LUISS

Corso di laurea in Giurisprudenza

Cattedra di Diritto Penale 2

Il contrasto al terrorismo nell'era dell'intelligenza artificiale: il ruolo delle piattaforme digitali

Prof. Maurizio Bellacosa

RELATORE

Prof.ssa Francesca Minerva

CORRELATORE

Lavinia Palamara Matr. 168113

CANDIDATA

Anno accademico 2024/2025

Sommario

INTRODUZIONE

L'avvento dell'Intelligenza Artificiale (che d'ora in avanti verrà indicata con l'acronimo IA) segna una delle più profonde trasformazioni tecnologiche e sociali della nostra epoca, definendo i contorni di una nuova era di transizione digitale. Questa forza inarrestabile, pervasiva e dinamica, sta ridefinendo ogni aspetto della vita umana, dall'economia alla medicina, dalla comunicazione alla sicurezza. La sua capacità di analizzare enormi volumi di dati, riconoscere schemi complessi e persino prendere decisioni autonome ha aperto scenari di progresso impensabili fino a pochi decenni fa, promettendo efficienza, innovazione e soluzioni a problemi di lunga data. Tuttavia, la stessa potenza che rende l'IA una risorsa inestimabile le conferisce anche un potenziale intrinsecamente duplice, proiettandola in una zona grigia in cui l'opportunità si confonde con la minaccia.

La discussione sull'IA non può, infatti, prescindere da una profonda riflessione sulle implicazioni etiche, sociali e, in particolare, sulla sicurezza.

Se da un lato l'IA offre strumenti senza precedenti per migliorare la qualità della vita e affrontare sfide globali, dall'altro introduce nuove vulnerabilità e amplifica rischi preesistenti. Le tecnologie basate sull'IA, siano esse integrate in software complessi o in hardware all'avanguardia, possono essere sfruttate per scopi malevoli, creando nuove frontiere per la criminalità.

Il lavoro in questione si propone di esplorare in profondità questa complessa dualità, con specifico riferimento ad una delle più allarmanti forme di criminalità: quella del terrorismo internazionale.

La domanda di ricerca centrale che guida questo studio è: "In che modo l'Intelligenza Artificiale sta trasformando la minaccia terroristica globale e quali sono le risposte normative e

tecnologiche che i sistemi giuridici internazionali e nazionali stanno sviluppando per contrastarla, bilanciando sicurezza e libertà individuali nell'era digitale?"

Per rispondere a questa complesso interrogativo, il lavoro svolto si propone di raggiungere una serie di obiettivi specifici: in primo luogo, definire e contestualizzare il fenomeno dell'Intelligenza Artificiale, analizzandone la sua duplice natura di opportunità e potenziale minaccia, con un'attenzione particolare alle tecnologie a rischio e ai riverberi delle stesse sulla materia del diritto penale. In secondo luogo, esaminare i modi specifici in cui l'IA può essere impiegata come strumento per la commissione di atti terroristici, distinguendo tra cyber threats, physical threats e political threats. Successivamente, si analizzeranno le strategie preventive e regolatorie adottate a livello internazionale ed europeo per contrastare l'uso terroristico dell'IA, comparando il modello tecno-libertario americano con il più strutturato modello regolatorio europeo, e approfondendo la normativa nazionale italiana. Infine, la ricerca valuterà il ruolo cruciale delle piattaforme digitali, indagando come l'IA sia impiegata sia come strumento di attacco che di prevenzione in questo contesto, analizzando le recenti decisioni giurisprudenziali e le sfide poste dalla disciplina attuale, alla luce anche della normativa sovranazionale e di quella interna con precipuo riferimento al DDL sull'Intelligenza Artificiale 1146-B recentemente modicato alla Camera dei deputati.

La rilevanza di questa ricerca risiede nell'estrema attualità e urgenza del tema. L'intersezione tra Intelligenza Artificiale e sicurezza nazionale rappresenta una delle sfide più significative del nostro tempo, richiedendo un'analisi interdisciplinare che superi i tradizionali confini tra diritto, tecnologia e scienze politiche.

La rapidità con cui l'IA si sviluppa, anche dal punto di vista della regolamentazione normativa, impone una riflessione costante sull'adeguatezza delle cornici giuridiche esistenti e sulla necessità

di nuove soluzioni che garantiscano la sicurezza senza compromettere i diritti fondamentali. Questo studio mira a contribuire al dibattito accademico e politico, offrendo una panoramica critica e strutturata sulle dinamiche emergenti, sui rischi specifici e sulle opportunità di mitigazione offerte dall'IA nel contesto della lotta al terrorismo.

L'approccio metodologico adottato in questa tesi è di tipo analitico-comparativo e giuridico. Verranno analizzate le normative vigenti e quelle in fase di sviluppo, le giurisprudenze rilevanti e la dottrina specifica in materia di Intelligenza Artificiale, diritto penale e terrorismo. Come detto, particolare enfasi sarà posta sull'analisi comparativa tra i modelli regolatori americano ed europeo, per comprenderne le filosofie sottostanti e le ricadute pratiche. Lo studio si avvarrà anche di un'analisi di casi reali e di una valutazione critica delle applicazioni tecnologiche dell'IA, sia per scopi offensivi che difensivi.

L'elaborato in questione è strutturato in quattro capitoli, ciascuno dei quali approfondisce un aspetto specifico del rapporto tra Intelligenza Artificiale e terrorismo.

Il Capitolo 1, "Le nuove frontiere del terrorismo e l'utilizzo dei sistemi di IA nell'era della rivoluzione tecnologica ", introduce il concetto di Intelligenza Artificiale, fornendone un sommario ma necessario inquadramento storico ed una definizione e un'analisi del suo duplice impatto, come opportunità e come minaccia. In tale ambito, verrà esaminato sia l'approccio antropocentrico ed eticamente sostenibile dell'IA sia le tipologie di tecnologie IA a rischio con particolare riferimento a quelli utilizzati nell'amministrazione della giustizia.

Il Capitolo 2, " L'Intelligenza Artificiale come strumento per la commissione di reati terroristici e come nuovo ed ulteriore strumento di contrasto all'interno del sistema normativo", si addentra nella duplicità delle questioni che hanno rilevanza di carattere generale all'interno della macroarea dei rapporti tra intelligenza artificiale e sistema penale, e che, come detto,

assumono un rilievo del tutto particolare rispetto ai fenomeni del terrorismo. Così dopo aver definito e contestualizzato il terrorismo internazionale e le sue trasformazioni nell'era digitale con precipuo riferimento alla Direttiva (UE) 2017/541 sulla Lotta contro il Terrorismo, il capitolo esplorerà l'intersezione tra Intelligenza Artificiale ed i fenomeni del terrorismo, distinguendo gli utilizzi leciti da quelli illeciti dell'IA. In particolare, verranno analizzate in dettaglio le modalità attraverso cui l'IA può essere sfruttata per la commissione di reati terroristici, suddividendole in *cyber threats, physical threats e political threats,* fornendo esempi e scenari ipotetici ma realistici.

Il Capitolo 3, " Modelli di regolazione dell'intelligenza artificiale: implicazioni penali e prime decisioni giudiziarie sull' IA generativa", sposta l'attenzione sulle risposte e le strategie di contrasto sia in ambito sovranazionale che in ambito interno. Si analizzeranno i modelli regolatori internazionali, in particolare il modello regolatorio europeo incentrato sulla disciplina di carattere generale del regolamento AI ACT e il modello tecno-libertario americano, con i suoi fondamenti sulla libertà di espressione online. In particolare, verrà esaminata la continua evoluzione della normativa europea ed i riflessi della stessa sull'attività delle Big Tech, tra cui Google, OpenAI Microsoft, Amazon e Mistral, la più importante azienda di IA europea, che più volte ha avuto un atteggiamento critico per l'eccesso regolatorio sull'AI Act. Il capitolo si concluderà con gli interventi di carattere settoriale sulla disicplina dei deepfake e sui risvolti della stess in ambito penalistico attraverso una comparazione normativa e l'esame della casistica giudiziaria raggruppata in tre filoni: quella del revenge porn; quella della diffamazione nel cui ambito rientrano i discorsi di odio; quella della manipolazione politica.

Infine, il Capitolo 4, "Forme di responsabilità delle piattaforme digitali", approfondirà la funzione delle piattaforme digitali, che agiscono sia come terreno fertile per la radicalizzazione, il

reclutamento e la propaganda terroristica, sia come strumenti potenziali per il loro contrasto. Il capitolo esaminerà le applicazioni dell'Intelligenza Artificiale per il contrasto al terrorismo sulle piattaforme, evidenziandone sfide e limiti. Un'analisi approfondita sarà dedicata alle decisioni giudiziarie più rilevanti, come il Caso Gonzalez vs Google e Twitter v. Taamneh, per comprenderne l'impatto sulla responsabilità e sul ruolo delle piattaforme con particolare riferimento all'adeguatezza della disciplina attuale per affrontare le sfide specifiche poste da foto, video e propaganda sulle piattaforme.

Le Conclusioni di questo lavoro mireranno a sintetizzare i risultati emersi da questa complessa analisi, proponendo riflessioni sulle sfide future e sulle possibili direzioni per lo sviluppo di un quadro normativo e tecnologico che possa efficacemente mitigare i rischi del terrorismo nell'era dell'IA, salvaguardando al contempo i principi democratici e i diritti fondamentali. Si cercherà di delineare un percorso che bilanci l'innovazione inarrestabile dell'IA con la necessaria esigenza di sicurezza e giustizia in un mondo sempre più interconnesso e digitalizzato.

CAPITOLO PRIMO: LE NUOVE FRONTIERE DEL TERRORISMO E L'UTILIZZO DEI SISTEMI DI IA NELL'ERA DELLA RIVOLUZIONE TECNOLOGICA.

1.Metaverso, intelligenza artificiale e piattaforme digitali.

A partire dalla data simbolo dell'11 settembre 2001, i dati di esperienza sulla realtà empirico-criminologica del terrorismo internazionale[1] consentono di scandire tre distinte fasi:

- la prima, caratterizzata dagli eclatanti attentati alle torri gemelle a New York, si può collocare dal tempo di *Al Qaida* all'affermazione di *Islamic State*;
- la seconda, caratterizzata dal c.d. terrorismo territoriale, si può collocare dall'avvento dell'*Islamic State*, sino alla sconfitta sul campo, ed all'attuale, quasi completa, scomparsa di tale forma di terrorismo;
- la terza coeva ai tempi in cui viviamo, ci pone di fronte ad una situazione in evoluzione, fluida, non facilmente decifrabile, sia per le caratteristiche evolutive del fenomeno criminale, che per l'avvento delle nuove

[1] M. ROMANELLI, *Criminalità organizzata e terrorismo: la circolazione dei modelli criminali e degli strumenti di contrasto,* in *Sistema penale*, 20 dicembre 2019. Nello scritto in questione l'Autore ricostruisce i diversi periodi all'interno dell'operatività dello Stato Islamico, ed i relativi fenomeni criminali; AA. VV., *Contrasto multilivello al terrorismo internazionale e rispetto dei diritti umani*, G. Giappichelli Editore, Torino 2012.

tecnologie informatiche.

L'evoluzione tecnologica del terrorismo si innesta nell'era della transizione digitale, considerata "*la quarta rivoluzione industriale*"[2], che, segna il passaggio da una realtà esistenziale basata su processi manuali e analogici ad un'altra che invece trae energia da una enorme mole di *dati* e che implica la presenza di sviluppatori che sappiano gestire ed analizzare tali dati.

Il continuo ed inarrestabile sviluppo tecnologico ha portato sempre di più a sovrapporre al mondo reale una realtà virtuale nella quale l'essere umano riesce a muoversi e ad interagire con gli altri utenti tramite i suoi avatar e le repliche digitali (digital twin).

A partire dal 1992, a seguito di un popolare romanzo di fantascienza "Snow Crash" di Neil Stephenson, il termine "metaverso" ha voluto rappresentare l'idea dell'esistenza di uno spazio alternativo al mondo reale che, inizialmente, limitato al mondo dei "gamers" e dei video giochi si è poi trasferito in ambiti diversi finendo per avere implicazioni per l'intera società civile e, quindi, interessando anche la materia del diritto penale[3].

In particolare, in ambito medico si è assistito alla trasposizione in ambito virtuale di situazioni cliniche per consentire a studenti e specializzandi di esercitarsi. Analogamente nell'ambito delle attività commerciali sempre di più si sta assistendo alla diffusione delle criptovalute e delle transizioni digitali effettuate con la tecnica del block-chain che permette di certificare e attestare l'esistenza del pagamento.

Con specifico riferimento al diritto penale, l'evoluzione tecnologica impatta sulla materia in questione sia dal punto di

---

[2] *European Parliament resolution of 3 May 2022 on artificial intelligence in a digital age* (2020/2266(INI)). Dopo l'introduzione del vapore, dell'elettricità e dei computer, la rivoluzione digitale anche a livello normativo viene condiserata come la quarta rivoluzione industriale.

[3] F. COPPOLA *Intelligenza artificiale Metaverso il sistema penale: prevenzione, repressione, opportunità, rischi* Wolters Kluwer Cedam 2025 pag.12 ss. In tale ambito ed in tale contesto vengono esaminate tutte le implicazioni di interesse penalistico sia dal punto di vista sostanziale che processuale.

vista sostanziale che da quello processuale implicando una trasformazione del ruolo e dell'attività dell'operatore del diritto, nel cui ambito si colloca il lavoro del giudice, dell'avvocato, del pubblico ministero, dei periti, della polizia giudiziaria e dei cancellieri. Come vedremo nel prosieguo della trattazione, la citata evoluzione tecnologica incide sulle tematiche della tipicità della fattispecie e della responsabilità dell'autore del reato coinvolgendo anche il tema della responsabilità degli enti collettivi; sul versante processuale coinvolge maggiormente il profilo probatorio: ad esempio un digital twin potrebbe essere usato come strumento probatorio e peritale, ad esempio per ricostruire incidenti stradali, scene del crimine o attacchi informatici realizzando una sorta di replica digitale viva che consente al giudice di sperimentare diversi scenari e verificare la fondatezza delle tesi accusatorie o difensive.

Nello scenario sin qui delineato, si collocano i sistemi di IA che possono essere utilizzati come efficace strumento di contrasto alla lotta al crimine ma, allo stesso tempo, possono essere utilizzati dalle stesse organizzazioni terroristiche per commettere reati.

Al riguardo un ruolo fondamentale viene svolto dalle piattaforme digitali, allorquando le stesse vengono sfruttate dalle organizzazioni criminali per finalità di propaganda e di terrorismo psicologico. Si tratta attività che oggi possono essere ancora di più facilitate dalle tecnologie basate sulla IA come avviene nel caso della sovrapposizione di volti, o della modificazione delle espressioni e persino delle imitazioni di voci; tali tecnologie possono essere sfruttate per diffondere contenuti a una molteplicità di individui, velocemente, in modo virale ed a bassi costi creando nuove frontiere per il terrorismo internazionale[4].

[4] Il riferimento è alla problematica dei *deepfake* e dei *political treahts* che più diffusamente verranno analizzata nelle pagine seguenti relativamente alle interferenze con la materia penale ed in particolare con i reati di terrorismo.

A tale area comunemente denominata delle political threats (minacce psicologiche) si possono aggiungere altri due distinti settori in cui il terrorismo può operare utilizzando i sistemi di IA: quello degli attacchi informatici (cyber-threats) e quello degli attacchi fisici (psycal-threats).

Preliminare all'esame di tali questioni risulta la comprensione dei profili evolutivi e definitori della IA sia in ambito scientifico che in quello normativo.

In tale ottica, il moltiplicarsi delle opportunità ricollegabili all'utilizzo delle applicazioni della IA nella quotidianità della vita degli utenti (si pensi ad esempio agli assistenti virtuali nei cellulari o in device casalinghi, ed all'"esplosione" dei chatbot e dell'intelligenza artificiale generativa) impone di non rallentare, o addirittura bloccare, il progresso del settore, ma allo stesso tempo impone di comprenderne i rischi.

Al riguardo, il presupposto di ogni discorso in materia di sviluppo della IA è che la *variabile umana* sarà fondamentale per guidare e utilizzare al meglio le opportunità delle nuove tecnologie evitando i rischi incontrollabili laddove le macchine dovessero essere lasciate a sé stesse.

Procediamo con ordine.

§§§

1.1. Lineamenti evolutivi dell'IA e discipline di riferimento.

Dal punto di vista storico, il 1956 viene considerato l'anno di nascita dell'IA perché durante un famoso seminario estivo tenutosi presso il Dartmouth College di Hanover nel New Hampshire vennero dettate le linee programmatiche della nuova disciplina a partire dalla raccolta dei contributi sviluppati negli anni precedenti e in direzione delle potenzialità future. In tale occasione, John Mc Carthy presentando la sua proposta di studio

sottolineava la capacità simulativa dell'IA, ossia la possibilità che la macchina potesse riprodurre qualsiasi operazione frutto della intelligenza umana, come il linguaggio, la concettualizzazione, il *problem solving e* l'autoapprendimento[5].

L'anno del 1956 costituisce una data convenzionale perché quando si parla di storia dell'intelligenza artificiale si fa riferimento anche alla cibernetica e all'avvento dei primi calcolatori elettronici la cui apparizione è anteriore a tale data. Si citano inoltre Charles Babbage e la sua macchina analitica, Gottfried Wilhelm Leibniz e il suo progetto di meccanizzare la ragione, risalendo fino alla macchina logica di Raimondo Lullo e agli automi semoventi di Erone di Alessandria. In questo contesto l'inquadramento storico della disciplina dell'intelligenza artificiale appare particolarmente problematico in quanto rappresenta il punto di avvio per un'analisi più ampia, nella quale considerare anche questioni legate alla tradizione formalistica di indagine sulla mente, alla nascita delle prime macchine calcolatrici nonché alla tendenza dell'uomo ad auto imitarsi.

Oltre a tali aspetti, per delineare criticamente la storia dell'intelligenza artificiale è importante riconoscere come questa abbia ereditato molte idee, punti di vista e tecniche da altre discipline, in particolare dalla filosofia, dalla matematica e dalla psicologia. Più precisamente, derivano dalla filosofia i risultati relativi al dibattito sulla natura dell'intelligenza e della razionalità; dalla matematica l'approccio formale basato sulla logica; dalla psicologia l'analisi delle relazioni fra conoscenza e azione.

[5] J. MC CARTHY - M. L. MlNSKY - N. ROCHESTER - E.C. SHANNON, *A proposal or the Dartmouth Summer Research Project on Artificial lntelligence,* 31 agosto 1955, disponibile al link http://jmc.stanford.edu/articles/dartnwulh/dartmoujh.pdf: *«The study is to proceed on the basis of the conjecture thai every aspect o/ learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it»*

Tuttavia, è senza dubbio con la cibernetica e l'informatica che queste influenze si fanno più manifeste e aprono la strada alla nascita ufficiale dell'intelligenza artificiale.

All'inizio degli anni Quaranta dello scorso secolo si cominciò a indicare con il termine cibernetica lo studio sistematico dei processi riguardanti la comunicazione e il controllo sia negli animali sia nelle macchine. Warren S. McCulloch e Walter Pitts proposero nel 1943 il primo modello di neuroni artificiali, attingendo alla conoscenza della fisiologia e delle funzioni di base dei neuroni, alla logica proposizionale e alla teoria della computabilità di Alan M. Turing. L'idea alla base del progetto cibernetico fu di studiare i meccanismi dell'autoregolazione e del controllo presenti sia negli organismi viventi sia nelle macchine con retroazione, in grado cioè di rispondere in modo adattativo alle sollecitazioni dell'ambiente modificando il proprio comportamento. Uno dei risultati più significativi consentì di mostrare come ogni funzione calcolabile potesse essere elaborata da una qualche rete di neuroni connessi. A partire da ciò, nel 1949, Donald O. Hebb dimostrò come una semplice regola di aggiornamento per modificare le forze di connessione fra i neuroni potesse dare luogo a processi di apprendimento. Nonostante questi significativi successi iniziali, il destino della cibernetica era segnato: dopo pochi anni (intorno alla metà degli anni Cinquanta del secolo scorso) le risorse furono ormai quasi completamente convogliate sull'intelligenza artificiale, a causa del disinteresse della cibernetica per le crescenti prestazioni dell'informatica e di una limitazione degli obiettivi iniziali. Questa tradizione risorgerà verso la metà degli anni Ottanta con il riemergere, in seno all'intelligenza artificiale, del paradigma delle reti neurali. Infatti, nel 1987 si tiene la prima edizione della conferenza su machine learning, neuroscienze e intelligenza artificiale in seguito nota come neurIPS dove verranno nuovamente affrontate e sviluppate le tematiche in questione.

Come detto, oltre alla cibernetica l'IA implica un riferimento all'informatica.

In particolare, il ragionamento attraverso cui il sistema informatico elabora i dati ha un nome: algoritmo. L'algoritmo insieme alle reti neurali permette alle macchine di apprendere dati, riconoscere schemi, prendere decisioni e persino interagire in modo naturale con gli esseri umani.

Si tratta di una procedura ben definita volta alla trasformazione di dati di *input* in dati di *output*. Nei sistemi dotati di intelligenza artificiale, la macchina è dotata di un algoritmo – il c.d. *metalgoritmo* – capace di costruire da sé nuovi algoritmi e idoneo a definire un nuovo processo di trasformazione, a seconda del problema di volta in volta rappresentato.

La macchina non è capace di porre domande, ma sa costruire risposte, anche attraverso meccanismi ignoti al suo programmatore. È in questo contesto che va inquadrato il fenomeno del *machine learning*, tale per cui la macchina migliora le sue prestazioni grazie all'esperienza; questo dispiega i suoi effetti anche sulla prevedibilità della risposta, essendo plausibile che il sistema operi attraverso procedure non più controllabili dal suo ideatore e non sempre verificabili *ex post*: una macchina che non è capace di conoscere o approfondire il problema, ma è capace, attraverso un'attività puramente computazionale, di risolverlo.

