VOCI
COSTITUZIONALI

ALESSIA FORTE

L'apparenza inganna: i dilemmi costituzionali del *deepfake*

Dottoranda in Diritto costituzionale
Università degli Studi di Milano Statale

01/04/2026

ALESSIA FORTE

L'apparenza inganna: i dilemmi costituzionali del *deepfake*

SOMMARIO: 1. Introduzione - 2. Dalle *fake news* ai *deepfake*: un salto qualitativo - 3. Il quadro normativo europeo - 4. Il recente intervento normativo nazionale e cenni conclusivi

ABSTRACT (italiano)
Il presente contributo analizza alcune delle problematiche sollevate dall'ormai inarrestabile avanzata dei *deepfakes* nel contesto del dibattito pubblico. In un primo momento, l'indagine si sofferma sul salto qualitativo compiuto da tale fenomeno rispetto a quello ancor più noto delle *fake news*, evidenziandone il peggioramento in termini percettivi. In un secondo momento, il lavoro si sofferma sul quadro normativo di riferimento, europeo e nazionale, ponendo a confronto l'approccio preventivo basato sulla trasparenza dell'AI Act con la recentissima scelta di criminalizzazione operata dal legislatore nazionale attraverso l'introduzione dell'art. 612-quater c.p.

ABSTRACT (inglese)
This paper examines the critical issues arising from the relentless rise of deepfakes within the context of public debate. The first part focuses on the qualitative leap this phenomenon represents compared to the well-known issue of fake news, highlighting its detrimental impact on sensory and cognitive perception. The second part analyzes the relevant legal framework at both the European and national levels. Specifically, it provides a comparative assessment of the preventive approach based on transparency established by the EU AI Act and the recent shift toward criminalization adopted by the Italian legislator through the introduction of Article 612-quater of the Criminal Code.

PAROLE CHIAVE (italiano)
Libertà d'espressione, *Deepfake*, manipolazione, *Information disorder*, dibattito pubblico.

PAROLE CHIAVE (inglese)
Freedom of expression, Deepfakes, Information disorder, Public speech.

1. Introduzione

L'essere compiutamente e realmente informati rappresenta una premessa indispensabile del buon funzionamento di un regime liberaldemocratico: non soltanto nel senso che il diritto di ricevere e di ricercare informazioni costituisce un presupposto ed uno strumento perché si renda effettiva una libertà cardinale come quella di espressione o di manifestazione del pensiero; ma, prima ancora, nel senso che si tratta di un'esigenza indeclinabile per evitare che sia distorto il formarsi della pubblica opinione[1].

Nell'attuale ecosistema digitale l'informazione è istantanea, pervasiva, globale. Una notizia, anche falsa, può nascere in un angolo remoto della rete e, nel giro di pochi minuti, diventare virale. L'"*information disorder*"[2] non è una semplice patologia del sistema informativo: è ormai una condizione strutturale che mette in discussione la capacità stessa dei cittadini di orientarsi nel dibattito pubblico.

La dottrina costituzionalistica più risalente aveva già intuito la coessenziale duplicità della libertà garantita dall'art. 21: essa è, al tempo stesso, diritto individuale e diritto funzionale, essenziale per la garanzia del pluralismo e la vitalità democratica[3]. La qualità dell'informazione è una condizione minima di esistenza della democrazia. Senza verità, o almeno senza un ragionevole grado di affidabilità, la formazione della pubblica opinione rischia di diventare fragile, manipolabile, esposta a distorsioni sistemiche. Le tecnologie contemporanee rendono questa lettura ancora più evidente.

2. Dalle *fake news* ai *deepfake*: un salto qualitativo

I rischi legati alla disinformazione non sono certo nuovi: le *fake news* hanno già dimostrato con evidenza la vulnerabilità del dibattito pubblico. Tuttavia, i *deepfake* rappresentano, per certi versi, un salto qualitativo rispetto alle tradizionali manipolazioni.

L'intelligenza artificiale è, senza dubbio, un potente fattore di progresso. Però, come tutte le tecnologie, è uno strumento e, in quanto tale, apparentemente neutrale. Può essere impiegata per finalità benefiche come per scopi talvolta tutt'altro che leciti.

Il *deepfake* in cui l'immagine e la voce di un individuo sono contraffate in maniera tale da renderli indistinguibili dall'originale ne è l'esempio più inquietante.

A titolo esemplificativo, il viso di una persona può essere inserito sul corpo di un'altra, o la voce di un individuo può essere associata ad un terzo soggetto. Anche i movimenti del corpo possono essere artefatti, fornendone una falsa rappresentazione. Come pure un soggetto può essere posizionato in luoghi e contesti non reali o, comunque, dal medesimo non conosciuti e non frequentati.

