actors, able to take advantage of the weaknesses of cyberspace to make propaganda or create instability.

As the work will show, the regulatory path initiated by the European Union, although not without criticism and still in need of further clarification on the interpretative level, represents a concrete step towards the balance between technological innovation and protection of rights.

A multi-level governance model, combining technological innovation and substantial protection of rights, is the only way to address the challenges of *deepfakes* without jeopardizing the solidity of the democratic system or curbing the potential of digital development