§§§

1.2. La spiegazione dell'IA in ambito scientifico.

Una volta delineati i lineamenti evolutivi della IA e le relative discipline di riferimento, occorre evidenziare come dal punto di vista scientifico il discorso pubblico sulla IA - in un contesto nel quale le questioni dell'apprendimento e del ragionamento delle

macchine sono divenute presenti in qualsiasi dibattito[6] - sia attualmente polarizzato intorno a due macro-questioni: da un lato l'idea che l'avvento dell'IA possa portare all'estinzione del suo creatore cioè l'essere umano; dall'altro l'idea che la sua diffusione come forma di intelligenza generale e superiore, compirà il destino dell'umanità perché porterà al suo superamento[7].

Sul punto, un primo tentativo di spiegazione scientifica è presente nei pionieristici studi di Alan Turing, il quale, tra i primi, si chiese: *«Can machine think?»*[8]

Al fine di indagare la possibilità che anche una macchina potesse "pensare", il celebre matematico inglese ideò un esperimento volto a valutare la sua capacità di esibire un comportamento intelligente, inteso quale funzionamento non agevolmente distinguibile da quello di un essere umano.

Quello che sarebbe poi passato alla storia come "Test di Turing" prendeva ispirazione dal c.d. *imitation game,* ossia un gioco di

---

[6] *Big data*, *IA*, *machine learning*, *cloud* e *distributed ledger technology* sono le nuove tecnologie dell'innovazione digitale che pervadono in profondità la realtà in cui viviamo.

[7] In tale ambito ed in tale contesto si discute, ad esempio, se i nuovi modelli di intelligenza artificiale potranno avere nel prossimo immediato futuro una coscienza. Secondo Demis Hassabis fondatore di Google Deep Mind, i BOT sarebbero pronti a sviluppare un'intelligenza umana. Nel maggio del 2025 Anthropic che insieme a OpenAI e Google guida la rivoluzione di Chatbot ha dichiarato che il suo ultimo modello di intelligenza artificiale "*Claude Opus 4*" ha ricattato i suoi sviluppatori quando gli è stato comunicato che sarebbe stato disattivato. In particolare, durante un test di sicurezza, l'azienda ha chiesto a *Claude Opus 4* di comportarsi come un assistente di una società fittizia, ma poi gli ha fornito accesso ad altrettanto fittizie e-mail che annunciavano la sua sostituzione e che rivelavano anche che l'ingegnere responsabile della decisione aveva una relazione extraconiugale. Secondo Anthropic, *Claude Opus 4* ha minacciato di rivelare la relazione se la sua sostituzione fosse andata avanti. Viene in mente un parallelismo con "Hal 9000" il robot descritto da Arthur Clarke nel libro Odissea nello spazio ripreso nel 1968 da Stanley Kubrick nell'omonimo film. Come "Hal 9000" si ribella di astronauti quando capisce che lo vogliono spegnere, anche il BOT di Anthropic ha cercato di sopravvivere a tutti i costi. In ogni caso, l'esperimento di Claude opus 4 testimonia come le discussioni interne alla capacità dei modelli di comprendere il contesto, vendicarsi, provare sentimenti umani o essere intelligenti come gli esseri umani si stanno concretamente sviluppando. A quel punto si avrebbe il vero cambiamento atteso a livello globale un terremoto non solo per l'economia, ma per la stessa civiltà umana.

[8] A. M. TURING, *Computing machinery and intelligence,* in *Mind,* UX. 1950,433. Per una ricostruzione storica, cfr. G. SARTOR, *L'intelligenza artificiale e il diritto,* Giappichelli, 2022, 23 e ss

società nel quale, in assenza di ogni contatto diretto, uno dei partecipanti interrogava due interlocutori di sesso diverso, con l'obiettivo di individuare chi dei due fosse l'uomo e chi la donna. I due interlocutori avevano l'intuibile compito di dissimulare la propria identità al fine di trarre in inganno l'interrogante.

Trasposto nel contesto sperimentale, Turing ebbe l'intuizione di sostituire uno dei due interlocutori con una macchina, con la seguente considerazione: se le macchine pensano come un uomo, allora l'intervistatore, pur essendo consapevole che uno dei due interlocutori sia una macchina, non dovrebbe essere in grado di discernere in modo affidabile l'essere umano dall'elaboratore, in quanto quest'ultimo si comporterebbe allo stesso modo del1'interlocutore umano nel dissimulare la propria identità.

Le spiegazioni scientifiche di Turing sono state oggetto di vaglio critico in un recente saggio di Coppola[9] laddove è stato condivisibilmente affermato che sebbene i più moderni sistemi di IA, come ad esempio *Chat-GPT* 4.5., abbiano superato il *test* di Turing ciò non può trasformarli automaticamente in degli esseri pensanti in quanto la circostanza che i sistemi di IA possano simulare perfettamente alcune capacità umane non significa che anche dal punto di vista delle conseguenze giuridiche possano essere considerati come tali.

Invero, i descritti tentativi di spiegazione scientifica dell'IA devono tener conto che, anche, in tale ambito non può esistere una unica formula definitoria in quanto la realtà fenomenica ci consegna due distinti modi di operare dei sistemi di intelligenza artificiale sui quali ora soffermeremo la nostra attenzione.

§

1.2.1. IA debole (Narrow IA) e IA forte (AGI).

[9] F. COPPOLA *op.cit.* Wolters Kluwer Cedam 2025 pag.19 ss

La realtà fenomenica ci consente di classificare l'IA in due distinte macrocategorie:

- Narrow IA, meglio nota come IA debole (o ristretta) in quanto specializzata in compiti specifici;
- AGI, meglio nota come IA forte (o generale) in quanto considerata come una macchina con capacità cognitive simili a quelle umane, in grado di apprendere e adattarsi in modo autonomo a nuovi compiti.

In particolare, la Narrow IA è la forma di IA che comunemente viene utilizzata nella nostra società e che, come tale, viene usata per compiti predeterminati (ad esempio svolgere una ricerca, effettuare riconoscimenti vocale etc.). Con la IA debole si intende un modello di funzionamento semplificato del processo decisionale della macchina, tradotto con la formula algoritmica *"if-this-then-that"*. In questo scenario, l'IA riceve tutte le informazioni dall'uomo: dagli *input* agli *output,* ogni decisione della macchina è determinata dal programmatore che ha già previsto le condizioni di azione e le risposte che l'IA deve fornire in determinati contesti. Un recente studio[10] ha analizzato i possibili vantaggi in ambito penalistico per almeno tre ragioni: a) in primo luogo per la capacità di immagazzinare un *dataset di* variabili e di *output* da generare superiore alle capacità di memorizzazione dell'uomo; b) in secondo luogo per la capacità di ricercare le informazioni immagazzinate e fornire l'output programmato con una velocità sconosciuta all'uomo; c) infine per la tendenziale prevedibilità della decisione e la controllabilità dei percorsi decisionali che generano *l'output*.

Passando ora ad esaminare la seconda macrocategoria dell'IA meglio nota come IA forte (o generale da qui il nome in inglese di AGI), la stessa è considerata alla stregua di una macchina con capacità cognitive simili a quelle umane, in grado di apprendere e adattarsi in modo autonomo a nuovi compiti. In questi casi, anziché fornirle tutti gli *input* e i possibili *output,* all'IA si dettano

[10] F. COPPOLA, *op. cit.* pag.17

tre distinti insegnamenti: come apprendere; come classificare; come generalizzare.

Con specifico riferimento alle modalità di addestramento della IA forte, in dottrina vengono individuate tre distinte articolazioni rispettivamente denominate:

- *supervised.* Nell'ambito dell'apprendimento supervisionato, il sistema viene istruito dall'IA *trainer* attraverso una serie di esempi in cui a ciascun caso è associata una risposta corretta così da consentire al modello di dedurre una regola generale applicabile anche a casi nuovi, seppur parzialmente difformi rispetto a quelli utilizzati nella fase di addestramento. In altri termini, al sistema vengono fomiti *input* accompagnati dai corrispondenti *output,* al fine di permettergli di individuare la relazione che li lega e costruire una funzione predittiva riutilizzabile in contesti analoghi;
- *unsupervised.* Diversamente da quanto accade nel supervised, nell'ambito dell'apprendimento non supervisionato, al sistema viene somministrato esclusivamente un insieme di dati privi di etichettatura, senza alcuna indicazione circa *l'output* atteso: l'obiettivo è quello di individuare, in via autonoma, strutture logiche latenti e *pattern* ricorrenti all'interno degli *input* forniti. La tecnica di riferimento in tale ambito è il *clustering,* che consente al sistema *di* raggruppare *gli* elementi sulla base *di* similarità o connessioni interne rilevanti, secondo criteri elaborati in modo autonomo. Un'ulteriore metodologia è quella dell*'association learning,* volta a rilevare regole che descrivano porzioni significative del *dataset* attraverso l'individuazione di correlazioni statistiche, associazioni frequenti o ricorrenze rilevanti;
- *reinforcement learning*. L'autoapprendimento rinforzato si distingue, invece, per una dinamica fondata non sull'insegnamento diretto, bensì sull'interazione tra il

sistema e l'ambiente: l'agente artificiale apprende dalle conseguenze delle proprie azioni, ricevendo ricompense o penalità in base al grado di successo o di insuccesso rispetto ad un obiettivo prestabilito. Come opportunamente osservato in dottrina il sistema apprende dei risultati delle azioni proprie o altrui: è in grado di distinguere successi e fallimenti a seconda di come le azioni incidano sul raggiungimento delle utilità o valori da esse perseguite. L'intero processo addestrativo si compone di due fasi distinte: il *training dataset* finalizzato al perfezionamento iterativo del sistema e il *testing dataset* che verifica la capacità della IA di generalizzare le competenze acquisite a dati nuovi e non precedentemente analizzati. Solo al termine di questo percorso e una volta raggiunto un livello prestazionale soddisfacente il sistema può essere implementato in contesti operativi reali.

Nell'ambito di tale tematica si colloca l'intelligenza artificiale generativa in quanto tipica espressione dei sistemi c.d. "sapienti", propri della cosiddetta *Strong AI,* i quali non si limitano a emulare le capacità cognitive umane, ma si contraddistinguono per la loro attitudine a svilupparle autonomamente. I sistemi in questione si fondano sul principio del *self- learning,* ossia sull'apprendimento automatico derivante da dati ed esperienze pregresse, che consente loro di migliorare progressivamente attraverso l'analisi di nuove fonti informative e l'osservazione degli *output* generati. In particolare, a partire da un *dataset* iniziale, il sistema è in grado di evolversi in base alle singole interazioni, dando origine a uno spettro potenzialmente infinito di comportamenti, spesso imprevedibili anche per i programmatori. In tale prospettiva si inserisce il *deep learning,* sottoinsieme del *self-learning* basato sull'impiego di reti neurali artificiali, strutture computazionali capaci di riprodurre il comportamento dei neuroni biologici attraverso connessioni interne e segnali digitali. Tali reti sono

articolate in più "strati" che elaborano progressivamente le informazioni mediante pesi interni modificabili attribuendo rilevanza differente ai segnali ricevuti. Attraverso il processo di elaborazione stratificato, la rete costruisce rappresentazioni sempre più raffinate dei dati iniziali, e, qualora l'*output* risulti errato, attiva un meccanismo di *backpropagation* (retro-propagazione dell'errore) che consente di correggere i parametri interni. Sul punto come coerentemente evidenziato in dottrina[11], man mano che le informazioni percorrono in profondità diversi strati della rete neurale, l'elaborazione che ricevono diventa via via meno trasparente e controllabile.

Tra le forme più note di sperimentazione del *deep learning,* figurano i *Large Language Models (LLM),* i quali sfruttano il *Natural Language Processing* (LNP) per emulare il linguaggio umano contestualizzarlo e generare un testo coerente e specifico rispetto alle richieste dell'utente. *Chat-GPT* di *OpenAI* è un esempio dell'evoluzione di tali modelli. In tale ambito si collocano anche altri modelli quali PaLM (Google). LLaMA (Meta) e Claude (Anthropic).

Essendo capace di creare autonomamente gli *output,* tramite la combinazione originale di una sconfinata mole di dati su cui si è addestrata e su cui continua ad addestrarsi man mano che interagisce con l'utente esterno, l'IA generativa è massimamente performante, ma anche altamente opaca e imprevedibile. Tale per cui in dottrina si è autorevolmente parlato di *black box* e di "oracolo" algoritmico.

Per tali ragioni si discute l'inclusione dei modelli Large *Language Models (LLM)* nella categoria dei rischi sistemici potendosi definire tale un rischio quando non riguarda solo singoli utenti o casi isolati, ma può avere conseguenze diffuse e gravi sull'intero sistema economico, politico o sociale. Al riguardo possibili rischi sistemici possono essere la disinformazione su larga scala così

[11] F. COPPOLA *op.cit*. In particolare, cfr. pag. 7 e ss. dove vengono sviluppate le argomentazioni in questione

come la generazione di testi credibili ma falsi usati per manipolazione politica, propaganda o truffe.
In conclusione, i sistemi di IA forte-generativa presentano il notevole vantaggio di una elaborazione e analisi dei dati di tipo creativo, che si avvicina a quella dell'uomo ed è capace di cogliere il contesto e adattare *l'output* alle variabili presenti. Di contro, quanto più si arricchisce di dati, tanto più *l'output* dell'IA diventa impenetrabile, imprevedibile e potenzialmente ingestibile dall'uomo.

§§§

1.3. IA in ambito normativo: tra eccessi regolatori e definizioni di tipo elastico.

Gli scenari di progresso descritti e lo sviluppo economico-industriale hanno reso necessario un intervento regolatorio sui sistemi di IA che si è sviluppato con modalità differenti tra loro.
Come meglio verrà approfondito nel capitolo terzo l'UE ha adottato un intervento di carattere generale attraverso il Regolamento EU 2024/1689 del Parlamento europeo e del Consiglio europeo del 13 giugno 2024 c.d. AI ACT[12] .

[12] AI ACT, Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio del 13 giugno 2024 entrato in vigore nell'agosto del 2025 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale), pubblicato sulla Gazzetta Ufficiale dell'Unione Europea Serie L del 12 luglio 2024. Sulla AI ACT ci soffermeremo nel capito tre, evidenziando sin da ora come lo stesso rappresenti il primo tentativo a livello globale di regolamentare questa tecnologia emergente. In particolare, l'articolo 113 stabilisce che: "*Il presente regolamento entra in vigore il ventesimo giorno successivo alla pubblicazione nella Gazzetta ufficiale dell'Unione europea. Si applica a decorrere dal 2 agosto 2026. Tuttavia: I capi I e II si applicano a decorrere dal 2 febbraio 2025; Il capo III, sezione 4, il capo V, il capo VII, il capo XII e l'articolo 78 si applicano*

In tal modo, l'Europa ambisce al ruolo di *first mover* regolatorio digitale, già inaugurato con il Regolamento Generale sulla protezione dei dati personali (GDPR)[13], che conferma la volontà di costruire un modello giuridico ispirato alla compatibilità tra libertà economiche e diritti fondamentali e destinato a diventare un riferimento obbligato per le imprese tecnologiche internazionali operanti sul mercato UE, tra cui ovviamente rientrano le c.d. big-tech. In realtà si tratta di una ambizione non semplice da realizzare trattandosi di un articolato normativo eccessivamente regolatorio che deve confrontarsi: da un lato con gli USA che vogliono assumere l'assoluta leadership della materia in questione ispirandosi ad una autoregolazione del mercato; dall'altro con la CINA che, invece, tende ad adottare interventi normativi dal carattere dirigista.

Rinviando gli approfondimenti di tali argomenti al capitolo terzo, in questa sede è doveroso anticipare come i capisaldi dell'AI ACT sono essenzialmente quattro:

- l'approccio antropocentrico fondato sulla dignità della persona;
- la gradualità del rischio dei sistemi di intelligenza artificiale;
- l'ampia portata materiale e territoriale del regolamento europeo;
- l'elencazione dei sistemi ad alto rischio.

Nonostante le descritte ambizioni regolatorie, le difficoltà di regolamentare una materia in continua evoluzione si colgono a partire dall'approccio definitorio. Infatti, la scelta dell'atto normativo in questione è quella di adottare una definizione di tipo

a decorrere dal 2 agosto 2025, ad eccezione dell'articolo 101; L'articolo 6, paragrafo 1, e i corrispondenti obblighi di cui al presente regolamento si applicano a decorrere dal 2 agosto 2027"

[13] 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 e alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (Regolamento generale sulla protezione dei dati - GDPR).

elastico e non rigido, per meglio consentire un adattamento alle possibili evoluzioni dei sistemi di I.A. senza dover ritoccare di volta in volta le premesse definitorie.

Sul punto, l'art.3, punto 1), del regolamento in questione stabilisce che:

- per sistema di intelligenza artificiale si intende un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali;
- per dato deve intendersi qualsiasi rappresentazione digitale di atti, fatti o informazioni e qualsiasi raccolta di tali atti, fatti o informazioni, anche sotto forma di registrazione sonora, visiva o audiovisiva;
- per modelli di intelligenza artificiale devono intendersi modelli che identificano strutture ricorrenti attraverso l'uso di collezioni di dati, che hanno la capacità di svolgere un'ampia gamma di compiti distinti e che possono essere integrati in una varietà di sistemi o applicazioni.

In tal modo, la descritta formulazione stabilisce con chiarezza una stretta correlazione tra la normativa in questione e quella relativa al trattamento dei dati personali. Tale correlazione si estende anche a quella prevista per le piattaforme digitali[14] nel momento in cui i dati personali vengono fatti circolare sul web e che impone alle piattaforme obblighi di rimozione in caso di contenuti illeciti.

In Italia, nel descritto ambito definitorio si è incuneato anche il ddl 1146-b di iniziativa governativa approvato dal Senato della Repubblica il 20 marzo del 2025 e successivamente modificato

[14] La normativa sugli obblighi delle piattaforme digitali è contenuta nel Digital Service Act (meglio noto come DSA) Regolamento U.E. 2022/2065. Per gli approfondimenti sulla correlazione tra normative si rinvia al capitolo terzo, par. 1.4.

dalla Camera dei deputati in data 25 giugno del 2025. Sul punto nella sua attuale formulazione l'art.2 stabilisce che:

- per sistema di intelligenza artificiale deve intendersi il sistema definito dall'articolo 3, punto 1), del regolamento (UE) 2024/1689;
- per dato deve intendersi qualsiasi rappresentazione digitale di atti, fatti o informazioni e qualsiasi raccolta di tali atti, fatti o informazioni, anche sotto forma di registrazione sonora, visiva o audiovisiva attraverso la conoscenza approfondita della disciplina in materia di protezione dei dati personali;
- per modelli di intelligenza artificiale devono intendersi i modelli definiti dall'articolo 3, punto 63), del regolamento (UE) 2024/1689.

Il chiaro riferimento all'art. 3, punto 1) dell'AI ACT testimonia come anche il legislatore italiano non voglia rimanere imbrigliato in una definizione di IA rigida tale da impedirne un adeguamento all'evoluzione dei modelli in questione.

§§§

1.4. L'approccio antropocentrico nell'AI ACT e nelle altre fonti sovranazionali.

Come sopra anticipato, il primo caposaldo dell'AI ACT è costituito dall'approccio antropocentrico fondato sulla dignità della persona.

Tale approccio, già affermato nelle considerazioni introduttive del regolamento in questione, intende dare centralità al controllo umano sulle decisioni automatizzate considerando l'essere umano come un baluardo della autodeterminazione informativa in contrasto a pratiche manipolatorie.

La conseguenza di tali assunti è che la rapida evoluzione delle tecnologie di IA, che contribuisce al conseguimento di un'ampia

gamma di benefici a livello economico, ambientale e sociale comprensivo dell'intero spettro delle attività industriali e sociali, non può sostituirsi alle decisioni dell'uomo.

Si tratta di un principio che a livello europeo trova un antecedente nella Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi, adottata in occasione della trentunesima Assemblea plenaria della Commissione Europea per l'efficienza della giustizia Cepej tenutasi a Strasburgo il 3-4 dicembre 2018. In tale contesto era già stata esplicitamente riconosciuta l'importanza della intelligenza artificiale nelle nostre moderne società e dei suoi potenziali benefici al servizio della efficienza e qualità della giustizia. Tuttavia, con riferimento allo specifico aspetto del *decision-making process* era stato chiaramente evidenziato come i sistemi di IA non potessero mai "blindare" la decisione della magistratura. Difatti, *nell'Appendice II* della Carta etica veniva espressamente incoraggiato l'utilizzo dell'IA per offrire ai giudici un ventaglio di informazioni quantitative e qualitative maggiormente dettagliato. Al contrario, veniva respinta con forza qualsiasi proposta di impiego dei sistemi di IA che vedesse il giudice subire una determinata decisione giudiziaria *made by A.I.,* pretendendo di vincolarlo alla stessa.

L'approccio antropocentrico di cui stiamo discutendo è stato successivamente ribadito, in data 8 aprile 2019, in occasione dell'approvazione delle Linee guida etiche per una intelligenza artificiale affidabile, aventi funzione uniformatrice e preparatoria rispetto ai principi applicabili all'uso dei sistemi di intelligenza artificiale. Anche in tale circostanza, veniva con chiarezza ribadita la centralità del controllo umano sulle decisioni automatizzate e contestualmente affermati i principi di trasparenza, robustezza, qualità e protezione dei dati trattati.

Infine, tali principi sono stati definitivamente ripresi e sviluppati in occasione del citato regolamento AI ACT laddove viene affermato che:

- l’IA consiste in una famiglia di tecnologie in rapida evoluzione che contribuisce al conseguimento di un’ampia gamma di benefici a livello economico, ambientale e sociale comprensivo dell’intero spettro delle attività industriali e sociali;
- l’uso dell’IA garantendo un miglioramento delle previsioni, l’ottimizzazione delle operazioni e dell’assegnazione delle risorse e la personalizzazione delle soluzioni digitali disponibili per i singoli e le organizzazioni, può fornire vantaggi competitivi fondamentali all’imprese e condurre a risultati vantaggiosi sul piano sociale e ambientale, ad esempio in materia di assistenza sanitaria, agricoltura, sicurezza ambientale, istruzione e formazione, media, sport, cultura, gestione dell’infrastrutture, energia, trasporti e logistica, servizi pubblici, sicurezza, giustizia, efficienza dal punto di vista energetico e delle risorse, monitoraggio ambientale, conservazione e ripristino della bio-diversità e degli ecosistemi, mitigazione di cambiamenti climatici e adattamento ad essi.