In sostanza, il *deepfake* rappresenta un salto di qualità importante in senso negativo rispetto alle stesse *fake news* poiché mentre delle notizie si può dubitare, si può mettere in dubbio la loro attendibilità, si può non avere fiducia nell'imparzialità della fonte da cui provengono, non si è abituati a dubitare di ciò che colpisce direttamente i sensi. L'elemento dirompente che caratterizza i *deepfake* è proprio la capacità di intaccare direttamente la percezione sensoriale: un contenuto audiovisivo realistico tende ad essere accolto come "prova": non è un caso, infatti, che nei giudizi

[1] Cfr. L. PALADIN, *La libertà dell'informazione*, Torino, UTET, 1979, p. 3, secondo cui «senza una libera informazione non potrebbe neppure fondarsi ed esistere una democrazia del tipo in cui si colloca l'Italia, perché il voto non sarebbe sostanzialmente libero».

[2] Invero, l'insieme fenomenico dei *deepfakes* e delle *fake news* dà vita a quello che è stato definito "*l'information disorder*". Sul tema si rimanda, in particolare, al contributo di S. MAHER, *Deep Fakes. Seeing and Not Believing*, in M. FILIMOWICZ (a cura di) *Deep Fakes. Algorithms and Society*, 2022, p. 1 ss.

[3] Sul tema si v., in particolare, il noto saggio di C. ESPOSITO, *La libertà di manifestazione del pensiero nell'ordinamento italiano,* Milano, Giuffrè, 1958, p. 9 ss.

penali la testimonianza oculare (coloro che hanno visto) goda storicamente di una forza persuasiva superiore alla testimonianza *de relato* (coloro che riferiscono di aver appreso un fatto o una circostanza)[4].

Il *deepfake* incrina proprio questa dimensione percettiva in cui, in un certo senso, non basta più "vedere per credere".

3. Il quadro normativo europeo

L'ordinamento europeo non ha adottato, almeno per ora, una disciplina *ad hoc* dedicata al fenomeno in commento. Ciò non significa che esso sia rimasto privo di attenzione normativa tanto che all'interno del Digital Services Act e, soprattutto, dell'Artificial Intelligence Act, possono trovarsi riferimenti più o meno espliciti, nonché disposizioni più o meno applicabili.

In particolare, il Regolamento sull'IA definisce il *deepfake* come un'«immagine o un contenuto audio o video generato o manipolato dall'IA che assomiglia a persone, oggetti, luoghi, entità o eventi esistenti e che apparirebbe falsamente autentico o veritiero a una persona» (art. 3, n. 60, AI Act).

Con riguardo alle tutele previste, l'art. 50 disciplina espressamente i *deepfake* tra gli obblighi di trasparenza, imponendo a fornitori e *deployer* di rendere riconoscibile quando un contenuto è stato artificialmente generato o manipolato. È tuttavia prevista un'eccezione a tale obbligo in caso di usi autorizzati dalla legge per rilevare, prevenire, indagare e perseguire reati.

Inoltre, se si tratta di creazioni artistiche, creative, satiriche o fittizie, l'obbligo di trasparenza è limitato e non può ostacolare "l'esposizione o il godimento dell'opera".

In proposito si potrebbe dire che non è manchevole l'attenzione riposta nel bilanciamento con altri diritti e libertà. È difatti espressamente previsto, per i contenuti *deepfake*, che tale obbligo di trasparenza non deve essere interpretato come un impedimento al diritto alla libertà di espressione e alla libertà delle arti e delle scienze come garantiti nella Carta dei diritti fondamentali dell'UE, in particolare quando il contenuto faccia parte di un'opera evidentemente creativa, satirica, artistica o di finzione, sebbene debbano comunque essere rispettate adeguate tutele per i diritti e le libertà di terzi[5].

Inoltre, sempre in tale disposizione, si può apprezzare «l'assenza di un atteggiamento eccessivamente "paternalistico"»: i *deepfake* non sono in sé vietati; ciò non toglie che lo diventino nel caso corrispondano a contenuti illeciti (come nel caso dei contenuti pornografici generati artificialmente associando immagini di persone reali, che formano la maggior parte dei *deepfake* esistenti)[6].

4. Il recente intervento normativo nazionale e cenni conclusivi

Fino a tempi recenti, l'ordinamento italiano non disponeva di una disciplina specifica sul tema dei *deepfake*. Pertanto, in caso di lesioni penalmente rilevanti, la dottrina tentava di ricondurre il

[4] Cfr., sul punto, l'intervento di F. POSTERARO, al Convegno *La minaccia del deepfake: come affrontarla in Italia?*, Roma, 9 dicembre 2019, consultabile su www.radioradicale.it.

[5] Cfr. S. TROZZI, *La dimensione costituzionale dell'Intelligenza artificiale generativa. La tutela della dignità umana nell'era del deepfake*, in *Diritto Pubblico Europeo Rassegna online*, 1/2024 p. 241.