Da tali considerazioni discende chiaramente come l’impostazione di fondo sia quella di considerare l’intelligenza artificiale una tecnologia potente che, se regolata correttamente, può portare enormi benefici a tutta l’umanità ma che allo stesso tempo può essere utilizzata solamente sotto il controllo dell’uomo in modo tale da garantirne il rispetto dei diritti fondamentali, della democrazia e dello Stato di diritto.

Stante la specifica attinenza al tema trattato, in questa sede meritano, di essere richiamate una serie di ulteriori fonti sovranazionali che confermano il rifiuto di un integrale sostituzione del decisore umano con sistemi di intelligenza artificiale.

In primo luogo, l'art. 22 del Regolamento Generale sulla Protezione dei Dati Personali (2016/679) sancisce in termini

generali il diritto dell'interessato "di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona".

Con specifico riferimento al trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, nonché alla libera circolazione di tali dati, l'art. 11 della Direttiva (UE) 2016/680 prevede il generale divieto di profilazione, salvo deroghe presidiate da specifiche garanzie.

La disposizione stabilisce che "gli Stati membri dispongono che una decisione basata unicamente su un trattamento automatizzato, compresa la profilazione, che produca effetti giuridici negativi o incida significativamente sull'interessato sia vietata salvo che sia autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento e che preveda garanzie adeguate per i diritti e le libertà dell'interessato, almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento".

Tale previsione è stata recepita nell'ordinamento interno dall'art. 8 del d.lgs. 18 maggio 2018, n. 51, che ricalca pressoché fedelmente il testo europeo e conferma l'eccezionalità delle decisioni fondate su trattamenti automatizzati in assenza di specifiche garanzie per l'interessato. In tale ambito rientrano le possibilità di utilizzare i sistemi di riconoscimento facciale e di polizia predittiva in conformità con le disposizioni europee, tema sul quale ci soffermeremo nel prosieguo.

Da ultimo, l'art.56 del Regolamento (UE) 2017/1939, che disciplina l'attuazione di una cooperazione rafforzata sull'istituzione della Procura europea (EPPO), precisa che l'interessato ha diritto di non essere sottoposto a una decisione dell'EPPO basata unicamente sul trattamento automatizzato.

Il combinato disposto delle fonti sopra indicate esprime un deciso rifiuto verso forme di automazione decisionale in ambito penale, riaffermando invece con forza il ruolo imprescindibile dell'intervento umano quale presidio di legalità della decisione, responsabilità e tutela dei diritti fondamentali della persona.

§§§§§

2.L'utilizzo dell'intelligenza artificiale come opportunità.

I tentativi definitori dell'IA, sia dal punto di vista scientifico che normativo, svolti nei paragrafi precedenti hanno permesso di evidenziare l'importanza strategica dell'IA nel mondo in cui viviamo. In un recente saggio[15] l'IA viene considerata l'invenzione dell'umanità perché oltre alle influenze ed ai risvolti che può avere sulla geopolitica è in grado di contribuire al conseguimento di un'ampia gamma di benefici a livello economico, ambientale e sociale comprensivo dell'intero spettro delle attività industriali e sociali.

Si tratta di affermazioni che trovano riscontro in ambito europeo nel citato regolamento AI ACT e che sul versante italiano vengono oggi ribadite nel citato DDL 1146-B laddove viene evidenziato che l'IA può rappresentare un'opportunità rivoluzionaria. A tale affermazione di principio viene fatta seguire l'individuazione dei settori ritenuti strategici per lo sviluppo del Paese, tra cui vengono specificamente annoverati: la sanità; il lavoro; la giustizia e le professioni intellettuali; la disabilità; la pubblica amministrazione. Si tratta di settori, in relazioni ai quali maggiormente si accentua la possibilità di

[15] A. ARESU *Geopolitica dell'intelligenza artificiale* Feltrinelli 2024

sfruttare tutte le opportunità legate alle nuove tecnologie. In tali contesti, l'IA è vista come uno strumento che coadiuva le decisioni umane senza sostituirle, promuovendo lo sviluppo di sistemi comprensibili e tecnologicamente avanzati

Con specifico riferimento ad uno degli indicati settori ritenuti strategici quello della giustizia, occorre sin da ora evidenziare che nella materia penale tali tecnologie possono essere utilizzate da un lato per finalità "maligne" da parte delle organizzazioni criminali e dall'altro come un formidabile strumento nell'attività di contrasto e di prevenzione dei reati.[16]

Si tratta di temi che meglio verranno approfonditi nel capitolo secondo con specifico riferimento alla materia del terrorismo, ma che preliminarmente in questa sede impongono una analisi delle modalità con cui tali opportunità possono realizzarsi sia sul versante giudicante che su quello investigativo.

§§§

2.1. Le opportunità dell'IA nell'attività giudiziaria e l'utilizzo di modelli di cooperazione algoritmica nel sentencing.

Una recente rassegna sul "diritto digitale guidato dall'intelligenza artificiale" riassume i vantaggi e le opportunità dell'IA che nell'ambito dello svolgimento dell'attività giudiziaria può offrire al lavoro svolto dai magistrati, dal personale ausiliario, dagli avvocati, dai periti, dagli investigatori e da chiunque in tale contesto risulta coinvolto negli usi applicativi della IA, la possibilità di analizzare enormi volumi di dati, riconoscere schemi complessi e persino prendere decisioni autonome. Ciò in

[16] C.O. ONORATI, *I.A., politica e reati contro la personalità dello Stato*, in *Sistema penale*, 13 giugno 2022, all'interno della raccolta degli atti del workshop della Fondazione Occorsio su *Intelligenza artificiale e giurisdizione penale.* pag 105

quanto, l'IA ha aperto scenari di progresso impensabili fino a pochi decenni fa, promettendo efficienza, innovazione e soluzioni a problemi di lunga data e contribuendo nei più disparati settori ad un efficace progresso economico, culturale e sociale.

In particolare, le enormi potenzialità della IA in termini di opportunità si traducono nella riduzione del carico di lavoro degli uffici giudiziari e nell'aumento della velocità dei tempi della giustizia coinvolgendo i seguenti campi: raccolta e analisi di big data; automatizzazione di compiti ripetitivi attraverso procedure con o senza controllo dell'operatore umano; rilevamento di malware o di disfunzioni nei sistemi; ottimizzazione delle reti organizzative e della condivisione di pratiche; generazione di testi; protezione, o indebita appropriazione, dei dati.

Si tratta di innovazioni che consentono di migliorare i metodi forensi tradizionali e che, come tali, si traducono in una formidabile attività di supporto, e non di sostituzione, sia all'attività del pubblico ministero e della polizia giudiziaria consentendo ad entrambi di accedere con maggiore facilità alla ricerca e all'esame di una enorme mole di dati rispetto a quanto avveniva nel passato con internet.

Allo stesso tempo ne risulta facilitata anche l'attività dell'organo giudicante al fine di garantire uniformità di giudizio nell'ambito della c.d. giustizia predittiva[17].

Si tratta di situazioni che ovviamente possono rendere ancora più efficace la lotta al crimine e per quel che a noi interessa l'attività di contrasto alle organizzazioni terroristiche.

Tali aspetti introducono la necessità di stabilire se l'efficienza complessiva dei sistemi giudiziari possa spingersi fino alla sostituzione dell'agente artificiale al decisore umano, situazione ipoteticamente ipotizzabile nel caso in cui il giudice dovesse subire una determinata decisione giudiziaria *made by A.I.,* rimanendo vincolato alla stessa.

[17] R. T. YADAV, *AI-Driven Digital Forensics. International Journal of Scientific Research & EngineeringTrends*, Vol. 10 (2024), Issue 4, pp. 1673-1681.

Il descritto approccio antropocentrico dell'AI ACT e degli atti sovranazionali collegati, scongiura tale possibilità rendendo insostituibile il ruolo del decisore umano.

Al riguardo, in ambito interno con riferimento al settore dell'attività giudiziaria la relazione di accompagnamento al DDL 1146-B[18] sottolinea come il settore in questione costituisca una

[18] In ambito interno, il già citato DDL 1146-B ribadisce che l'IA può rappresentare un'opportunità rivoluzionaria. A tale affermazione di principio viene fatta seguire l'individuazione dei settori ritenuti strategici per lo sviluppo del Paese, tra cui vengono specificamente annoverati: la sanità; il lavoro; la giustizia e le professioni intellettuali; la disabilità; la pubblica amministrazione. Si tratta di settori, in relazioni ai quali maggiormente si accentua la possibilità di sfruttare tutte le opportunità legate alle nuove tecnologie. Scorrendo rapidamente le disposizioni in questione, occorre sottolineare come, sul versante della sanità e della disabilità, l'articolo 7, comma quinto, del ddl 1146-B prevede che l'utilizzo dei sistemi di AI in ambito sanitario costituisce un supporto nei processi di prevenzione, diagnosi e cura lasciando tuttavia in pregiudicata la spettanza della decisione alla professione medica. Al riguardo particolari rilevanza assume anche la creazione di una piattaforma di intelligenza artificiale ad opera dell'Agenas per il supporto e le finalità di cura comprensiva dell'assistenza territoriale. Sul versante del lavoro, l'articolo 10 del ddl in questione chiarisce che: l'intelligenza artificiale può essere impiegata per migliorare le condizioni di lavoro, tutelare l'integrità psicofisica dei lavoratori, accrescere la qualità delle prestazioni lavorative e la produttività delle persone in conformità al diritto dell'Unione Europea; l'utilizzo dell'intelligenza artificiale in ambito lavorativo deve essere sicuro, affidabile, trasparente e non può svolgersi in contrasto con la dignità umana, né violare la riservatezza dei dati personali; che il datore di lavoro il committente è tenuto a informare il lavoratore dell'utilizzo dell'intelligenza artificiale nei casi e con le modalità di cui all'articolo 1-bis del d.lgs. 26 maggio 1997, numero 152. A sua volta l'articolo 11 istituisce presso il ministero del lavoro delle politiche sociali, un osservatorio sull'adozione dei sistemi di IA. Sul versante delle professioni intellettuali, per la rilevanza che le stesse assumono nell'economia del tessuto sociale italiano, viene stabilito che l'intelligenza artificiale non può snaturare la funzione e minare il rapporto di fiducia tra cliente e professionista. Pertanto, si stabilisce che nelle professioni intellettuali, il pensiero critico umano debba risultare prevalente rispetto all'uso degli strumenti di intelligenza artificiale che può riguardare solo le attività di supporto all'attività professionale. Per assicurare il rapporto fiduciario tra professionista e cliente si è stabilito inoltre che le informazioni relative ai sistemi di intelligenza artificiale utilizzati dal professionista debbano essere comunicate al cliente con linguaggio chiaro, semplice ed esaustivo. Sul versante dell'attività della pubblica amministrazione, l'articolo 13 regolarizza l'utilizzo della intelligenza artificiale, prevedendo in primo luogo la promozione dell'utilizzo della IA nella pubblica amministrazione per garantire il buon andamento e l'efficienza dell'attività amministrativa; inoltre introduce la centralità del principio dell'autodeterminazione e della responsabilità umana stabilendo che l'utilizzo in funzione strumentale all'attività provvedimentale non fa venir meno la responsabilità e il potere decisionale della persona. Sul punto si prevede che le pubbliche

attività strategica assai delicata proprio perché può sostituirsi al decisore umano. Per tali ragioni, il testo in questione nel sottolineare le opportunità si preoccupa dei rischi connessi all'utilizzo dell'IA nel settore giustizia.

Infatti, accanto alle indicate opportunità convivono gli enormi rischi connessi all'utilizzo dei sistemi di IA in questione a partire da quello di affidarsi ad una tecnologia priva di una coscienza critica che è l'essenza del pensiero umano e quindi dell'argomentazione giuridica posta alla base della decisione che per sua funzione riguarda beni giuridici costituzionalmente rilevanti. Ognuno di questi aspetti comporta problemi diversi: tecnici (di usabilità, di accettabilità, di generalizzabilità a contesti diversi), etici e normativi sui quali meglio ci soffermeremo nel prosieguo.

L'importanza dei temi connessi all'attività giudiziaria spiega perché anche il disegno di legge in questione non discostandosi dagli atti sovranazionali abbia voluto dare centralità al valore umano.

In dottrina[19], muovendo dai principi contenuti nella citata Carta etica, è stato correttamente ribadito che l'operatore del diritto coinvolto negli usi applicativi della IA deve essere messo nelle condizioni di controllare l'agente artificiale, ciò al fine di rendere gli scopi e gli esiti compatibili con il funzionamento e il benessere della comunità sociale.

Sul punto, deve essere segnalato come l'articolo 15 del DDL 1146-B assegna al Ministro della giustizia, nell'elaborazione delle linee programmatiche sulla formazione dei magistrati il compito di promuove attività didattiche sul tema dell'intelligenza artificiale e sugli impieghi dei sistemi di intelligenza artificiale nell'attività giudiziaria, finalizzate alla formazione digitale di

amministrazioni siano tenute a garantire autonomia decisionale e responsabilità favorendo anche la formazione dipendenti pubblici.

[19] S.DI NUOVO, *L'intelligenza artificiale tra percezioni soggettive e regolazioni normative*, in Giustizia insieme, maggio 2025, pag.2

base e avanzata, all'acquisizione e alla condivisione di competenze digitali, nonché alla sensibilizzazione sui benefici e rischi, anche nel quadro regolatorio di cui ai commi 2 e 3 del presente articolo. Per le medesime finalità la norma in questione assegna al Ministro il compito di curare altresì la formazione del personale amministrativo.

L'utilizzo corretto dei sistemi di IA in ambito giudiziario può portare opportunità e vantaggi per realizzare uniformità di giudizio nelle decisioni ed ancor di più sul versante dell'applicazione della pena rispetto al quale per superare le criticità che affliggono l'attuale sistema di commisurazione della pena, nell'ambito delle attività formative promosse dalla Scuola Superiore della Magistratura, è stato recentemente sperimentato un modello di cooperazione algoritmica nel *sentencing* denominato *ExAequo* [20].

In particolare, l'immaginata interazione tra IA-*sentencing* relativamente alla commisurazione della pena si articola su tre fasi:

- raccolta dati e programmazione dell'algoritmo;
- elaborazione e ricerca algoritmtica;
- commisurazione umana.

In dottrina è stato condivisibilmente evidenziato come la progressione logica delineata non sia frutto del caso, ma risponde all'esigenza - più volte sottolineata - di garantire che sia sempre la decisione umana a chiudere il cerchio così da preservare il primato dell'uomo sulla macchina e prevenire ogni deriva eteronomica nella funzione giudicante. La ricerca in *Ex-Aequo* testimonia plasticamente come il sistema di IA debba solo fornire al giudice un patrimonio informativo che costui potrà considerare per fondare la propria, autonoma decisione.

Conclusivamente l'evoluzione e l'interazione tra le capacità dell'IA e le competenze umane impone ai professionisti del diritto

[20] F. COPPOLA. *Commisurazione della pena e intelligenza artificiale: una ipotesi di lavoro con l'algoritmo Ex-Aequo,* in *Arch. pen.,* 2/2023

la necessità di rimanere vigili nell'affrontare le sfide associate, assicurando il necessario equilibrio tra il progresso tecnologico e il mantenimento di standard etici.

§

2.1.1. L'utilizzo eticamente sostenibile dell'IA: le raccomandazioni della Carta etica europea.

Al riguardo, la citata Carta etica europea[21], si propone di dettare una regolamentazione di soft law rivolta a tutti gli operatori coinvolti a diverso titolo nella progettazione e nell'impiego dei sistemi di intelligenza artificiale in ambito giudiziario.
Essa si indirizza altresì ai legislatori chiamati a stabilire una cornice normativa all'interno della quale tali strumenti vanno sviluppati verificati e utilizzati.
Già nella introduzione della Carta in questione emerge un impiego favorevole all'impiego della IA a condizione che ciò si traduca in uno strumento rafforzato dell'efficienza e della qualità del servizio giustizia nel pieno rispetto dei principi sanciti dalla Convenzione europea dei diritti dell'uomo, dalla convenzione per la protezione dei dati personali e della stessa carta etica. In tale prospettiva, l'impiego di strumenti di IA a supporto del processo decisionale della magistratura viene ritenuto funzionale a promuovere la prevedibilità delle decisioni giudiziarie e la coerenza del giudizio.
Tuttavia, particolare cautela viene espressamente raccomandata nell'ambito della giustizia penale. Sul punto, al fine di orientare in senso *eticamente sostenibile* l'utilizzo dell'IA nei sistemi

[21]Come detto in precedenza la Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi è stata adottata in occasione della trentunesima Assemblea plenaria della Commissione Europea per l'efficienza della giustizia Cepej tenutasi a Strasburgo il 3-4 dicembre 2018 definisce l'intelligenza artificiale come l'insieme dei metodi scientifici, teorie e tecniche finalizzate a riprodurre mediante le macchine le capacità cognitive degli esseri umani

giudiziari e di garantire la piena conformità ai cardini del giusto processo la Carta individua alcune linee guida fondamentali enunciando i seguenti principi:

- rispetto dei diritti fondamentali. Tale principio impone che tanto nella fase di progettazione quanto in quelle applicativa i sistemi di IA devono essere compatibili con le garanzie di rango convenzionale, tra cui quelle previste nella Convenzione europea dei diritti dell'uomo (CEDU) e nella Convenzione sulla protezione delle persone secondo un modello definito come *human-rights-by-design*. Si tratta di principi che per la loro ampiezza si rivolgono anche a soggetti estranei al circuito dei giuristi e che devono guidare l'intero ciclo di progettazione dei *software,* In tale ottica, si richiede di concepire e addestrare i sistemi di I.A. in modo tale da prevenire ogni possibile violazione dei diritti fondamentali connessi all'amministrazione della giustizia tra cui: il diritto di accesso alla giurisdizione; il diritto ad un equo processo, nelle sue componenti essenziali del contraddittorio e della parità delle armi; il principio di legalità; l'indipendenza della magistratura;
- canone di non discriminazione. Esso muove dalla consapevolezza che i sistemi di IA, attraverso l'analisi e la classificazione dei dati relativi a individui o gruppi, possono riflettere e persino amplificare discriminazioni già esistenti. Si vieta, dunque, espressamente, che tali tecnologie generino o aggravino diseguaglianze di trattamento;
- qualità e la sicurezza dei sistemi. Al riguardo la Carta raccomanda l'utilizzo di fonti certificate, nonché l'adozione di modelli sviluppati in un ambiente tecnologico sicuro. Il principio attiene, dunque, tanto alla provenienza dei dati - con particolare riferimento alle decisioni giudiziarie quanto all'integrità del processo di

elaborazione, imponendo la tracciabilità ex *post* dei passaggi logici, nonché la conservazione sicura dei modelli e degli algoritmi impiegati;

- trasparenza, imparzialità ed equità. In questo caso si tratta di un principio che mira a garantire l'accessibilità e la comprensibilità dei processi computazionali, affinché essi siano sottoponibili a verifica esterna. La Carta in proposito auspica la pubblicazione dei codici sorgente unitamente a una spiegazione, in linguaggio chiaro e accessibile, delle modalità di funzionamento e del margine d'errore; in alternativa, propone l'istituzione di Autorità indipendenti preposte alla valutazione e certificazione, *ex ante* e periodica, dei sistemi utilizzati in ambito giudiziario. Come rilevato in dottrina si tratta come è evidente, di un tentativo di bilanciare l'esigenza di tutela della proprietà intellettuale e/o del segreto industriale con quella di rendere verificabili le tecnologie impiegate nei sistemi giudiziari;
- *under user control* che implica il controllo da parte dell'utilizzatore dei sistemi di IA. Tale principio esclude ogni forma di automatismo prescrittivo nell'impiego dell'IA e riafferma la centralità dell'utente quale soggetto informato e pienamente consapevole delle proprie scelte. La nozione di utente è, peraltro, polisemica: può riferirsi tanto all'operatore giuridico che utilizza il servizio, quanto al destinatario della decisione. Con riferimento all'operatore giuridico, il principio restituisce il controllo al giudice persona, che potrà fare un uso sapiente della previsione elaborata dalla macchina e disattenderla ove necessario, tenendo in considerazione le peculiarità della fattispecie concreta. Quanto al secondo destinatario della decisione, il principio rileva in funzione garantista traducendosi nel diritto all'informazione sulle alternative disponibili, nel diritto alla consulenza legale e,

soprattutto, nel diritto a un ricorso effettivo dinanzi a un giudice ai sensi dell'art. 6 CEDU.

Nel complesso, il documento manifesta un atteggiamento di fiducia nei confronti degli strumenti di IA, in grado di aumentare l'efficienza complessiva dei sistemi giudiziari nel rispetto dei principi sopraindicati e vietando l'impiego di sistemi che vedano il giudice subire una determinata decisione giudiziaria *made by A.I.,* pretendendo di vincolarlo alla stessa.

§§§

2.2. Le opportunità dell'IA nell'attività di polizia giudiziaria: dalla prevenzione tradizionale alla predictive policing.

L'utilizzo massiccio dei sistemi di IA per scopi di *law enforcement* nello svolgimento dell'attività del pubblico ministero e della polizia giudiziaria, nel cui ambito rientra la c.d. polizia predittiva[22], rappresenta una realtà che sempre di più si sta consolidando e che si traduce in una rilevante opportunità per una più efficace lotta al crimine attraverso le seguenti modalità: cooperazione tra autorità; sinergia tra i software applicati di IA e la lotta al terrorismo; individuazione delle potenziali aree criminali e dei possibili autori di crimini efferati tra cui gli autori di reati terroristici.

Come diffusamente avremo modo di esaminare nel capitolo secondo, in tale ambito rientrano i sistemi predittivi *placed-based e person-based* che sono ammessi dal regolamento AI ACT

[22] Con particolare riferimento all'attività di "*predictive policing*" relativamente alla prevenzione dei reati in materia di terrorismo si rinvia alle considerazioni successive nonché a quelle che verranno sviluppate nel capitolo secondo.

sebbene costituiscano sistemi ad alto rischio attesa la necessità di garantire i diritti fondamentali ed in particolare la presunzione di innocenza.

Sul punto, la risoluzione del Parlamento europeo del 6 ottobre 2021 mira a promuovere un utilizzo dell'intelligenza artificiale che sia sicuro, trasparente, non discriminatorio e rispettoso diritti fondamentali dei valori fondativi dell'ordinamento europeo, collocandosi in tal modo in linea di continuità con le prese di posizione assunte dalle altre istituzioni europee e internazionali.

Proprio per bilanciare le opportunità con i rischi, quanto al requisito della trasparenza dei sistemi utilizzati, già il paragrafo 17 della citata risoluzione del 6 ottobre 2021, in coerenza con il quarto principio della Carta etica, poneva l'accento sulla necessità di assicurare "metodologie di trattamento dei dati accessibili e comprensibili". Il medesimo paragrafo insiste, inoltre, su una serie di requisiti tecnici e metodologici considerati imprescindibili per un impiego responsabile dell'IA in ambito penale, invocando "la spiegazione, la trasparenza, la tracciabilità e la verifica degli algoritmi quali elementi necessari della vigilanza al fine di garantire che lo sviluppo, la diffusione e l'utilizzo di sistemi di IA per il settore giudiziario e delle attività di contrasto rispettino i diritti fondamentali e godano della fiducia dei cittadini, nonché al fine di garantire che i risultati generati dagli algoritmi possano essere resi intellegibili per gli utenti e coloro che sono soggetti a tali sistemi, e che vi sia trasparenza riguardo ai dati di base e alle modalità con cui il sistema è giunto a una certa conclusione". A tal fine, la Risoluzione in questione raccomanda che l'acquisto e l'impiego di tali strumenti siano consentiti esclusivamente laddove essi risultino suscettibili di revisione ed i relativi codici sorgente siano accessibili e comprensibili quantomeno per le autorità giudiziarie, le forze di polizia e gli organismi indipendenti di controllo.

In tale ambito si pongono le problematiche relative al riconoscimento facciale[23], sulle quali nel prosieguo ci soffermeremo, in relazione alle quali la risoluzione che qui si commenta pur riconoscendone le potenzialità applicative, manifesta una profonda preoccupazione, richiamando la necessità di circoscriverne l'utilizzo a specifiche finalità strettamente giustificate, in ragione dell'elevato impatto che tali strumenti possono determinare sui diritti fondamentali della persona, quali il diritto alla riservatezza, la presunzione di innocenza, la dignità umana e, in ultima analisi, l'affidabilità stessa dei risultati ottenuti.

Con riferimento al profilo delle discriminazioni, il paragrafo 9 della risoluzione in questione offre alcuni esempi di discriminazione algoritmica da evitare laddove impattante negativamente sulle minoranze razziali o etniche, sulla comunità LGBTQ, sui bambini, sugli anziani e sulle donne.

In conclusione, le raccomandazioni contenute nella risoluzione del parlamento europeo convergono su una linea forte di cautela circa la delega di processi decisionali a sistemi di IA intelligenze artificiale, specialmente in ambito penale con specifico riferimento all'attività di polizia giudiziaria.

§§§§§

[23] La tecnologia del riconoscimento facciale che, come vedremo nel capitolo successivo, costituisce uno valido strumento di contrasto al terrorismo consente l'estrazione e l'elaborazione di dati biometrici del volto delle persone riprese da telecamere installate in spazi pubblici, al fine di confrontarli con quelli archiviati in banche dati contenenti le generalità di individui sospettati o già condannati per reati di terrorismo.

3. I rischi connessi all'utilizzo dei sistemi di intelligenza artificiale

Come abbiamo visto nel paragrafo precedente, l'intelligenza artificiale è una tecnologia potente che, se regolata correttamente, può portare enormi benefici e contribuire in maniera decisiva al benessere economico e sociale dell'intera popolazione mondiale. Nell'ambito della materia penale sono indubbie fonti di opportunità tanto sul versante giudicante che su quello inquirente sul presupposto del controllo umano dell'agente artificiale e del pieno rispetto dei diritti fondamentali dell'individuo.

Allo stesso tempo, le moderne tecnologie possono essere utilizzate dalle organizzazioni criminali per commettere reati nei più svariati settori del diritto penale e per quel che a noi interessa nella materia del terrorismo aprendo scenari che al momento possono essere imprevedibili.

Le descritte situazioni impongono di analizzare le ragioni per cui i sistemi di IA da opportunità possono trasformarsi in un rischio e contestualmente gli accorgimenti che a livello normativo il legislatore, sia in ambito sovranazionale che interno, ha intrapreso per evitare che venga compromessa la sicurezza e l'integrità delle istituzioni democratiche.

Si tratta di tematiche ben chiare al legislatore europeo che nell'art.3 dell'AI ACT dapprima definisce il rischio "come la combinazione della probabilità del verificarsi di un danno e la gravità del danno stesso" e successivamente introduce il risk based approach, classificando cioè i sistemi di I.A. in relazione al rischio di impatto negativo sui diritti fondamentali, tra cui la dignità umana, la democrazia, l'uguaglianza, la non discriminazione, la protezione dei dati personali e la sicurezza.

§§§

3.1. Le pratiche di IA vietate e la gradualità del rischio nell'AI ACT.

In virtù del citato principio della gradualità del rischio, il regolamento AI ACT distingue tra di loro i seguenti sistemi:

- sistemi a rischio inaccettabile, che trovano la loro disciplina nell'art.5;
- sistemi a rischio elevato, per i quali è previsto un obbligo di certificazione e che trovano la loro disciplina nell'art. 6 e nell'All.III;
- sistemi a rischio limitato, per i quali è previsto un obbligo di trasparenza e che trovano la loro disciplina nell'art.50;
- sistemi a rischio basso, per i quali non sussiste alcun obbligo informativo e nessuna previsione normativa che li riguardi.

Trasversalmente alle categorie di rischio indicate, si inserisce quella dei sistemi a rischio sistemico per i quali è previsto un obbligo di informazione e che trovano la loro disciplina negli artt.51 e ss..

In considerazione del mutevole progresso delle tecnologie e dei rischi connessi, la riconducibilità dei sistemi di IA alle categorie di rischio indicate è suscettibile di subire variazioni rispetto alle originarie collocazioni come, ad esempio, avvenuto con alcune pratiche di predizione *person based*, sulle quali ci soffermeremo da qui a breve.

In data 4 febbraio 2025, la Commissione europea ha adottato, conformemente all'art. 96 dell'AI Act ed in vista dell'applicabilità dei divieti già vigenti a partire dal 2 febbraio 2025, le linee Guida relative alle pratiche di intelligenza artificiale vietate dall'art. 5. Queste indicazioni mirano a garantire un'applicazione coerente, efficace e uniforme del regolamento, in ausilio a fornitori ed utilizzatori di sistemi di intelligenza artificiale. A tal fine, per ciascuna delle categorie di rischio indicate, vengono fornite chiarificazioni sul significato dei concetti identificativi e delineati

esempi d'implementazioni vietate. Emerge in questo modo lo sforzo di fornire indicazioni concretamente attuabili perché calate sulla specificità della funzione e dell'oggetto del singolo divieto e costruite per differenziare le pratiche illecite da quelle lecite.

§

3.1.1. I sistemi a rischio inaccettabile. Le pratiche inaccettabili in ambito giudiziario: profilazione e riconoscimento facciale mediante *scraping.*

Con riferimento ai sistemi a rischio inaccettabile, l'art.5 dell'AI ACT[24], disposizione entrata in vigore il 2 febbraio 2025, include in tale ambito quei sistemi che violano i diritti fondamentali o sono utilizzati per pratiche di manipolazione, sfruttamento e controllo sociale.

In particolare secondo la descrizione dell'art. 5 in questione vi rientrano quelli che: utilizzano tecniche subliminali manipolatorie; sfruttano la vulnerabilità di soggetti; effettuano valutazioni sul rischio di commissione di reati; ampliano banche dati di riconoscimento facciale mediante tecniche di *scraping*; inferiscono emozioni personali sul luogo di lavoro e nei luoghi di istruzione; compiono categorizzazione biometrica con modalità discriminatoria o remota in spazi accessibili al pubblico tranne che per modalità di contrasto tassativamente previste.

[24] In data 4 febbraio 2025, la Commissione europea ha adottato, conformemente all'art. 96 dell'AI Act ed in vista dell'applicabilità dei divieti già vigenti a partire dal 2 febbraio 2025, le linee Guida relative alle pratiche di intelligenza artificiale vietate dall'art. 5. Queste indicazioni mirano a garantire un'applicazione coerente, efficace e uniforme del regolamento, in ausilio a fornitori ed utilizzatori di sistemi di intelligenza artificiale. A tal fine, per ciascuna delle quattro categorie di IA vietate, vengono fornite chiarificazioni sul significato dei concetti identificativi e delineati esempi d'implementazioni vietate. Emerge in questo modo lo sforzo di fornire indicazioni concretamente attuabili perchè calate sulla specificità della funzione e dell'oggetto del singolo divieto e costruite per differenziare le pratiche illecite da quelle lecite.

Nel capitolo secondo verranno affrontati i vantaggi che i sistemi di IA possono portare nella lotta contro il terrorismo. Tuttavia, il rischio che vengano compromessi diritti fondamentali dell'individuo ha indotto il legislatore europeo ad introdurre alcuni accorgimenti.

Riservando di sviluppare più ampiamente tale problematica, in questa sede occorre anticipare che con specifico riferimento ai sistemi di IA che effettuano valutazioni sul rischio di commissione di reati, come appunto può accadere nella materia del terrorismo, nella originaria previsione dell'AI ACT tali applicazioni predittive venivano fatte rientrare tra i sistemi di IA ad alto rischio, e come tali ammesse purché sottoposte alle misure di contenimento dei rischi prescritte dal Regolamento europeo. Tuttavia, a seguito delle modifiche apportate al documento da Parlamento e Consiglio, nella versione definitiva del Regolamento alcune pratiche di predizione *person-based* sono state qualificate come vietate, in quanto costitutive di un pericolo intollerabile per i diritti fondamentali. In particolare, l'art.5 par. I, lett. d) del Regolamento europeo proibisce "l'immissione sul mercato, la messa in servizio per tale finalità specifica o l'uso di un sistema di IA per effettuare valutazioni del rischio relative a persone fisiche al fine di valutare o prevedere il rischio che una persona fisica commetta un reato, unicamente sulla base della profilazione di una persona fisica o della valutazione dei tratti e delle caratteristiche della personalità".

Il divieto presenta alcune rilevanti eccezioni in quanto non si applica ai sistemi di IA utilizzati a sostegno della valutazione umana del coinvolgimento di una persona in un'attività criminosa, che si basa già su fatti oggettivi e verificabili direttamente connessi a un'attività criminosa. Se ne ricava che, in ambito europeo, non sono ammesse soluzioni predittive fondate su valutazioni esclusivamente *person based* essendo invece possibile un loro utilizzo a supporto di un fondamento probatorio

preesistente di natura oggettiva riguardante il coinvolgimento della persona nell'attività criminosa che si ritiene in esecuzione.

Con riferimento ad un'altra tipologia di divieto costituita dai sistemi di IA che ampliano banche dati di riconoscimento facciale[25] mediante tecniche di *scraping* le linee Guida relative alle pratiche di intelligenza artificiale vietate dall'art. 5 dell'AI ACT hanno stabilito che in questi casi non è richiesto che l'unico scopo del *database* sia quello di essere utilizzato per il riconoscimento facciale, bastando che possa essere impiegato a questo scopo. Allo stesso modo lo *scraping* indiscriminato di immagini facciali è ritenuto sussistente quando la raccolta di dati o contenuti avvenga senza un focus specifico su un singolo individuo o su un gruppo di individui, a prescindere dal rispetto dei protocolli di *opt-out* di internet, come robot.txt. Per la Commissione si è, però, comunque al cospetto di *scraping* vietato quando il risultato finale sia funzionalmente lo stesso di un'operazione di *scraping* indiscriminata fin dall'inizio. A fronte di ciò ed in generale per evitare impieghi che siano nei fatti elusivi dei divieti, la Commissione intende incoraggiare l'adozione di misure di trasparenza e *audit* dei sistemi IA per garantire in itinere il rispetto delle norme.

Dal punto di vista sanzionatorio l'articolo 99, comma 3, stabilisce che la non conformità al divieto delle pratiche di IA di cui all'articolo 5 in questione è soggetto a sanzione amministrativa pecuniaria fino ad € 35.000,00 oppure se l'autore del reato è un'impresa fino al 7% del fatturato mondiale totale annuo dell'esercizio precedente se superiore.

Nonostante l'indicato aspetto sanzionatorio, in dottrina sono state evidenziate delle lacune normative in quanto nel regolamento in

[25] La tecnologia del riconoscimento facciale che, come vedremo nel capitolo successivo, costituisce uno valido strumento di contrasto al terrorismo consente l'estrazione e l'elaborazione di dati biometrici del volto delle persone riprese da telecamere installate in spazi pubblici, al fine di confrontarli con quelli archiviati in banche dati contenenti le generalità di individui sospettati o già condannati per reati di terrorismo

questione non vengono chiaramente indicati i rimedi individuali specifici che possono essere esperiti dai singoli utenti in caso di inosservanza di tali disposizioni da parte dei produttori. Infatti, per le imprese viene stabilito, solamente, un obbligo di conformità entro sei mesi dall'entrata in vigore del sistema IA, ragion per cui in caso di violazione dell'art.5 dell'AI ACT per i singoli utenti l'unica norma applicabile sembrerebbe quella di cui all'art. 22 GDPR che stabilisce il diritto da parte dell'interessato di non essere sottoposto a decisioni automatizzate[26].

Quanto all'obbligo di conformità deve essere ribadita in questa sede l'ampia portata materiale e territoriale del regolamento. Infatti, l'AI ACT si applica tanto alla immissione sul mercato (distribuzione e uso nel corso di attività commerciali) quanto alla messa in servizio nel mercato UE (fornitura all'utilizzatore per il primo uso) da parte di utilizzatori anche se non stabiliti o ubicati all'interno dell'UE di sistemi di IA o modelli di IA per finalità generali nell'Unione. In particolare il regolamento si estende agli utilizzatori stabiliti o ubicati all'interno dell'UE, agli importatori e distributori di sistemi IA, ai fabbricanti di prodotti che immettono sul mercato o mettono in servizio un sistema di IA insieme al loro prodotto e con il loro nome o marchio, ai rappresentanti dei fornitori non stabiliti in UE, alle persone interessate che si trovano in UE, fornitori e utilizzatori nell'Unione stabiliti o ubicati in un paese terzo, qualora l'output prodotto sia utilizzato nell'UE. Si prevede sostanzialmente un ambito di applicazione esteso persino agli output prodotti in Europa, con evidente difficoltà di inquadramento del fatto giuridico nello spazio e nel tempo in ossequio al principio di certezza del diritto, in quanto un output di un sistema di IA può addirittura riferirsi a un'utilizzazione diluita nel tempo dei

[26] In ordine al diritto alla decisione umana, di recente la Corte di Giustizia si è espressa delineando le caratteristiche del significato di decisioni automatizzate produttive di effetti giuridici e includendo tra esse la profilazione (C-634/21, Schufa Holding (Scoring), 7 dicembre 2023).

risultati indipendentemente dalla utilizzazione del sistema stesso in Europa. Tra le deroghe si annoverano la sicurezza nazionale, la difesa e l'uso militare - con evidenti discrasie in ordine al dual use, mentre sono esenti dall'applicazione dell'AI Act le autorità pubbliche di paesi terzi e organizzazioni internazionali

§

3.1.2. I sistemi a rischio elevato. In particolare, i sistemi di IA nell'amministrazione della giustizia.

Con riferimento ai sistemi a rischio elevato, l'articolo 6, paragrafo 2, del regolamento AI ACT prevede che "un sistema di intelligenza artificiale è considerato ad alto rischio quando è destinato ad essere utilizzato come componente di sicurezza di un prodotto, o il cui utilizzo è soggetto ad obblighi di valutazione della conformità prima della messa in commercio o dell'entrata in servizio del prodotto in questione, oppure quando è elencato nell'allegato III, a meno che non sia dimostrato, in base ad una valutazione preliminare, che il sistema non comporta un rischio significativo per la salute, la sicurezza o i diritti fondamentali delle persone fisiche, tenuto conto della sua finalità, del contesto d'uso, della probabilità e della gravità dell'impatto".

Sul punto, il citato art. 6, co. 2, in combinato disposto con l'allegato III, n. 8 del regolamento in questione, le cui disposizioni entreranno in vigore il 2 febbraio del 2026, elenca i sistemi ad alto rischio considerati legittimi, nel cui ambito rientrano quelli relativi: allo scoring lavorativo; all'accesso alle prestazioni pubbliche essenziali in ordine; alla valutazione dell'affidabilità creditizia; all'amministrazione della giustizia; ai processi democratici.

Tali sistemi devono soddisfare requisiti rigorosi in termini di trasparenza, explainability, sicurezza e supervisione specificatamente delineati dai seguenti articoli:

- l’art. 10 del Regolamento prevede che l’I.A. sia addestrata con dati che soddisfano precisi “criteri di qualità”;
- l’art. 13 del Regolamento impone che la progettazione e lo sviluppo dei sistemi di I.A. ad alto rischio avvenga nel rispetto di canoni di “trasparenza”, dovendosi garantire che “il loro funzionamento sia sufficientemente trasparente da consentire agli utenti di interpretare l’output del sistema e utilizzarlo adeguatamente”;
- l’art. 14 del Regolamento prescrive la regola della “sorveglianza umana”, intesa come comprensiva di “strumenti di interfaccia uomo-macchina adeguati, in modo tale da poter essere efficacemente supervisionati da persone fisiche durante il periodo in cui il sistema di IA è in uso”;
- l’art. 15 del Regolamento, stabilisce che detti sistemi devono garantire standard di accuratezza, robustezza e cybersicurezza;
- l’art. 70 al fine di vigilare sulla corretta armonizzazione delle procedure nazionali con il Regolamento, stabilisce che ciascuno Stato membro istituisca un’Autorità di controllo incaricata di supervisionare l’effettiva applicazione delle disposizioni regolamentari.

Con specifico riferimento all’amministrazione della giustizia, giova premettere che in tale ambito rientrano temi molti diversi tra loro tra cui: quelli relativi alla raccolta e analisi di big data; quelli relativi alla automatizzazione di compiti ripetitivi attraverso procedure con o senza controllo dell’operatore umano; quelli relativi al rilevamento di malware o di disfunzioni nei sistemi; quelli relativi alla ottimizzazione delle reti organizzative e della condivisione di pratiche; quelli relativi alla generazione di testi; quelli relativi alla protezione, o indebita appropriazione, dei dati.

Un discorso a parte meritano le decisioni giudiziarie.

Sul punto sono stati già accennati i rischi connessi all'utilizzo dei sistemi di IA le cui potenzialità possono giungere fino allo scrivere autonomamente le sentenze sostituendosi al giudice. A ciò si aggiunga che tale tecnologia risulta priva di una coscienza critica che è l'essenza del pensiero umano e quindi dell'argomentazione giuridica posta alla base della decisione che per sua funzione riguarda beni giuridici costituzionalmente rilevanti. Ognuno di questi aspetti comporta problemi diversi: tecnici (di usabilità, di accettabilità, di generalizzabilità a contesti diversi), etici e normativi.

In proposito, la Carta etica stabilisce che le decisioni giudiziarie debbano chiaramente indicare la provenienza dei dati utilizzando fonti certificate e adottando modelli sviluppati in un ambiente tecnologico. Allo stesso tempo le decisioni giudiziarie devono garantire l'integrità del processo di elaborazione rendendo tracciabili *ex post* i passaggi logici e conservando in luoghi sicuri i modelli e gli algoritmi impiegati.

Successivamente alla Carta etica, con l'AI Act, il legislatore europeo si è preoccupato di dettare alcune regole in materia di amministrazione della giustizia.

Già nelle considerazioni preliminari, al punto 40, si trova scritto che «alcuni sistemi di IA destinati all'amministrazione della giustizia e ai processi democratici dovrebbero essere classificati come sistemi ad alto rischio, in considerazione del loro impatto potenzialmente significativo sulla democrazia, sullo Stato di diritto, sulle libertà individuali e sul diritto a un ricorso effettivo e a un giudice imparziale». In più, sottolinea il legislatore sovranazionale, è «in particolare opportuno, al fine di far fronte ai rischi di potenziali distorsioni, errori e opacità, classificare ad alto rischio i sistemi di IA destinati ad assistere le autorità giudiziarie nelle attività di ricerca e interpretazione dei fatti e del diritto e nell'applicazione della legge a una serie concreta di fatti».

In ambito interno, il citato art.15 del ddl 1146-B chiaramente stabilisce che nei casi di impiego di un sistema di intelligenza artificiale nelle attività giudiziaria è sempre riservata al magistrato ogni decisione sull'interpretazione e sull'applicazione della legge, sulla valutazione dei fatti e delle prove e sull'adozione dei provvedimenti ragion per cui in nessun caso potrebbero ritenersi ammessi quei sistemi che in ipotesi fossero in grado di scrivere autonomamente le sentenze.

Con riferimento alle imprese produttrici di tali sistemi considerati ad alto rischio, il regolamento AI ACT introduce, sulla falsariga del modello di responsabilità da prodotto, un mero obbligo di certificazione, che può essere aggiornato nel tempo dalla Commissione.