[6] In questi termini, si v. G.E. VIGEVANI, *Potere politico e mezzi di comunicazione*, in *Rivista AIC*, 2/2025, p. 268 ss. ove vengono efficacemente indagate le scelte del legislatore europeo sul tema dell'Intelligenza Artificiale.

fenomeno a fattispecie esistenti: sostituzione di persona (art. 494 c.p.), frode informatica (art. 640-ter, comma 3, c.p.), diffamazione (art. 595 c.p.), *revenge porn* (art. 612-*ter* c.p)[7].

Oggi, è di particolare interesse è l'art. 26 della legge 23 settembre 2025, n. 132, "Disposizioni e deleghe al Governo in materia di intelligenza artificiale", il quale introduce, nell'alveo delle fattispecie penali, l'art. 612-*quater* c.p. rubricato "Illecita diffusione di contenuti generati o manipolati artificialmente".

L'introduzione di tale fattispecie rappresenta certamente una novità di rilievo: la norma incrimina l'illecita diffusione di contenuti generati o manipolati artificialmente, segnando il passaggio da una fase di adattamento di fattispecie esistenti a una tutela autonoma e tipizzata.

Tuttavia, questa scelta solleva alcune riflessioni critiche.

Anzitutto, ci si può chiedere se fosse davvero necessario ricorrere al diritto penale. Non pochi commentatori avevano osservato che gli strumenti già esistenti – sostituzione di persona, diffamazione, trattamento illecito di dati personali, *revenge porn* – avrebbero potuto coprire la maggior parte delle ipotesi. L'introduzione di una nuova incriminazione rischia di alimentare quella tendenza alla ipertrofia penale che, nel nostro ordinamento, mal si concilia con il principio di *extrema ratio* della sanzione penale.

Un secondo profilo riguarda la determinatezza della fattispecie. Come definire, in concreto, la soglia tra manipolazione penalmente rilevante e manipolazione lecita? Si pensi al ruolo della satira, della parodia, della creazione artistica: confini labili che, se non interpretati con equilibrio, potrebbero generare un effetto di dissuasione anche su attività pienamente lecite.

C'è poi il problema probatorio: come accertare in giudizio l'artificialità di un contenuto? Con quali strumenti tecnici si potrà distinguere un *deepfake* ingannevole da un contenuto genuino o da una semplice rielaborazione creativa? Qui si aprono questioni complesse, che richiederanno competenze tecnico-forensi forse ancora poco diffuse nei tribunali.

Infine, il raffronto con l'Europa mette ulteriormente in luce il carattere peculiare della scelta italiana.

L'UE, come evidenziato, ha privilegiato una prospettiva diversa: non quella repressiva, ma quella preventiva e di trasparenza, con obblighi che ricadono soprattutto sui fornitori e sugli utilizzatori professionali dei sistemi di IA. L'Italia, invece, ha imboccato la via della penalizzazione. Due approcci che non necessariamente si escludono, ma che dovranno essere coordinati, per garantire un'applicazione coerente e rispettosa dei principi sovranazionali.

Ad ogni modo, anche alla luce di queste considerazioni, credo che il punto centrale sia uno: la battaglia contro i *deepfake* non potrà essere vinta solo sul terreno delle incriminazioni. Servono, certo, regole chiare e strumenti di tutela, ma serve anche molto altro: tecnologie di autenticazione in grado di certificare l'origine dei contenuti digitali; responsabilità delle piattaforme nella segnalazione e rimozione di contenuti manipolati; e, soprattutto, educazione critica dei cittadini, perché nessuna norma potrà mai sostituire la capacità di ciascuno di interrogarsi sulle fonti, di dubitare, di esercitare quello spirito critico che è la migliore garanzia per una democrazia vitale.

In fondo, il rischio maggiore dei *deepfake* non è tanto quello di ingannarci una volta, quanto quello di renderci incapaci di credere a qualsiasi cosa (in sostanza, ciò che la dottrina americana definisce il *liar's dividend*)[8]. Ecco perché la risposta giuridica, pur necessaria, deve sempre accompagnarsi a una riflessione culturale più ampia: proteggere non solo le vittime immediate della manipolazione digitale, ma la stessa fiducia collettiva nell'informazione quale bene pubblico essenziale.

---

[7] A titolo esemplificativo, si v. il contributo di M. PICCARDI, *L'impatto del deepfake sul diritto penale*, in AA.VV. *Cybercrime*, UTET, 2023, p. 1453 ss.

[8] Per un approfondimento sul tema, si v., il contributo di J.A. GOLDSTEIN, A. LOHN, *Deepfakes, Elections, and Shrinking the Liar's Dividend*, 2024, consultabile su www.brennancenter.org, in cui gli autori sondano le preoccupazioni relative alla diffusione di *deepfake*s ingannevoli e, soprattutto, al c.d. *liar's dividend*.