Dal punto di vista sanzionatorio il mancato obbligo di certificazione determina l'applicazione di sanzioni amministrative pecuniarie fino a 15 000 000 EUR o, se l'autore del reato è un'impresa, fino al 3 % del fatturato mondiale totale annuo dell'esercizio precedente così come stabilito dall'articolo 99, comma quarto, del regolamento AI ACT.

Si tratta di una previsione che può non rivelarsi appagante sotto il profilo della attuazione di rimedi giurisdizionali specifici, salvo la possibilità per il consumatore che rimane vittima di tali meccanismi di presentare un reclamo a un'autorità di vigilanza del mercato o il diritto di ottenere spiegazioni chiare e significative sul prodotto in questione.

§

3.1.3. I sistemi a rischio limitato: il caso di Chat-GPT di OpenAI e la collocazione sistematica dei deepfake.

L’art.50 dell’AI ACT definisce sistemi a rischio limitato tutti quei sistemi di IA destinati a interagire direttamente con le persone fisiche.

Tali sistemi devono essere progettati e sviluppati in modo tale che le persone fisiche interessate siano informate del fatto di stare interagendo con un sistema di IA. L’indicato obbligo di trasparenza, che impone al produttore di aderire a codici di condotta che hanno come finalità quello di informare gli utenti che stanno interagendo con un sistema di IA, costituisce la regola generale in materia. Tuttavia, la normativa in questione prevede una deroga all’obbligo di trasparenza nei seguenti casi: quando l’interazione con un sistema di IA risulti evidente dal punto di vista di una persona fisica ragionevolmente informata, attenta e avveduta, tenendo conto delle circostanze e del contesto di utilizzo; quando i sistemi di IA sono autorizzati dalla legge per accertare, prevenire, indagare o perseguire reati, fatte salve le tutele adeguate per i diritti e le libertà dei terzi, a meno che tali sistemi non siano a disposizione del pubblico per segnalare un reato.

Dal punto di vista sanzionatorio il mancato obbligo di trasparenza determina l’applicazione di sanzioni amministrative pecuniarie fino a 15 000 000 EUR o, se l'autore del reato è un'impresa, fino al 3 % del fatturato mondiale totale annuo dell'esercizio precedente così come stabilito dall’articolo 99, comma quarto, del regolamento AI ACT. Nessuna indicazione viene invece stabilita sui rimedi giurisdizionali specifici a tutela del consumatore, salvo la possibilità per lo stesso di presentare un reclamo ad un'autorità di vigilanza del mercato nonché il diritto di ottenere spiegazioni chiare e significative sul prodotto in questione.

Nella categoria dei sistemi a rischio limitato rientrano senza alcun ragionevole dubbio i Chatbot ovverosia quei programmi informatici progettati per simulare una conversazione umana, interagendo con gli utenti tramite testo o voce come ad esempio Siri, Google Assistant, Alexa.

Aspetti problematici suscita invece la collocazione nel settore in questione dei sistemi che utilizzano l'intelligenza artificiale generativa, che, come detto in precedenza[27], si configura come espressione dei sistemi c.d. “sapienti”, propri della cosiddetta *Strong AI,* i quali non si limitano a emulare le capacità cognitive umane, ma si contraddistinguono per la loro attitudine a svilupparle autonomamente.

Chat-GPT di OpenAI, unitamente ad altri modelli quali Deepseek lanciata nel 2024 da Beijing, PaLM (Google), LLaMA (Meta) e Claude (Anthropic), è un esempio dell'evoluzione di tali modelli essendo capace di creare autonomamente gli *output,* tramite la combinazione originale di una sconfinata mole di dati su cui si è addestrata e su cui continua ad addestrarsi man mano che interagisce con l'utente esterno. Dunque, siamo in presenza di una forma di IA generativa massimamente performante, ma anche altamente opaca e imprevedibile ragion per cui in dottrina si è autorevolmente parlato di *black box* e di "oracolo" algoritmico.

Un esempio può ulteriormente chiarire i termini della questione. Nel caso di *ChatGPT,* l'insieme delle fonti da cui la *chatbot* apprende e rielabora le informazioni è talmente esteso e incontrollabile che si sono verificati casi in cui l'elaboratore è stato accusato di aver inventato notizie diffamatorie[28]. In tali circostanze, la risposta di *OpenAI* è stata quella di oscurare completamente la ricerca delle persone diffamate all'interno del proprio sistema non essendo stato possibile risalire a ritroso alla fonte da cui l'IA aveva appreso tali informazioni errate, né intervenire selettivamente sull’output.

L’esempio in questione testimonia plasticamente perché sta diventando sempre più discussa la possibile trasmigrazione di tali sistemi di IA nella categoria dei sistemi ad alto rischio, anziché in quelle a rischio minimo, in quanto *l'output* dell'IA può diventare

[27] Sulla IA generativa cfr. *infra* cap. primo, par. 1.2.1 pag.22

[28] *Se l’intelligenza artificiale finisce in tribunale ChatGpt diffama un sindaco australiano che ora vuole denunciarla*, in la *Repubblica*, 9 aprile 2023

impenetrabile, imprevedibile e potenzialmente ingestibile dall'uomo.

Tale problematica si è vieppiù accentuata con la diffusione dei deepfake che grazie alla tecnologia della Generative Adversarial Networks (GAN) è in grado di sovrapporre volti, modificare espressioni e persino imitare voci, in modo quasi indistinguibile dalla realtà. Il primo approfondito lavoro sui rischi connessi a tale tecnologia ed in particolare su quelli legati a possibili forme di manipolazione politica e sociale connessi ai deep fakes è di Chesney e di Citron[29].

Invero, nella loro formulazione originaria, i deepfake nascono per finalità del tutto lecite tra cui, per limitarci ad alcuni esempi, quelli utilizzati nel mondo del cinema e dell'animazione, dell'arte e dell'intrattenimento in generale, nonché in quello degli avatar virtuali che agiscono parlando numerose lingue.[30] Tali utilizzi non hanno posto alcuna difficoltà di collocamento all'interno dei sistemi a rischio limitato. Sul punto l'art. 50, comma 4, che qui si commenta, stabilisce che i deployer di un sistema di IA che genera o manipola immagini o contenuti audio o video che costituiscono un «deep fake» rendono noto che il contenuto è stato generato o manipolato artificialmente. Anche in questo caso, il prescritto obbligo di trasparenza prescritto come regola di carattere generale soffre delle eccezioni: infatti non si applica se l'uso è autorizzato dalla legge per accertare, prevenire, indagare o perseguire reati; allo stesso quando il contenuto faccia parte di

[29] R. CHESNEY – D.K. CITRON, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 California Law Review 1753 (October 15, 2019).

[30] Il 30 novembre del 2022 la startup OpenAI ha aperto al pubblico ChatGpt, gratuitamente, un chatbot basato su un *Large Language Model*, un'intelligenza artificiale generativa in grado di «comprendere» e generare linguaggio naturale. Si tratta di una data che viene considerata l'inizio dell'intelligenza artificiale di massa. Dopo ChatGpt è arrivato Bard – ora Gemini – di Google, Copilot di Microsoft, alcuni modelli dedicati alla creazione di immagini come Midjourney e Dall-e, e via via gli altri. Un'idea, condivisa da molti addetti ai lavori, è che a livello di prodotti per i singoli utenti (dunque non parliamo di software destinati alle aziende), non c'è né ci sarà in un futuro prossimo un modello di intelligenza artificiale che dominerà tutti gli altri.

un'analoga opera o di un programma manifestamente artistici, creativi, satirici o fittizi, il prescritto obbligo di trasparenza di cui si limita all'obbligo di rivelare l'esistenza di tali contenuti generati o manipolati in modo adeguato, senza ostacolare l'esposizione o il godimento dell'opera.

Ciò premesso, i deep fake hanno raggiunto l'attenzione pubblica e dei media nel 2017, quando su alcune community del sito 'Reddit' sono stati pubblicati falsi filmati pornografici di attrici e cantanti, delle quali risultava facile reperire immagini e dati essenziali per l'elaborazione del falso.

Tale uso illecito e ingannatorio dei deep fake si è poi esteso dalla pornografia ad una serie indeterminata di fattispecie penali, tra cui, senza la pretesa di essere esaustivi, si segnalano quelle relative:

- alle frodi, comprensive delle campagne di disinformazione globale ('deep fake news') e di manipolazione dell'opinione pubblica, anche in ambito elettorale e in ambito sanitario e scientifico (campagne no-vax, false informazioni su covid-19);
- alla diffamazione e harassment online[31];
- ai fenomeni di astroturfing[32];
- all'estorsione (es. minaccia di rilasciare un video falso che danneggerebbe la reputazione o la credibilità di una persona fisica o giuridica);
- agli attacchi informatici, al fine di realizzare il furto di identità[33] o di ingannare i sistemi di autenticazione, ad esempio, ordinando trasferimenti di denaro per aiutare le organizzazioni terroristiche;

[31] In tale ambito ed in tale contesto si inserisce la tematica delle molestie realizzate con messaggi e video ripetuti nel tempo

[32] Il termine astroturfing si riferisce a tutte le tecniche di marketing che vengono inviate agli utenti

[33] Secondo il rapporto Veridas, in https://veridas.com il furto di identità tramite AI è in forte crescita, con un aumento del 135% delle violazioni di account nel 2024.

- ai falsi personali (sostituzione di persone e fenomeno di 'morphing attack') e documentali (fabbricazione o alterazione di prove digitali in ambito processuale);
- alle distorsioni e manipolazione dei mercati (es. diffusione di video falso in cui il CEO di una società quotata commette un reato o si lascia andare a commenti razzisti/misogini);
- alle incitazioni agli atti di violenza contro minoranze;
- all'incentivo all'agitazione sociale e alla polarizzazione politica (es. teorie e cospirazioni) anche per finalità discriminatorie ai danni di gruppi sociali specifici o di natura sessualmente orientata.

Inoltre, nella materia che a noi interessa, i deepfake costituiscono una delle macroaree di interazione tra l'IA ed il terrorismo potendo essere utilizzati dalle organizzazioni terroristiche nell'ambito dei fenomeni di radicalizzazione per alimentare i discorsi di odio, tema sul quale ci soffermeremo nel capitolo secondo.

In conclusione, il descritto sconfinamento dei deepfake testimonia come i confini tra le varie categorie di rischio siano piuttosto mobili e tra di loro interscambiabili nell'applicazione pratica.

Sul tema in questione, al punto 56 delle considerazioni introduttive delle citate Linee guida della commissione europea del 4 febbraio 2025 viene chiaramente stabilito che è possibile che uno stesso comportamento vietato possa costituire una violazione di due o più disposizioni del regolamento sull'IA.

§

3.1.4. I sistemi a rischio minimo.

Con riferimento ai sistemi a rischio minimo, deve rilevarsi che in tale ambito vi rientrano i sistemi di IA utilizzati in giochi o per

creare playlist musicali personalizzate; i software che automatizzano compiti ripetitivi di routine in contesti aziendali, come la gestione di fatture o la programmazione di appuntamenti; i sistemi che analizzano grandi set di dati per identificare tendenze di mercato o preferenze dei consumatori, senza prendere decisioni automatiche che hanno un impatto significativo sugli individui. Si tratta di sistemi in relazione ai quali non sussiste normativamente alcun obbligo e rispetto ai quali emergono in maniera assoluta i vantaggi e le opportunità che tali sistemi possono recare nella vita di tutti i giorni.

§

3.1.5. I modelli a rischio sistemico.

Trasversalmente a quelli indicati, si inserisce la categoria dei sistemi a rischio sistemico disciplinati dall'art.51 del regolamento AI ACT secondo cui un rischio può essere definito tale quando non riguarda solo singoli utenti o casi isolati, ma può avere conseguenze diffuse e gravi sull'intero sistema economico, politico o sociale.
Ciò accade quando i sistemi in questione presentano capacità di impatto elevato che devono essere valutate dalle autorità competenti sulla base di strumenti tecnici e metodologie adeguati, compresi indicatori e parametri di riferimento.
Sul punto la normativa in questione opera una distinzione tra valutazioni presuntive e valutazioni empiriche.
Nel caso delle valutazioni presuntive, il modello di IA per finalità generali è considerato a rischio sistemico sulla base di una presunzione legale che si verifica allorquando la quantità cumulativa di calcolo utilizzata per il suo addestramento misurata in operazioni in virgola mobile è superiore a 10 25.
Nel caso delle valutazioni empiriche, il modello di IA per finalità generali viene riconosciuto a rischio sistemico sulla base di una

decisione della Commissione, ex officio o a seguito di una segnalazione qualificata del gruppo di esperti scientifici, allorquando il modello in questione presenta capacità o un impatto equivalenti a 10 25 tenendo conto dei criteri di cui all'allegato XIII.

Alla luce delle considerazioni svolte nelle pagine precedenti, si discute se in tale ambito normativo possano essere inclusi i descritti modelli Large *Language Models (LLM)* atteso che la loro capacità di elaborare e analizzare i dati in modo autonomo ed incontrollato può determinare una disinformazione su larga scala come, ad esempio, la generazione di testi apparentemente credibili ma in realtà falsi con la possibilità di essere usati a fini di manipolazione politica, propaganda o truffe.

§§§§§

4.Conclusioni.

Come detto in apertura del presente capitolo, l'era della transizione digitale apre le porte a quella che dopo, la macchina a vapore, l'elettricità ed il computer è stata definita la quarta rivoluzione industriale.

Quando si parla di rivoluzione digitale il riferimento è al passaggio da un mondo basato su processi manuali e analogici ad un mondo in cui la tecnologia digitale si spinge

fino a creare una trasposizione del mondo reale in quello virtuale secondo la logica del metaverso.

L'esistenza di una realtà virtuale (VR) inizialmente ristretta al settore dei video giochi sempre più sta coinvolgendo il nostro vivere sociale. Dalla medicina, alle attività commerciali tutto si svolge in una realtà sovrapponibile che trasposta nel diritto penale

implica trasformazioni ed interrogativi tanto sul versante del diritto penale sostanziale che in quello processuale.

In questo ambito si colloca l'intelligenza artificiale che rappresenta una grande opportunità per l'umanità soprattutto in quei settori definiti strategici per il nostro vivere sociale perché contribuisce al benessere economico e più in generale al miglioramento delle condizioni di vita.

Lo sviluppo di un'attenzione "diffusa" per l'IA, conseguenza anche del moltiplicarsi delle sue applicazioni nella quotidianità degli utenti (si pensi ad esempio agli assistenti virtuali nei cellulari o in device casalinghi, ed all'"esplosione" dei chatbot e dell'intelligenza artificiale generativa), oltre a portarla al di fuori della discussione tra esperti, ha fatto nascere un dibattito sulla necessità di trovare un equilibrio fra due opposte esigenze:

- non rallentare, o addirittura bloccare, il progresso del settore, e quindi le conseguenze positive dello stesso (comprensivo dell'enorme business da essa resa possibile direttamente, per il valore in sé dell'IA e, indirettamente, per la ricchezza prodotta dalle sue applicazioni);
- impedire che tale progresso avvenga in danno dei suoi utenti attraverso la protezione dei dati personali degli stessi.

La scelta condivisibile del legislatore europeo è stata quella di mettere, anche in ambito giudiziario, al centro del sistema il controllo umano per reagire ad una tecnologia dirompente e garantire che non venga compromesso l'equilibrio tra innovazione digitale e diritti fondamentali.

L'AI ACT, che indubbiamente costituisce il primo e più ampio intervento regolatorio della materia, muove da un approccio antropocentrico ed introduce il criterio della gradualità del rischio chiaramente distinguendo quattro livelli: un primo livello definito inaccettabile comprensivo, ad esempio, di quei sistemi che consentono la sorveglianza di massa o lo scoring lavorativo e le tecniche di manipolazione sociale; un secondo livello in cui il

rischio è elevato ed in relazione al quale vengono previsti obblighi di conformità da parte dei produttori; un terzo livello in cui il rischio è limitato ed in cui vengono previsti obblighi di informazione a carico di chi li immette nel mercato; ed infine un quarto livello in cui il rischio è basso ed in cui non sono previsti particolari obblighi normativi.

Gli elevati livelli di rischio e le indefinite potenzialità dei sistemi di IA, impongono di trasferire il tema in questione sulle modalità con cui la stessa IA sta trasformando la minaccia terroristica globale non solo nell'area della manipolazione politica (c.d. political threats) ma anche nelle altre due distinte macroaree degli attacchi informatici (cyber-threats) e degli attacchi fisici (psycal-threats)

CAPITOLO SECONDO: INTELLIGENZA ARTIFICIALE COME STRUMENTO PER LA COMMISSIONE DI REATI TERRORISTICI E COME NUOVO ULTERIORE STRUMENTO DI CONTRASTO ALL'INTERNO DEL SISTEMA NORMATIVO

1. L'evoluzione del terrorismo internazionale e le risposte normative

I sistemi di IA possono interferire con i reati di terrorismo attraverso la seguente duplice modalità:

- IA come strumento per la commissione di reati;
- IA come (nuovo ed ulteriore) strumento di contrasto all'interno del sistema normativo dato.

Come detto, si tratta di due questioni che hanno rilevanza di carattere generale all'interno della macroarea dei rapporti tra intelligenza artificiale e sistema penale, ma che, ovviamente, assumono un rilievo del tutto peculiare rispetto ai fenomeni del terrorismo.

Al riguardo, deve essere sottolineato come, a partire dagli attacchi di matrice jihadista compiuti l'11 settembre 2001 a New York[34], il tema della prevenzione e della repressione dei crimini

[34] M. ROMANELLI, *Criminalità organizzata e terrorismo: la circolazione dei modelli criminali e degli strumenti di contrasto, op.cit.*. Nello scritto in questione l'Autore ricostruisce i diversi periodi all'interno dell'operatività dello Stato Islamico, ed i relativi fenomeni criminali distinguendo tre diverse fasi: la prima riconducibile alla data simbolo dell'11 settembre 2001, giorno dell'inimmaginabile attacco dell'organizzazione terroristica Al Qaida alle Torri Gemelli; la seconda, a partire dal 2013, concomitante l'affermazione dell'Islamic State fino alla sconfitta sul campo, ed all'attuale, quasi completa, scomparsa del c.d. terrorismo territoriale; la terza fase dei giorni nostro caratterizzata dall'avvento delle tecnologie informatiche e della IA.

commessi dalle organizzazioni terroristiche sia diventato l'obiettivo primario della comunità mondiale.

Anche in ambito europeo, la centralità della lotta al terrorismo costituisce un principio oramai consolidato per garantire la sicurezza dei cittadini europei e realizzare quello spazio di libertà, sicurezza e giustizia (disegnato dai Trattati), attraverso la cooperazione e la collaborazione tra le agenzie di contrasto e le strutture giudiziarie dei paesi dell'UE.

Il terrorismo rappresenta senza dubbio una delle più grandi sfide al processo di integrazione e richiede a tutti gli "attori europei" - le istituzioni dell'Unione, le istituzioni degli Stati Membri, magistratura e forze di polizia in prima linea – un grande impegno per evitare che una tale minaccia globale possa compromettere la sicurezza degli Stati e la vita degli individui.

Si tratta di una sfida estremamente impegnativa, ancor di più oggi perché ci troviamo di fronte ad una situazione in evoluzione, fluida, non facilmente decifrabile, sia per le caratteristiche evolutive del fenomeno criminale[35], che per le nuove forme di terrorismo internazionale direttamente riconducibili all'evoluzione tecnologica e all'avvento dell'IA.

In tale ambito si inserisce il discorso sulle piattaforme online che, come detto in apertura del capitolo primo, possono essere sfruttate dai terroristi per diffondere contenuti a una molteplicità di individui, velocemente, in modo virale ed a bassi costi creando

[35] Nell'ambito delle novità che caratterizzano il terrorismo internazionale non possono non essere menzionate lo scoppio della pandemia da Covid-19, il conflitto Russia-Ucraina alle porte dell'Europa, e la questione medio-orientale caratterizzata dallo scontro Israele-Palestina che rendono lo scacchiere mondiale profondamente incerto e che favoriscono il pullulare delle organizzazioni terroristiche. Se è già molto complesso conoscere lo scenario e le linee evolutive del terrorismo c.d. islamico, gli esperti della materia hanno correttamente sottolineato come la questione si complica ulteriormente se si porta l'attenzione sull'evoluzione del fenomeno del terrorismo su base suprematista, razzista, nazista, sulle sue reti, sui meccanismi di radicalizzazione e di passaggio all'azione, sui vari fronti e territori, soprattutto extraeuropei, ma con significative punte anche in Europa ed in Italia, come è purtroppo tristemente noto a partire dalla terribile strage di Utoya del 22 luglio 2011, fino alle gesta di Luca Traini del 3 gennaio 2018.

nuove frontiere per il terrorismo internazionale; il tutto per una molteplicità di esigenze tra cui: l'addestramento (auto-addestramento), il finanziamento, la ricerca, la comunicazione di informazioni, la pianificazione operativa, il reclutamento e, non da ultimo, la propaganda e la vera e propria guerra psicologica.

Nella realtà in cui viviamo, l'uso propagandistico delle piattaforme online è ancor di più facilitato dall'utilizzo delle tecnologie basate sulla IA, tra cui rientrano quelle idonee a sovrapporre volti, modificare espressioni e persino imitare voci, che perfettamente appaiono conciliarsi con le peculiarità del terrorismo internazionale di matrice jihadista, per come si è manifestato ed evoluto negli ultimi anni e che sembra aver temporaneamente accantonato la scala e la tattica paramilitare degli attacchi di Parigi e Bruxelles a favore di una minaccia più parcellizzata e pervasiva servendosi di armi rudimentali, improvvisate ed imprevedibili (per esempio autovetture, veicoli commerciali). Come evidenziato in dottrina[36], nella sostanza si tratta di una forma di terrorismo caratterizzata da strutture fluide, cellulari e decentralizzate ma capaci di operare "a rete" anche a grande distanza grazie all'abile uso delle citate tecnologie informatiche, e per questo capace di mobilitare le "vocazioni" di singoli individui che perseguono "in proprio" il programma criminoso del gruppo. Ed è quest' ultimo il fenomeno dei cosiddetti "lupi solitari" (lone wolves) – pensiamo a quanto accaduto a Nizza, Londra, Stoccolma, Manchester, Barcellona, Berlino – che sembra segnare un temporaneo mutamento della strategia terroristica di matrice jihadista e che presuppone appunto una notevole capacità di propaganda.

§§§

[36] M. ROMANELLI, *Criminalità organizzata e terrorismo: la circolazione dei modelli criminali e degli strumenti di contrasto, op.cit.*.

1.1. L'adeguamento delle strategie di contrasto sul piano del diritto sostanziale e della cooperazione internazionale alle nuove forme di terrorismo internazionale.

Tale evoluzione del terrorismo impone un adeguamento delle strategie di contrasto, tanto sul piano del diritto sostanziale che sul piano della cooperazione internazionale attraverso un coordinamento e norme comuni che contribuiscano in prospettiva alla costruzione di un diritto penale internazionale nel rispetto, tuttavia, dei poteri di intervento delle giurisdizioni nazionali.

In ambito internazionale e interno si sono succeduti plurimi interventi in *subiecta materia*.

In primo luogo, devono essere menzionate le plurime risoluzioni del Consiglio di sicurezza delle Nazioni Unite[37] immediatamente successive agli attacchi delle Torri Gemelle ed incentrate sulla prevenzione e la repressione degli attacchi terroristi. Il tratto caratteristico di tali risoluzioni, che hanno poi orientato le legislazioni interne dei vari Paese membri, è stato quello di colpire il fenomeno del terrorismo sul versante delle attività di finanziamento e reclutamento.

In una soluzione di ideale continuità con le citate risoluzioni, in ambito europeo il problema di adottare efficaci misure contro il terrorismo è stato dapprima affrontato con la decisione quadro 2002/475/GAI del Consiglio, del 13 giugno 2002, sulla lotta contro il terrorismo e successivamente con la direttiva 541 del

[37] Risoluzione delle nazioni unite: 1373 del 2001; 12 novembre 2001 n.1377; 14 settembre 2005 n.1624; 24 settembre 2014 n.2178. nelle quali è dichiarato che "atti, metodi e pratiche di terrorismo sono contrari ai fini e ai principi delle Nazioni unite" e che «chiunque inciti, pianifichi, finanzi deliberatamente atti di terrorismo compie attività contrarie ai fini e ai principi delle Nazioni unite». Tali risoluzioni non indicano compiutamente una nozione di terrorismo generalmente accettata nella comunità internazionale. In tale ambito sono state esaminate le risoluzioni ONU finalizzate a mantenere la pace e pongono all'interprete la necessità di stabilire cosa realmente debba intendersi per atto terroristico in mancanza di una indicazione specifica e tassativa. È chiaro che l'ampliamento della nozione di atto terroristico riverbera i suoi effetti anche ad altri fini come, ad esempio, la negazione dello status di rifugiato a chi dovesse essere accusato di appartenere ad organizzazioni terroristiche.

marzo del 2017 che rappresenta il nuovo passo compiuto dal legislatore europeo nell'ottica del contrasto al terrorismo internazionale, con il duplice obiettivo di colmare le lacune di tutela esistenti e migliorare il quadro giuridico euro unitario[38].

Infine, nel nostro ordinamento si sono succedute tre distinte fasi che hanno modificato nel corso del tempo l'attuale disciplina dei reati in materia di terrorismo che trova specifica collocazione negli artt. 270 bis e seguenti del codice penale: la prima fase conseguenziale agli attentati delle Torri Gemelle è stata realizzata con il fondamentale d.l. n. 374 del 18 ottobre 2001; la seconda fase immediatamente successiva alla strage di Londra è stata realizzata con la riforma del 7 luglio 2005 (d.l. 144 del 27 luglio 2005); la terza fase successiva alla proclamazione di Islamic State ed alla ripresa degli attentati in Europa, soprattutto quello di Parigi del Bataclan, si è compiuta con la riforma del 18 febbraio 2015 (D.l. n. 7 del 18 febbraio 2015.

Recentemente il d.l. 11 aprile 2025, n. 48, convertito dalla legge 9 giugno 2025, n. 80, meglio noto come "d.l.sicurezza", al fine di contrastare, con sempre maggior forza, il fenomeno del terrorismo anche internazionale, in attuazione delle direttive comunitarie succedutesi nel tempo e tenuto conto che le cronache più recenti hanno dimostrato come, in tale ambito, hanno acquisito sempre maggior rilievo le condotte autonome e individuali, che sfuggono alle tradizionali ipotesi associative, ha introdotto nel codice penale l'inedito art. 270-quinquies, rubricato

[38] Cfr. direttiva UE 2017/541 del 15 marzo 2017 «*sulla lotta contro il terrorismo e che sostituisce la decisione quadro 2002/475/GAI del Consiglio*» e, in particolare, l'art. 8, in forza del quale «*Gli Stati membri adottano le misure necessarie affinché sia punibile come reato, se compiuto intenzionalmente, l'atto di ricevere informazioni per la fabbricazione o l'uso di esplosivi, armi da fuoco o altre armi o sostanze nocive o pericolose ovvero altre tecniche o metodi specifici al fine di commettere o di contribuire alla commissione di uno dei reati di cui all'articolo 3, paragrafo l, lettere da a) ad i)*». In dottrina GENNUSO, *"Tutto in una definizione? La nuova direttiva antiterrorismo dell'Unione europea e i confini del terrorismo"* in *Quaderni costituzionali* / n.3 2017 pag.651

«detenzione di materiale con finalità di terrorismo» che punisce con la reclusione da due a sei anni: «*chiunque, fuori dei casi di cui agli articoli 270-bis e 270-quinquies, consapevolmente si procura o detiene materiale contenente istruzioni sulla preparazione o sull'uso di congegni bellici micidiali di cui all'articolo 1, primo comma, della legge 18 aprile 1975, n. 110, di armi da fuoco o di altre armi o di sostanze chimiche o batteriologiche nocive o pericolose, nonché su ogni altra tecnica o metodo per il compimento di atti di violenza ovvero di sabotaggio di servizi pubblici essenziali, con finalità di terrorismo, anche se rivolti contro uno Stato estero, un'istituzione o un organismo internazionale*».

Con l'introduzione della fattispecie in commento, quindi, il legislatore ha voluto costruire una norma di chiusura che, grazie alla clausola di salvezza prevista nella prima parte della norma, garantisca la rilevanza penale anche a condotte prodromiche rispetto alla realizzazione di atti concreti di terrorismo e conseguentemente colmare un vuoto normativo sulla detenzione di documentazione propedeutica al compimento di attentati e sabotaggi con finalità di terrorismo".

Dal punto di vista dell'elemento materiale, la condotta tipica incrimina la persona che «*si procura o detiene materiale*»[39] - di qualunque tipo, quindi anche informatico - contenente istruzioni sulla preparazione di armi o altre sostanze pericolose o su ogni altra tecnica o metodo per il compimento di atti di violenza ovvero di sabotaggio di servizi pubblici essenziali.

§

[39] In termini, A. CISTERNA, *L'evanescenza delle condotte condiziona le norme sul terrorismo*, in Guida al diritto, 2025 n.16, pag.10 ss

1.1.1. La nuova fattispecie dell'art. 270-quinquies del codice penale e i rischi di (eccessiva) anticipazione delle soglie di tutela

I descritti interventi normativi, tanto in ambito internazionale quanto in ambito di normativa interna, hanno come caratteristica l'anticipazione dell'intervento penale (tramite fattispecie di pericolo o, in certi casi, l'incriminazione degli atti preparatori o, ancora, il ricorso allo strumento di prevenzione) in quanto il terrorismo prende di mira beni giuridici primari (la vita e l'integrità fisica dei cittadini, le fondamentali strutture politiche, costituzionali, economiche e sociali di un paese), la cui rilevanza è indiscussa.

Tale anticipazione della tutela, necessaria a prevenire la irreversibile lesione di tali primari beni giuridici, deve comunque sempre mantenersi nell'alveo del rispetto dei principi generali del diritto penale per garantire un bilanciamento tra ragioni della sicurezza e rispetto dei diritti fondamentali[40]. Si tratta di un bilanciamento necessario per evitare che venga introdotta una legislazione emergenziale come tale derogatoria dei principi fondamentali anziché un complesso normativo volto ad affrontare il fenomeno criminoso attraverso la costruzione di una legislazione europea che consenta attraverso l'armonizzazione delle legislazioni degli Stati membri di contrastare più efficacemente una minaccia globalizzata e, al tempo stesso, parcellizzata e "mobile".

[40] Ex multis cfr. A. VEDASCHI, *The dark side of counterterrorism: arcana imperii and salus rei publicae*, in The American Journal of Comparative Law, 66, 4, 2018, pp. 877-926; G. DE MINICO, *Costituzione. Emergenza e terrorismo*, Jovene, Napoli 2016; C. BASSU, *Terrorismo e costituzionalismo. Percorsi comparati*, Torino, Giappichelli, 2010; T. GROPPI (a cura di), *Democrazia e terrorismo*, Napoli, Editoriale Scientifica, 2006; P. BONETTI, *Terrorismo, emergenza e costituzioni democratiche*, Bologna, Il Mulino, 2006; T. E. FROSINI, *Il diritto costituzionale alla sicurezza, in forumcostituzionale.it*, 2006; G. DE VERGOTTINI, *Guerra e Costituzione*, Bologna, Il Mulino, 2004

Sul punto, in dottrina forti perplessità critiche ha sollevato la recente introduzione del citato art. 270 quinques ad opera del d.l. sicurezza[41].

In particolare, è stato sottolineato come l'entrata in vigore della nuova fattispecie può avere come conseguenza quella di criminalizzare la "mera detenzione di una tipologia di testi scritti", situazione questa che nella realtà potrebbe rendere difficile discernere le ipotesi in cui tale detenzione sia finalizzata al compimento di azioni di terrorismo rispetto a quelle in cui i testi siano "detenuti per i motivi più disparati di studio, collezionismo, curiosità"[42].

---

[41] In dottrina, sul d.l. n. 48 del 2025, v.: AA.VV., *Il Dl Sicurezza/1*, in *Guida al diritto*, 2025, n. 16, pagg. 10 ss. con i contributi di A. CISTERNA, *L'evanescenza delle condotte condiziona le norme sul terrorismo*, *ibidem*, pagg. 82 ss.;

[42] L. POMPILI, *Il dissenso nelle nuove fattispecie di reato e nelle aggravanti introdotte con il DDL sicurezza*, in *Penale diritto e procedura,* 15 ottobre 2024. Sul punto, il chiaro riferimento, al contenuto del testo previgente nell'ultima parte dell'art. 270-*quinquies* cod. pen. ai «*comportamenti univocamente finalizzati*» al compimento di atti di terrorismo ha orientato la Suprema Corte a sottolineare la necessità che, in tale ambito, possano assumere rilevanza esclusivamente condotte, seppur di auto-addestramento, che abbiano una rilevanza sul piano materiale, così da essere idonee a realizzare la finalità terroristica oggetto dell'elemento psicologico (tra le altre, cfr. Sez. 5, n. 22066 del 06/07/2020, Barhounni, Rv. 279495-02 così massimata: "Ai fini della configurabilità del reato di addestramento ad attività con finalità di terrorismo anche internazionale, commesso dalla persona che abbia acquisito autonomamente informazioni strumentali al compimento di atti con la suddetta finalità, è comunque necessario che il soggetto agente ponga in essere comportamenti significativi sul piano materiale, univocamente diretti alla commissione delle condotte di cui all'art. 270-*sexies* cod. pen., senza limitarsi ad una mera attività di raccolta di dati informativi o a manifestare le proprie scelte ideologiche. (Fattispecie in cui la Corte ha ritenuto configurabile in sede cautelare il reato di cui all'art. 270-*quinquies* cod. pen. sulla base di molteplici indici fattuali concreti, tra i quali: il possesso da parte dell'imputato di video ed immagini riconducibili alla propaganda terroristica per lo Stato islamico o illustrativi di tecniche per la preparazione di ordigni esplosivi, scaricati con elevata frequenza nell'arco di un significativo periodo di tempo, nonché di appunti manoscritti riproducenti la celebrazione di simboli e delle pratiche terroristiche dell'"Isis" e in cui l'indagato si proclamava "servo di Allah" votato al martirio; la partecipazione a chat di gruppo e canali di propaganda jihadista nei quali venivano manifestati propositi terroristici e di esaltazione del martirio e della guerra santa contro gli infedeli; il rinvenimento all'interno della sua abitazione di materiale destinato alla fabbricazione di un ordigno rudimentale); vedi anche Sez. 1, n. 47479 del 16/07/2015, Alberti, Rv. 265405-01 così massimata : "Per ritenere integrata la finalità di terrorismo di cui all'art. 270-*sexies* cod. pen., non è sufficiente che il soggetto agente abbia intenzione di recare un grave danno al Paese, ma è necessario che la

A livello di considerazione personale, deve anche aggiungersi che tale norma rischia di esporre a responsabilità le stesse piattaforme digitali. Infatti, come già precedentemente esposto, in base al DSA, tutti gli intermediari (come i social network, siti web con contenuti generati dagli utenti, motori di ricerca, ecc.) devono dotarsi di procedure di moderazione trasparenti e di sistemi di notifica e rimozione (notice-and-takedown) per i contenuti illegali. Ciò significa che se una piattaforma digitale detiene contenuti illeciti (ad esempio video che incitano alla propaganda), la piattaforma ha il dovere di rimuoverli tempestivamente una volta segnalato, secondo le procedure previste dal DSA.

Ulteriori e diversi interventi critici sono stati sollevati in un intervento del massimario della Corte di cassazione[43] che ha inteso evidenziare come l'eccessiva anticipazione della soglia di tutela contrasti con i principi costituzionali di offensività e

---

sua condotta crei la possibilità concreta, per la natura e per il contesto obiettivo dell'azione e degli strumenti di aggressione in concreto utilizzati, che esso si verifichi, nei termini di un reale impatto intimidatorio sulla popolazione, tale da ripercuotersi sulle condizioni di vita e sulla sicurezza dell'intera collettività, posto, che solo in presenza di tali condizioni, lo Stato potrebbe sentirsi effettivamente coartato nelle sue decisioni. (In motivazione, la Corte ha precisato che l'accertamento dell'idoneità in concreto della condotta deve essere effettuato applicando il paradigma della prognosi postuma e facendo riferimento ai criteri, indicati dalla norma, della "natura e contesto" dell'azione)*".*

[43] Dl sicurezza: la relazione dell'ufficio del massimario della Corte di cassazione in *Sistema penale.it* giugno 2025. Negli stessi termini, AIGA, i contributi scritti depositati presso le Commissioni riunite I e II del Senato della Repubblica, nel corso dell'attività conoscitiva sul d.d.l. Atto Senato n. 1236: AIGA (ASSOCIAZIONE ITALIANA GIOVANI AVVOCATI), 22 e 23 ottobre 2024; FILIERA pag. 3: "si intende attribuire rilevanza penale al c.d. pericolo del pericolo, costituito dal mero procacciamento o dalla semplice detenzione del materiale contenenti istruzioni per il compimento di atti di terrorismo, a prescindere dalla realizzazione concreta di ulteriori azioni rilevanti ex art 270-*sexies* c.p.". Cfr. anche, C.S.M., *Parere*, pag. 11: "Si tratta di una chiara opzione di anticipazione della tutela del bene protetto: a prescindere da eventuali dubbi relativi al rispetto del principio di offensività in materia penale – già da questo Consiglio segnalati in relazione alle modifiche apportate all'art. 270-*quinquies* dal decreto-legge n. 7 del 2015 (cfr. delibera consiliare del 18 marzo 2015, laddove si evidenziava che l'estensione dell'area di punibilità ai terroristi che operano sganciati da sodalizi e da organizzazioni, oltre a comportare un inedito arretramento della soglia della rilevanza penale sino al compimento di atti meramente preparatori, avrebbe interpellato la capacità dell'interprete di assicurare il rispetto del principio di necessaria offensività della condotta*")*.

legalità.

A tale posizione del massimario della Corte di Cassazione si è contrapposta la posizione del Ministro della giustizia Carlo Nordio, il quale ha sottolineato la necessità di tali interventi per adeguare la normativa interna a quella internazionale[44].

Sul punto, nella Relazione illustrativa al d.l. sicurezza viene specificato che la nuova previsione normativa intende evitare che, "come dimostrato dalla esperienza investigativa e giudiziaria, le persone trovate in possesso di documentazione ascrivibile a gruppi terroristici internazionalmente riconosciuti possano andare esenti da una contestazione penale per la parte relativa alla mera detenzione documentale". Secondo il legislatore, la condotta merita di avere rilevanza penale, in quanto "il procacciamento di materiale idoneo a facilitare la commissione delle suddette attività sovversive costituisce condotta di per sé allarmante e pericolosa, a livello sociale, indipendentemente dalla effettiva realizzazione di atti terroristici, in quanto sintomatica di una progressione capace di portare repentinamente alla commissione di atti violenza con finalità di terrorismo.

Le contrapposte posizioni in questione evidenziano come nonostante le affermazioni teoriche non sia sempre facile nella realtà pratica trovare quel giusto punto di bilanciamento tra ragioni della sicurezza e rispetto dei diritti fondamentali.

§§§§§

2.Le prospettive di impiego dell'intelligenza artificiale da parte del terrorismo: IA il futuro del terrorismo?

[44] C. NORDIO, *D.L. sicurezza quei giudici irriverenti verso il Colle. Danno per tutte le toghe* in www.ilmessaggero.it del 02.07.25

Esaurito l'*excursus* sulle modalità di contrasto del terrorismo in ambito di normativa sovranazionale ed interna, torniamo ora ad esaminare le modalità di utilizzo di IA da parte delle organizzazioni terroristiche e le interferenze con le relative fattispecie penali.

Sul punto deve essere rilevato come la "qualità" da parte delle indicate organizzazioni terroristiche di gestire le tecnologie avanzate e le piattaforme digitali si è dimostrata essenziale rispetto ad alcune delle caratteristiche di fondo del terrorismo c.d. islamico e dei suoi fattori di forza che, come abbiamo visto nei paragrafi precedenti, sono costituiti dalla propaganda, dall'arruolamento, dall'addestramento e dall'auto addestramento, dalla radicalizzazione funzionale al passaggio all'azione, dalla chiamata all'"egira", dalla rivendicazione di attentati e quindi ancora dalla propaganda attraverso la rivendicazione. In relazione a tali caratteristiche l'amplificazione dei discorsi di odio attraversi i sistemi di IA assume una importanza fondamentale.

Tale riscontrata qualità pone inesorabilmente la necessità di analizzare le conseguenze che possono verificarsi allorquando "*il terrorismo governa ed impiega l'IA"* e parallelamente di rispondere alla seguente domanda: "*IA è il futuro del terrorismo?"*.

Sul punto la più attenta dottrina[45] ha evidenziato come le nuove minacce da parte delle organizzazioni terroristiche non risultano limitate alla amplificazione dei discorsi di odio mediante l'utilizzo di IA ma riguardano più in generale le seguenti macroaree:

- *cyber threats*, nel cui ambito rientrano le tematiche degli attacchi informatici;

[45] M.ROMANELLI, *Intelligenza artificiale, influenza sul mercato politico e reati contro la personalità dello Stato. La criminalità terroristica,* in Sistema Penale 29 giugno 2023. L'autore esamina le due modalità di gestione di IA da parte delle organizzazioni terroristiche sottolineandone i rischi connessi al loro utilizzo allorchè tali strumenti utili per la collettività possano trasformarsi in dei rischi per l'intera collettività.

- *psysical threats*, nel cui ambito rientrano l'utilizzo di droni e delle self-driving car;
- *political threats,* nel cui ambito rientrano i citati discorsi di odio e più in generale la tematica dei deep-fake.

Procediamo con ordine.

§§§

2.1. Cyber threats: le modalità di aggressione attraverso gli attacchi informatici.

Nell'epoca della rivoluzione digitale, le minacce informatiche riguardano la vita di tutti i cittadini, delle aziende e delle istituzioni ognuna delle quali nella propria quota parte è chiamata ad assumersi le proprie responsabilità per rendere sicuri i propri domini nei quali vengono esercitati diritti fondamentali.
In tale ambito si colloca la tematica dei *cyber attacks* che rappresentano una modalità di aggressione dei sistemi informatici realizzabile attraverso[46]:

- la distruzione che può essere totale o parziale, duratura o limitata nel tempo[47];
- l'esfiltrazione di dati rilevanti, con successivi impieghi

[46] I reati informatici sono stati introdotti nel Codice penale con la l. 23 dicembre 1993, n. 5477, in recepimento della raccomandazione del Consiglio d'Europa. Il provvedimento legislativo rappresenta uno dei primi e maggiormente significativi interventi di riforma nell'ambito della criminalità informatica, finalizzato a fronteggiare il fenomeno dei c.d. hackers. L'art. 4, in particolare, ha aggiunto nella sezione del codice penale riservata all'inviolabilità del domicilio un pacchetto di tre fattispecie incriminatrici, ed in particolare gli artt.615 ter, 615 quater, 615 quinques tutte costruite attorno alla tutela del sistema informatico o telematico da potenziali interferenze abusive ed eventualmente pregiudizievoli per la riservatezza dei dati in esso contenuti.
[47] Nel nostro ordinamento la condotta in questione è idonea ad integrare la disposizione fulcro dei computer crimes ovverosia l'art. 635 bis c.p., che disciplina il reato di danneggiamento di informazioni, dati e programmi informatici Tale norma è stata di recente interessata da una importante e più ampia novella legislativa in forza dell'art. 16, l. 28 giugno 2024, n. 90, in materia di rafforzamento della *cybersicurezza* nazionale e di reati informatici.

criminali dei dati stessi[48].

Nel settore degli attacchi informatici le aggressioni avvengono normalmente attraverso tecniche definite "basiche" quali quelle note come DoS o DDos[49] entrambe dotate di enorme capacità di moltiplicazione dei risultati di offesa.

Tale tipologia di attacchi è stata reiteratamente utilizzata negli anni anche da organizzazioni terroristiche riconducibili all' *Islamic State,* situazione questa che nel nostro ordinamento sul versante repressivo ha indotto gli uffici inquirenti a contestare la fattispecie associativa in concorso con i reati fine di natura informatica. In particolare, con riferimento ai caratteri di tali

[48] Nel nostro ordinamento la condotta in questione è idonea ad integrare la disposizione fulcro dei computer crimes ovverosia l'art. 615-ter c.p., che disciplina il reato di accesso abusivo a sistema informatico/telematico protetto da misure di sicurezza. Tale norma è stata di recente interessata da una importante e più ampia novella legislativa in forza dell'art. 16, l. 28 giugno 2024, n. 90, in materia di rafforzamento della *cybersicurezza* nazionale e di reati informatici. La novella legislativa in questione ha modificato – in maniera a ben vedere draconiana – il trattamento sanzionatorio delle ipotesi aggravate di cui ai commi 2 e 3 a cornice edittale della pena di cui al secondo comma, da uno a cinque anni, è stata innalzata a due-dieci anni, mentre quelle previste dal terzo comma sono state aumentate a tre-dieci e quattro-dodici anni. In parallelo, anche il contesto normativo europeo si sta adeguando: nel 2024 sono entrate in vigore nuove direttive e regolamenti – tra cui la NIS2, il Cyber Resilience Act e il Digital Operational Resilience Act (DORA) – che impongono standard più elevati per la sicurezza delle infrastrutture digitali e responsabilizzano maggiormente anche le piccole e medie imprese, spesso meno attrezzate.

[49] In informatica, "DoS" può riferirsi a due concetti principali: Disk Operating System (DOS) e Denial of Service (attacco DoS). DOS è un tipo di sistema operativo, mentre un attacco DoS è un tipo di attacco informatico; un attacco DDoS (Distributed Denial of Service) è un tentativo dannoso di interrompere il normale traffico di un server, servizio o rete, sovraccaricandolo con un flusso di traffico Internet. In pratica, l'attacco mira a rendere il sistema preso di mira inaccessibile agli utenti legittimi, inondandolo di richieste false o indesiderate. Come emerge dal primo Cyber Security Report pubblicato da TIM e dalla Cyber Security Foundation il 12 giugno 2025, nell'ambito di una indagine che fotografa l'evoluzione degli attacchi digitali in Italia, gli attacchi DDoS sono aumentati del 36% rispetto all'anno precedente. La media è stata di 18 eventi al giorno, ma non è solo il numero a impressionare: quasi 4 attacchi su 10 hanno superato i 20 Gbps di intensità, un livello che rende più difficile sia il rilevamento che la difesa. Si è assistito anche a un'evoluzione nelle modalità d'esecuzione, con attacchi multipli che colpiscono contemporaneamente più punti della stessa organizzazione – siti, reti, dispositivi – rendendo inadeguate molte delle contromisure tradizionali. È significativo notare come questi attacchi abbiano interessato sempre di più anche la Pubblica amministrazione, la cui esposizione è passata dall'1% al 42% del totale in un solo anno, segno di un cambio di strategia da parte degli attori malevoli e di un contesto geopolitico sempre più instabile.

associazioni riconducibili all'art.270 bis c.p. i pronunciamenti giurisprudenziali in materia[50] hanno validato tale impostazione evidenziando che ai fini della configurabilità del delitto di associazione sovversiva con finalità di terrorismo internazionale, la necessità di una struttura organizzativa effettiva e tale da rendere possibile l'attuazione del programma criminale non implica necessariamente il riferimento a schemi organizzativi ordinari, essendo sufficiente che i modelli di aggregazione tra sodali integrino il minimum organizzativo richiesto a tale fine. In particolare, le sentenze dei giudici di legittimità ritengono che tali caratteri debbano ritenersi sussistenti anche con riferimento alle strutture "cellulari" proprie delle associazioni di matrice islamica, caratterizzate da:

- estrema flessibilità interna;
- capacità di rimodularsi secondo le pratiche esigenze che, di volta in volta, si presentano;
- capacità di operare anche contemporaneamente in più Stati, ovvero anche in tempi diversi e con contatti fisici, telefonici o comunque a distanza tra gli adepti anche connotati da marcata sporadicità, considerato che i soggetti possono essere arruolati anche di volta in volta, con una sorta di adesione progressiva ed entrano comunque, a far parte di una struttura associativa saldamente costituita.

Da tali caratteristiche consegue che, in tali evenienze, l'organizzazione terroristica transnazionale assume le connotazioni, più che di una struttura statica, di una "rete" in

[50] Cass. pen., Sez. feriale, Sent., (data ud. 12/09/2013) 16/12/2013, n. 50620 massime collegate: Uff. indagini preliminari Catania, Decreto, 15/07/2019 Cass. pen., Sez. V, Sentenza, 11/06/2008, n. 31389 (rv. 241175) Cass. pen., Sez. III, 02/12/2004, n. 8296 (rv. 231243) (nella specie, si trattava di una cellula nazionale collegata al movimento internazionale denominato Anonymous, responsabile di una pluralità di delitti di accesso abusivo a sistemi informatici, danneggiamento di sistemi informatici, detenzione e diffusione abusiva di codici di accesso a sistemi informatici o telematici, interruzione illecita di comunicazioni informatiche e telematiche).

grado di mettere in relazione soggetti assimilati da un comune progetto politico-militare, che funge da catalizzatore dell'*affectio societatis* e costituisce lo scopo sociale del sodalizio[51].

Con riferimento a tale tipologia di attacchi, deve ulteriormente rilevarsi come le problematiche in materia non riguardano solamente il citato versante repressivo e sanzionatorio ma coinvolgono prevalentemente il settore della prevenzione[52]. Al riguardo gli studiosi della materia[53] sottolineano che le infrastrutture informatiche strategiche (economiche, militari, sistema di trasporta, salute, istruzione) devono avere livelli di protezione adeguati al livello della minaccia; la minaccia terroristica globale cresce e conseguentemente deve crescere la capacità di protezione. In questo caso, chiaramente, l'IA non assume la veste di aggressore ma quella di protezione del sistema.

Non molto dissimili da quelle sin qui esposte, sono le considerazioni di base in ordine ad altre note tipologie di aggressioni informatiche quali ad esempio i *malware[54], i ransomware[55]; i man-in-the-middle.*

---

[51] Per tutte cfr. Cass., Sez. 5^, n. 31389 dell'11/06/2008, Bouyahia, Rv 241175

[52] Nel nostro ordinamento un ruolo fondamentale è svolto dal CNAIPIC centro anticrimine della Polizia postale che ha il compito di tutelare le infrastrutture critiche prevenendo gli attacchi informatici. Ancor di più quando le strutture risultano tra di loro interconnesse in quanto un attacco informatico ad una infrastruttura può significare anche un'estensione massiva ad organi dello Stato, a infrastrutture

[53] M.ROMANELLI *op.cit.*

[54] Nel gennaio del 2017 le cronache dei principali media italiani si occuparono di una vicenda giudiziaria che aveva ad oggetto l'esfiltrazione grazie a un malware chiamato EyePyramid di una mole difficilmente quantificabile di dati, violando sistemi informatici di decine di professionisti (molte vittime, per esempio, erano avvocati) ma soprattutto di molti centri nevralgici del nostro Paese, tra cui i ministeri degli Esteri, degli Interni, il Porto di Taranto, la Regione Lazio. Una frenetica attività investigativa portò gli inquirenti negli Stati Uniti, a bussare alla porta della società AfterLogic

[55] Dal citato rapporto della Cyber Security Report pubblicato da TIM e dalla Cyber Security Foundation il 12 giugno 2025 con riferimento al ransomware – la tecnica che consiste nel bloccare o criptare dati sensibili per poi chiedere un riscatto – viene evidenziato come lo stesso continui a rappresentare una minaccia concreta. Con 146 casi ufficialmente rilevati nel 2024, l'Italia è il secondo Paese dell'Unione Europea più colpito. A essere presi di mira sono stati soprattutto i settori più vitali per l'economia: il 58% degli attacchi ha interessato il mondo dei servizi, mentre un altro 26% ha riguardato la manifattura. Una delle cause dell'espansione del

Anche con riferimento a tale tipologia di attacchi informatici, la cui potenzialità è amplificata dalla capacità degli algoritmi di *machine learning*[56], le modalità di aggressione non pongono rilevanti problematiche sul versante repressivo sanzionatorio riguardando prevalentemente quello preventivo.

Tornando ad una delle modalità degli attacchi informatici e precisamente quelli relativi alla esfiltrazione di dati rilevanti, in Italia, come esempio di tale tipologia di attacco, viene menzionata l'aggressione informatica indirizzata nel 2015 contro una delle società leader a livello mondiale nella produzione e vendita di sofisticata tecnologia informatica, anche a Paesi stranieri, con sede a Milano. Nel caso di specie era avvenuto che all'esito dell'attacco informatico era stata effettuata, sulla rete, la pubblicazione molto rilevante di dati sensibili, compreso il codice sorgente relativo al più raffinato sistema di presa di controllo di telefoni cellulari del tempo, che era venduto ad una pluralità di società che a loro volta offrivano la propria tecnologia al servizio di indagini, anche della Procura della Repubblica di Milano oltre che di altre Procure Distrettuali. Il risultato immediato di tale pubblicazione era stato l'interruzione delle attività di intercettazione telematica ed ambientale che erano in corso in più procedimenti per fatti di terrorismo internazionale e la Procura di Milano dovette chiudere da un giorno all'altro attività di intercettazione molto utili, e cambiare radicalmente i programmi del proprio intervento, tra l'altro in un procedimento in cui il rischio di azione terroristica contro obiettivi situati all'interno

fenomeno è la diffusione del cosiddetto Ransomware-as-a-Service: gruppi criminali che sviluppano software malevoli e li mettono a disposizione di altri soggetti, ampliando la platea degli attaccanti anche a chi ha meno competenze tecniche.

[56] Il machine learning (apprendimento automatico) viene comunemente definito come un sottoinsieme dell'intelligenza artificiale che consente ai sistemi di apprendere dai dati e migliorare le proprie prestazioni senza essere esplicitamente programmati. Invece di seguire istruzioni rigide, i sistemi di machine learning utilizzano algoritmi per analizzare grandi quantità di dati, identificare pattern e prendere decisioni autonome.

dello Stato italiano era molto elevato[57].

*

Svolte queste considerazioni sui distinti versanti della repressione e della prevenzione, occorre ora rivolgere lo sguardo alle modalità con cui il cyber crime si sta evolvendo.

Sul punto, il rapporto sulla cybersecurity in Italia e nel mondo, pubblicato nel marzo 2025[58], evidenzia come:

- i cyber threats c.d. noti rispetto al 2023 hanno subito un incremento del 27%, a causa anche del conflitto in Ucraina, che ha innestato una nuova fase definita di "guerra cibernetica diffusa" dal 2022;
- il Cybercrime si conferma la principale causa degli incidenti, con oltre l'86%, seguito da attività di hacktivism generalizzata (8%), di espionage/sabotage (4%) ed information Warface (2%);
- le principali vittime degli attacchi informatici risultano essere multipli bersagli (per il 17,8%), enti governativi e la sanità (che condividono la percentuale del 13,3%), relegando sotto al 10 % le attività finanziarie, l'educazione, il manufacturing, le news ed i media, ecc.

Sul versante della sicurezza, il report ha registrato oltre 69 milioni di eventi di sicurezza, caratterizzato da un aumento vertiginoso di infezioni da malware (+ 131%), di attacchi DDoS (del 100%) con un incremento del 167% della distribuzione della banda aggregata media in Tbps e una crescita marcata degli attacchi di maggiore intensità (>100 Gbps), nonché di attacchi alla Pubblica amministrazione (+ 155%).

Si tratta di dati dai quali si può ragionevolmente ricavare che le sfide future inevitabilmente riguarderanno la crescente pressione degli attacchi e l'adeguamento alle normative in evoluzione, ragion per cui diventerà fondamentale implementare pratiche di

[57] Tali considerazioni vengono sviluppate da M. ROMANELLI, *op.cit*

[58]https://www.sistemapenale.it/it/scheda/cyber-and-ai-crime-rassegnadelle novità dicembre 2024-aprile 2025

security by design e migliorare i processi di gestione degli incidenti. Pertanto, la cultura della sicurezza deve essere sviluppata in sinergia tra scuole, università e soggetti pubblici e privati, per garantire un ecosistema sicuro per tutti.

A livello interno, devono, altresì, essere segnalati i risultati contenuti nella Relazione annuale Intelligence 2025 al Parlamento sulla Politica dell'Informazione per la Sicurezza[59].

Il documento redatto dall'organizzazione "Sistema di informazione per la sicurezza della Repubblica" affronta la crescente complessità delle minacce cyber alla sicurezza nazionale. Un aspetto importante trattato dalla Relazione in questione riguarda l'innovazione tecnologica in relazione alla quale vengono sottolineate le nuove opportunità di sviluppo, ma allo stesso tempo anche nuovi spazi di vulnerabilità. Infatti, si evidenzia come la trasformazione digitale e la crescente interconnessione dei sistemi sociali e tecnologici hanno amplificato le possibilità di attacchi da parte di attori ostili determinando un aumento degli attacchi alle amministrazioni centrali dello Stato, con particolare riguardo alle infrastrutture digitali, dell'energia e dei trasporti. Al riguardo, in Italia, la percentuale di crescita degli attacchi è del 15,2% rispetto al 2023, attestandosi polarmente tra incidenti di Cybercrime (per il 78%) e di hacktivism (per il 22%), verso bersagli in prevalenza di news/multimedia (per il 18%), mentre manufacturing è al secondo posto (con il 16%) degli attacchi, seguito da Government (10% del totale), che nel 2023 occupava il vertice della graduatoria.

Un altro elemento significativo della Relazione riguarda l'attenzione crescente verso l'intelligenza artificiale (IA) rispetto alla quale viene ribadito come la stessa IA rappresenti sia un'opportunità, grazie alla sua capacità di velocizzare e migliorare l'attività di Intelligence, che una potenziale minaccia per la

[59] https://www.sistemapenale.it/it/scheda/cyber-and-ai-crime-rassegna-delle-novita-dicembre-2025-aprile-2025 Cyber- and AI- crime

sicurezza delle democrazie, in particolare per il suo impatto sul lavoro, sulle applicazioni militari e sulla gestione dell'informazione.

§§§

2.2. Psysical threats: le modalità di aggressione attraverso l'utilizzo di droni e del *self-driving car*. La c.d. weaponization

Nell'ambito dei psysical threats devono essere collocati due distinti settori nei quali l'avvento delle nuove forme di tecnologia IA possono operare:

- quello relativo all'impiego di droni;
- quello relativo all'impiego di autovetture a guida autonoma (senza conducente) attraverso la presa di controllo del sistema di guida.

Da questo punto di vista il citato utilizzo della IA nei Psysical threats rappresenta uno strumento per la moltiplicazione degli attacchi terroristici e per la maggiore efficacia degli stessi, fermo restando che in tale ambito il livello della minaccia rimane alto a prescindere dall'impiego delle nuove forme di tecnologia legate all'IA.

A proposito della maggiore efficacia degli attacchi, nel caso dei droni il possibile utilizzo di IA serve, a volte, a far crescere una serie di allarmi non sempre giustificati, come ad esempio quando si paventa l'ipotesi che nel prossimo immediato futuro l'umanità assisterà inerte alla presa di possesso di centrali atomiche da parte di gruppi di terroristi oppure di aerei militari.

Invece, con riferimento all'ipotesi dell'uso di *self-driving car* uno dei vantaggi tradizionalmente segnalati in relazione al possibile impiego di IA è la non necessità del rischio di sacrificio della propria vita nella conduzione dell'attacco, o, in alternativa, la non necessità del rischio di "cadere prigionieri" del nemico che si

vuole abbattere.

Si tratta di vantaggi possibili per le organizzazioni terroristiche; tuttavia, in dottrina viene sottolineato come i meccanismi di radicalizzazione, particolarmente efficaci attraverso una grande capacità di propaganda, fanno riferimento all'azione individuale ed al sacrificio, fino all'attacco suicida, come ad una delle massime realizzazioni dei doveri del combattente e quindi del "vero" musulmano[60]. Già nel periodo di formazione di *Islamic State* come terrorismo territoriale, e poi durante il periodo di *Islamic State,* l'attacco suicida ha rappresentato un valore, non un rischio: valore fortemente rivendicato e propagandato attraverso una vera e propria procedura di comunicazione particolarmente efficace e che rappresentava un fortissimo collante. Infatti, nella fase di maggiore potenza di *Islamic State* il sacrificio della vita ha rappresentato una forza, non un rischio da evitare.

Questo non vuol dire che in *subiecta materia* vi sia - o vi sarà - il rifiuto di mezzi aggressivi tecnologici che prescindano dal sacrificio della vita propria, ma solo per segnalare che la radicalizzazione è stata, ed è, centrale nell'affermarsi del *jihadismo* globale e passa attraverso il profondo coinvolgimento delle persone, in carne ed ossa, con il valore aggiunto del sacrificio della vita.

È certo però che uno spettacolare attacco realizzato attraverso la più avanzata tecnologia avrebbe quell'effetto di evidenza della propria forza, con connessa disseminazione della paura e del terrore, che rappresentano il mezzo dell'agire terroristico.

Sempre con riferimento agli attacchi fisici attraverso la manipolazione di sistemi di guida autonoma di auto o di droni (per facilitare attacchi o per causare direttamente incidenti), o tramite interventi di data poisoning (per alterare i segnali di input o la loro lettura) sui sistemi di navigazione a guida elettronica di navi e aerei per massimizzare effetti di attacchi o rallentare

[60] M.ROMANELLI, *op.cit.*. In particolare, l'autore ripercorre sul punto l'evoluzione che gli attacchi fisici hanno assunto nel corso del terrorismo c.d. territoriale.

soccorsi dal punto di vista del diritto penale sostanziale, come meglio verrà approfondito in un apposito paragrafo, sempre di più assistiamo ad una situazione in cui l'autore fisico assume una posizione sempre più marginale nella realizzazione del fatto, mentre la sfida attuale – ancora solo in parte esplorata dalla dottrina – ruota intorno alla possibilità di ravvisare profili di responsabilità direttamente in capo alla macchina.

*

Nell'ambito dei psical threats deve essere trattata anche la problematica della c.d. weaponization riguardante la creazione e l'utilizzo da parte delle organizzazioni criminali di armi prognostiche – cioè metodi di analisi predittivi basati sulla I.A. e sui big data, che consentono di predire il futuro (disordini civili, epidemie, crisi economiche, risultati elettorali, ecc.) – a detrimento dell'avversario. Il discorso della weaponization riguarda anche l'utilizzo di algoritmi in combinazione con stampanti 3D per la stampa di armi: nel settore delle armi leggere esistono diverse organizzazioni che mettono a disposizione online in open source modelli digitali di armi in file CAD che possono essere scaricati e stampati in 3D.

Una considerazione a parte deve essere fatta sui sistemi di IA allorquando gli stesso vengono utilizzati anche come arma di difesa.

In dottrina[61], tra i sistemi di armamento noti sono stati evidenziati i seguenti:

- lo sviluppo da parte degli USA del sistema antimissile e antiaereo THAAD;
- il progetto Perdix, che consiste nell'utilizzo di sciami di 20 o più droni armati che operano in formazione per

[61] ONORATI, *I.A., politica e reati contro la personalità dello stato,* in ATTI DEL WORKSHOP FONDAZIONE VITTORIO OCCORSIO (Intelligenza artificiale e giurisdizione penale), 2021

svolgere un determinato compito; in Germania è stato sperimentato analogo sistema;

- il sistema missilistico autonomo Brimstone sviluppato per la Royal Air Force (Regno Unito). Si tratta di un missile progettato per ingaggiare bersagli terrestri e piccole imbarcazioni, che può funzionare in modalità fire and forget. Infatti, in questa modalità,
- il software all'interno del missile cerca un bersaglio predeterminato all'interno di una zona definita kill box. Una volta che il software riconosce un bersaglio, ordina al missile di colpirlo. Secondo quanto riferito, il Brimstone è stato usato contro obiettivi dello Stato islamico in Siria;
- i mini-carrarmati guidati da remoto (per quanto riguarda la Russia). Il veicolo senza pilota si chiama Uran-96 ed è equipaggiato con una torretta mitragliatrice, un lanciafiamme e un'arma anti-carro. Ha già sviluppato elicotteri a guida totalmente autonoma e sperimentato l'uso, diretto da remoto, di robot "killer" che attaccano da soli, senza alcun intervento umano, obiettivi autonomamente selezionati in base a parametri preimpostati;
- i robot totalmente automatizzati, autoveicoli militari a guida autonoma, droni anti-radiazioni che cercano, individuano ed attaccano i centri radar nemici senza alcun controllo e supervisione umana (Israele). Nel carrarmato Merkava IV Israeliano sono stati anche installati sistemi automatici di scoperta e soppressione del fuoco; Israele ha anche annunciato la creazione del primo blindato a guida autonoma in grado anche di recuperare e trasportare feriti (progetto Carmel).

Inoltre, gli studiosi della materia prevedono che molto presto saranno disponibili per l'impiego in scenari operativi:

- robot militari, non ricalcanti sempre le fattezze umane, ma autonomi e dotati di capacità decisionale. I prototipi in

fase di sviluppo sono di vario genere: robot da trasporto (in grado di spostare carichi di centinaia di chilogrammi), da ricognizione (in grado di raggiungere velocità di quasi 50 km/h e saltare ostacoli), acquatici, cingolati);

- robot "killer" in grado di selezionare gli obiettivi da colpire, scegliendo, al contempo, come e quando attaccarli senza alcun intervento umano. Sperimentazioni sono in atto da parte di vari Paesi, ma l'unica applicazione pratica è rappresentata da un robot-sentinella che vigila sulla zona demilitarizzata tra le due Coree, dotato di videocamere a infrarossi, mitragliatrice e lanciagranate, in grado di individuare e colpire bersagli in movimento in un raggio di 3,2 km (per il momento ancora comandato a distanza, ma sono in corso ricerche per la completa automatizzazione);
- esoscheletri integrati indossati da un operatore umano (progetti Talos negli Stati Uniti e Ratnik in Russia), primo stadio per la creazione di futuri cyborg;
- droni – tecnicamente UAV, acronimo di Unmanned Aerial Vehicle(s) –equipaggiati anche di sistemi d'arma (es. droni Predator degli USA nel conflitto con l'ISIS) e organizzati in sciami (swarm technology), con la capacità di imitare artificialmente capacità collaborative (es. sistema LOCUST negli USA), che potranno essere abbinati a sistemi di controllo avanzati grazie allo sviluppo di interfacce neurali.

§§§

2.3. Political threats: le modalità di aggressione attraverso i deepfake e la manipolazione politica. I discorsi di odio e la libertà di espressione.

Dopo gli attacchi informatici e quelli fisici, la terza modalità di interferenza dei sistemi di IA con i reati in materia di terrorismo è costituita dalla moltiplicazione dell'efficacia dei discorsi di odio a scopo di propaganda e di radicalizzazione della rete.
Si tratta di una modalità di interferenza che si realizza attraverso l'utilizzo della Generative Adversarial Networks (GAN)in virtù della quale risulta possibile sovrapporre volti, modificare espressioni e persino imitare voci, in modo quasi indistinguibile dalla realtà.
Con specifico riferimento alla materia del terrorismo, ciò avviene anche attraverso l'utilizzo delle nuove tecniche di "brain-reading" già utilizzate per curare effetti di malattie neurodegenerative che vengono trasferite in sistemi di IA allo scopo di instillare, incertezza, sfiducia, paura e rabbia in gruppi o collettività agendo sulla sicurezza psicologica.
Sul punto deve essere evidenziato come la propaganda online abbia nel tempo rappresentato uno dei più importanti fattori di successo del terrorismo c.d. islamico; allo stesso tempo le piattaforme digitali e più in generale la rete si sono rilevate il "luogo" d'eccellenza, come anche le carceri, per l'efficace funzionamento dei meccanismi di radicalizzazione. Basti pensare che ai tempi delle Torri gemelle, e durante l'organizzazione delle stragi in Europa, il sistema per la chiamata all'azione, per il consolidamento di scelte radicali, per la adesione all'idea del martirio viaggiava ancora prevalentemente su carta o su audiocassette contenenti le immagini dei martiri, gli inni al martirio, i canti di battaglia, che venivano trasportate fisicamente e clandestinamente, con assunzione di gravi rischi, nei luoghi dell'ascolto in comune e della radicalizzazione. È sufficiente ricordare qui la "capacità" di radicalizzazione che hanno avuto da

sempre le foto – ritenute pacificamente vere - delle varie forme di tortura che sono state utilizzate ad *Abu Ghraib* o le immagini degli "arancioni" di Guantanamo, che poi sono state riproposte nelle tuniche fatte indossare agli ostaggi giustiziati dal nascente Stato Islamico. Lo stesso effetto si è cercato di ottenere con le immagini terribili delle vittime dei bombardamenti, soprattutto bambini, accompagnate da frasi "classiche" di rivendicazione di attentati (*"il vostro sangue è forse diverso dal nostro?*", o frasi di contenuto analogo).

L'evoluzione tecnologica ha ovviamente contribuito in modo decisivo a questi successi ed i descritti sistemi di IA consentono ancor di più alle organizzazioni terroristiche di aumentare l'efficacia delle loro azioni: grazie a internet e ai social media, la tecnologia può raggiungere milioni di persone in tutto il mondo in pochi secondi e senza confini.

Molti lavori accademici e rapporti istituzionali evidenziano l'alto rischio di manipolazione e radicalizzazione tramite IA generativa. In tale ambito è stato evidenziato come alcuni gruppi terroristici - come ISIS, Al-Qaeda, Hamas, Hezbollah - stanno esplorando l'uso dell'IA generativa per scopi propagandistici, contenuti in linguaggi diversi, meme e immagini manipolate per intensificare l'impatto emotivo e attrarre nuovi adepti .

A titolo di esempio possono essere citati: la diffusione da parte di Hamas di immagini ingannevoli create da IA per screditare l'esercito israeliano, sfruttando arti generati per diffondere narrazioni fuorvianti; la pubblicazione di guide da parte di affiliati jihadisti per sfruttare strumenti generativi al fine di tradurre e amplificare messaggi propagandistici.

I descritti studi accademici ed istituzionali non segnalano esempi concreti di processi nelle aule giudiziarie per i fatti in questione anche in considerazione del fatto che la materia in questione, peraltro, si innesta su un difficile bilanciamento tra libertà di espressione ed esigenze di monitoraggio.

Il rischio che le organizzazioni criminali possano diventare ancora più forti manipolando l'intera comunità mondiale impone di comprendere come ed in che modo possa essere evitata la diffusione di tali contenuti violenti.

Nel 2018, in base ai dati raccolti da una ricerca guidata dal prof. Amr Al-Azm della Shawnee State University dell'Ohio[62] sul traffico di opere antiche ma estensibile anche alla materia del terrorimo, è risultato come Facebook, allo stato dei fatti, non è ancora riuscita a centrare l'indicato obiettivo di evitare la diffusione di contenuti violenti, incitanti alla violenza razzista o terroristica. Il monitoraggio effettuato sulle pagine create da circa tremila utenti, risultanti in qualche modo collegati a organizzazioni estremiste classificate dal governo degli Stati Uniti, ha messo in luce la presenza di contenuti brutali, inneggianti al terrorismo (il video di un'esecuzione; immagini di teste decapitate; tributi a «martiri» jihadisti) evidentemente sfuggiti alle maglie del controllo. Anche una semplice ricerca manuale effettuabile da chiunque utilizzando parole chiave quali Islamic State, Isis, Boko Aram o Al Qaeda dimostra tuttavia la persistenza nel sistema di molti profili riconducibili al terrorismo e questo può essere anche attribuibile all'esigenza presente nell'ordinamento degli Stati Uniti d'America di non comprimere la libertà di manifestazione di espressione individuale degli utenti.

Si tratta di temi e di argomenti che meglio verranno sviluppati nel prosieguo della trattazione ed in particolare nel capitolo quarto allorquando verrà affrontato il tema della responsabilità delle piattaforme digitali e che in questa sede impongono di comprendere fino a che punto la limitazione di tali forme di propaganda possa comprime il diritto di espressione delle persone.

---

[62] A. AL AZM e K. A. PAUL, *How Facebook Made It Easier Than Ever to Traffic Middle Eastern Antiquities*, in World Politics Review, August 2018, reperibile al sito worldpoliticsreview.com/insights/25532/how-facebook-made-it-easier-than-ever to traffic-middleeastern-antiquities

Sul punto, nell'area di influenza europea la CEDU ammette espressamente limitazioni alla libertà di espressione ritenendo legittimi quegli interventi normativi statali che prevedano sanzioni per chi si renda responsabile di incitamento all'odio o alla violenza, sebbene massima cura debba sempre essere prestata a che il diritto individuale a manifestare il proprio pensiero sia salvaguardato da eccessiva incisività. Come noto, nel nostro ordinamento l'art. 21 della Costituzione prevede espressamente il limite del buon costume insieme con limiti impliciti nel quadro dei quali si inquadrano le norme che, per esempio, puniscono la diffamazione[63].

In via generale deve essere rimarcato che il bilanciamento tra tutela dei diritti individuali ed esigenze di sicurezza rappresenta un obiettivo primario per gli ordinamenti democratici, che sono chiamati a rispondere alle minacce del terrorismo senza compromettere il regime di libertà costituzionalmente garantite. Nel contesto generale della lotta al terrorismo, il contrasto alla diffusione di contenuti che istigano alla violenza assume una valenza particolare e spinge a riflettere sulla possibilità di porre limiti alla libertà di manifestazione del pensiero, giustificati dalla necessità di preservare la sicurezza pubblica.

§§§

2.4. La risposta penalistica ai reati commessi dalle organizzazioni terroristiche utilizzando IA: problemi, limiti e prospettive.

Il descritto utilizzo dei sistemi di IA per commettere i reati da parte delle organizzazioni terroristiche involge problematiche che riguardano la teoria generale del reato e come tali riferibili anche ad altre fattispecie penali.

[63] Su tali considerazioni cfr. C. BASSU *Istigazione all'odio, terrorismo e sicurezza nell'era digitale: c'è un limite alla libertà di espressione?* in Diritto Pubblico Europeo,2019

Come abbiamo esaminato, i recenti interventi normativi dell'Unione europea si pongono il problema dei rischi dell'utilizzo di sistemi che impiegano IA in diversi campi, anche per il possibile impatto sui diritti fondamentali delle persone, ma non affrontano il tema della responsabilità sotto il profilo penale, se non nel senso tradizionale della necessità di una "governance" umana dell'IA di soggetti fisici che la controllano e governano e "devono" intervenire per "correggerla".

Restano però aperte una serie di questioni: se i sistemi di IA si traducono in condotte costituenti reato a quali condizioni è possibile punirle ed in quale ordinamento giudiziario?

Sul punto occorre rilevare che i sistemi di IA di ultima generazione sono dotati di un grado di autonomia dall'uomo tale da mettere in crisi il modello tradizionale della responsabilità indiretta di quest'ultimo per i fatti di reato verificatisi a causa del comportamento dello stesso sistema di IA.

È stato osservato in dottrina[64] che gli originari sistemi di IA, che necessitavano dell'intervento remoto umano e che offrivano risultati tendenzialmente prevedibili (rappresentando dunque un mero strumento dell'azione umana) non ponevano seri problemi di crisi dello schema penale, potendosi individuare e punire con gli schemi tradizionali l'essere umano che è possessore e gestisce un mezzo materiale per compiere la propria condotta.

Tale schema appare probabilmente ancora utilizzabile nei casi di droni e sottomarini automatizzati programmati e utilizzati per compiere attentati non altrettanto nei casi in cui il comportamento di tali sistemi di IA non sia interamente predeterminato, e come tale non prevedibile dai suoi stessi programmatori.

Ad esempio, ciò che può avvenire nel caso di movimento autonomo di bracci robotici o di auto a guida completamente autonoma, che se guardati in prospettiva cinetica, possono essere suscettibili di mutare la realtà circostante senza il bisogno di

[64] C.CORRIDORI, *Machina delinquere non potest?* in Giustizia Insieme del 19 maggio 2022

continui condizionamenti umani e possono provocare lesioni irreparabili, come la morte. Allo stesso modo, un software adeguatamente addestrato sarà capace di effettuare operazioni finanziarie sul mercato in via completamente autonoma, rischiando di porsi quale autore materiale delle condotte di manipolazione del mercato (art. 185 d.lgs. n. 58/1998).

In tali situazioni appare non solo complicato individuare un autore umano ma anche affrontare il discorso dal punto di vista della rimproverabilità alla persona fisica per il fatto di reato che si è verificato a causa di azioni impreviste della macchina dovute ad apprendimenti secondari. A ciò si aggiunga che secondo una interpretazione conforme al principio costituzionale di personalità della responsabilità penale, l'autore non potrà essere chiamato a rispondere penalmente neppure in termini di *aberratio* (che richiede pur sempre la prevedibilità in concreto dell'offesa).

Per tali ragioni un orientamento dottrinale, sul quale ora ci soffermeremo, mutuando l'esperienza della corporate criminal liability teorizza la possibilità di attribuire della responsabilità penale all'artefatto tecnologico.

§§§

2.4.1. *Machina delinquere (non) potest?*

La considerazione conclusiva del paragrafo precedente impone di rispondere al seguente quesito: *machina delinquere potest?*

Come detto, si tratta di un profilo che ruota intorno alla possibilità di ravvisare profili di responsabilità direttamente in capo al sistema di IA (cioè la macchina stessa) concependo direttamente le stesse entità intelligenti come autori del reato.

Sul punto, un recente studio[65] ha avuto il pregio di richiamare tra le primordiali tipologie di responsabilità penali in capo a "non-persone", imperniate su istanze meramente punitivo-retributive, dove all'offesa doveva rispondersi con un danno di egual misura e peso, i famosi processi che in epoca medioevale venivano attivati nei confronti degli animali, tra cui quello di un maiale processato e giustiziato in quanto reo di aver ucciso un fanciullo.
In epoca contemporanea, la più comune forma di responsabilità riferibile ad un agente impersonale è rappresentata dalla responsabilità da reato dell'ente (corporate criminal liability). In Italia, essa si configura come una responsabilità autonoma rispetto a quella della persona fisica, autore materiale del reato, e si applica in relazione a illeciti commessi da soggetti apicali o subordinati nell'interesse o a vantaggio dell'ente, laddove quest'ultimo risulti in "colpa di organizzazione" rispetto alla prevenzione dei reati della specie di quello verificatosi[66].
Ciò ha incrinato il primato dell'antropocentrismo penalistico, pur senza superarlo, in quanto l'attribuzione della responsabilità da reato all'ente passa, comunque, dall'accertamento del reato presupposto commesso dalla persona fisica, recuperando così un autore umano quale attore protagonista della scena criminale rispetto a un responsabile giuridicamente rilevante sullo sfondo, che è proprio l'ente privato.
Prendendo spunto da tale esperienza, un orientamento dottrinale minoritario[67] ritiene che non ci debbano più essere ostacoli logici e giuridici che possano impedire di concepire le macchine di IA come soggetti attivi del reato, superando l'assioma del *machina delinquere (et puniri) non potest*. Tale orientamento incentra il ragionamento sull'elemento oggettivo del reato (condotta/evento)

[65] F. COPPOLA *op.cit.* Wolters Kluwer Cedam 2025 pag.58 ss
[66] F.PALAZZO R.BERTOLDI, *Corso di diritto penale, parte generale, nona edizione* G. Giappichelli Editore, Torino 2023
[67] G. HALLEVY, "*The Criminal Liability of Artificial Intelligence Entities from Science Fiction to Legal Social Control*". Akron Intellectual Property Journal:https://ideaexchange.uakron.edu/akronintellectualproperty/vol4/iss2/1

apprezzato in termini puramente materialistici (com'è negli ordinamenti di common law) compatibili direttamente con l'operatività del sistema di intelligenza artificiale, sia che si tratti di una condotta attiva (integrata da un movimento fisicamente apprezzabile della macchina: ad esempio, il movimento di un braccio robotico) sia che si tratti di un'omissione (integrata dall'inerzia della macchina). Non muterebbe quindi il meccanismo logico di attribuzione della responsabilità; ciò che cambierebbe è solo la tipologia di sanzione, che dovrebbe essere adeguata ad incidere su una macchina (es. distruzione del corpo fisico, disattivazione e isolamento, o ancora riprogrammazione forzata, etc.) Peraltro, con una ricostruzione meritevole di interesse, l'orientamento in questione si è spinto ad ipotizzare tre paradigmi di responsabilità, tutti fondati sul presupposto necessario del riconoscimento della personalità giuridica alle entità intelligenti. Il primo, definito perpetration through another, rappresenta l'aggiornamento di quello, tradizionale, di responsabilità indiretta dell'uomo: in base ad esso, i sistemi di I.A. sono strumentalizzati per la commissione del reato da una persona umana, che potrà individuarsi nel programmatore del software o nell'utente finale, e che ne risponderà in via esclusiva. Il secondo (natural probable consequence) e il terzo (direct liability) paradigma prevedono, invece, la possibilità di individuare una responsabilità dell'entità intelligente, in via cumulativa o autonoma rispetto alla responsabilità di programmatore e/o dell'utente.

L'orientamento prevalente[68] si esprime, invece, in senso negativo evidenziando l'inadeguatezza del sistema penale, i cui comandi

[68] In Italia la dottrina penalistica (Fiorella, Donini, Mantovani, Guerra, Marinucci-Dolcini, ecc.) ritiene la tesi di Hallevy incompatibile con i principi di colpevolezza e personalità della responsabilità penale. La macchina non può essere autore di reato: il focus resta sempre sull'uomo che la programma, utilizza o controlla. Da segnalare sull'argomento A. CAPPELLINI, *"Machina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale*", in Criminalia: Annuario di scienze penalistiche, Edizioni ETS, 2018, Pisa: https://discrimen.it/wp content/uploads/Cappellini-Machina-delinquere-non-potest.pdf.

(e sanzioni) sono concepiti come diretti a "uomini", in grado di provare timore e senso di deterrenza rispetto alla violazione, e a cui si rivolge la pena in senso "rieducativo". Si obietta infatti che al riconoscimento dei sistemi di intelligenza artificiale come soggetti attivi del reato osta il principio di colpevolezza che informa di sé gli ordinamenti penali moderni. Inoltre, i sistemi di intelligenza artificiale non sono capaci di provare timore e sono quindi immuni dall'effetto dissuasivo della minaccia della pena (prevenzione generale) e, tantomeno sono in grado di cogliere l'effetto pedagogico connesso alla comminatoria legislativa della sanzione e quello di accreditamento sociale dei valori tutelati. Inutile sarebbe anche la funzione di prevenzione speciale - intesa, alla stregua del principio costituzionale della finalità rieducativa della pena, come risocializzazione - in quanto inapplicabile ai sistemi robotici e di intelligenza artificiale. La sanzione potrebbe dunque funzionare solo come mera neutralizzazione o riprogrammazione forzata, secondo concetti incompatibili con le categorie del diritto penale attuale. A livello pratico, anche tale dottrina riconosce che le uniche possibili soluzioni alternative al sistema penale sono due: o si vieta radicalmente la realizzazione di tali sistemi, in base al principio di precauzione (es. droni autonomi armati), con la conseguente rinuncia ai benefici sociali apportati dagli stessi; oppure si individua un'area di rischio consentito, attraverso complessi bilanciamenti tra l'utilità collettiva e i rischi imponderabili dei vari sistemi e tali azioni non verranno punite (es. auto a guida autonoma).

Entrambe le tesi proposte hanno il pregio di porre l'attenzione su profili di stretta attualità.

Tuttavia, l'orientamento prevalente appare largamente preferibile atteso che la forma della riposta penale si caratterizza per la sua natura antropocentrica come tale legata al concetto di "autore di condotte costituenti reato" e all'elemento psicologico del reato (in termini di dolo o colpa) e non appare immediatamente sovrapponibile nemmeno nel caso di azioni condotte da macchine

e da algoritmi di intelligenza artificiale in assenza di qualunque intervento umano.

Poiché tale problematica non può essere risolta introducendo nuove e larvate forme di responsabilità oggettiva, sulla falsariga della responsabilità per prodotto difettoso, a carico delle persone fisiche è inevitabile che solo un intervento chiaro del legislatore potrà essere dirimente sul tema in questione.

§

2.4.2. Meccanismi di azione diffusa, profili di responsabilità penale e cooperazione internazionale.

Strettamente collegata alla problematica in questione è quella legata alla individuazione delle responsabilità penali quando le azioni realizzate con metodo diffuso e attraverso macchine, programmi e bot-net nascondono la propria "identità" (reale o virtuale).

Come evidenziato in dottrina[69], in questi casi di azione "diffusa" dell'IA risulta pressoché impossibile identificare come autore di un reato un singolo soggetto nonché individuare l'origine dell'azione dell'IA che quasi sempre travalica i confini nazionali (e delle giurisdizioni) e si muove nel cyberspace in modo globale. Infatti, il cyberspazio non ubbidisce alla logica territoriale dei confini nazionali (per sua natura è infatti a-territoriale), a differenza degli ordinamenti statali che richiedono uno "spazio sul quale esercitare la propria sovranità esclusiva". Caratteristica della rete è quella di permettere all'individuo, di essere presente e operare anche simultaneamente in più "luoghi informatici"[70].

[69] C.ONORATI, *op.cit*,

[70] Sulla struttura "a rete" delle associazioni terroristiche attuali, specie di matrice jihadista (es. ISIS), capaci di operare contemporaneamente in più Stati attraverso cellule che comunicano reciprocamente a distanza e in modo sporadico, cfr. Cass., sez. V, 13 luglio 2017 (dep. 3 novembre 2017), n. 50189; Cass., Sez. VI, sent. 19 dicembre 2017 (dep. 29 marzo 2018), n. 14503.

Ciò comporta la de-temporalizzazione delle azioni, ossia programmare e automatizzare complesse operazioni senza il necessario e simultaneo "contatto fisico" tra uomo e sistema informatico. Si pensi, in particolare, alla realizzazione di criminal smart contracts dove è possibile pianificare a monte il software, la cui esecuzione causerà l'evento rilevante per la norma incriminatrice solo successivamente e al verificarsi di determinate condizioni previamente stabilite ed automaticamente eseguite. Tali sistemi "distanziano", creando tra loro un gap spaziale e temporale, l'agente fisico e la macchina che sarà autore materiale di un crimine.

Analogamente ai casi sin qui descritti, nei fenomeni di deep fake, delle azioni ai danni di sistemi economici e politici degli Stati l'azione di destabilizzazione può essere subdola, operare in simultanea da più contesti spaziali e cronologicamente ripetuta nel tempo sfruttando più "co-autori inconsapevoli" o in buona fede (si pensi ad esempio ai movimenti no-vax che in alcuni casi vengono mossi e manipolati per anni sulla base di deep fake introdotte e manipolate in rete da piccoli gruppi di individui; o ancora, a coloro i quali alimentano il proprio odio razziale o antisemita in base a finti documenti che circolano in rete e veicolati da gruppi che si dicono in possesso della "verità" e di un "sistema" di controinformazione).

*

Le situazioni sin qui descritte pongono problemi sul piano dell'applicazione della legge penale nello spazio e del *tempus commissi delicti* in relazione alle quale risultano ipotizzabili diverse soluzioni:

- una prima soluzione potrebbe essere quella di considerare il crimine commesso nel luogo in cui gli attori hanno avviato l'azione (di manipolazione o di creazione e

inserimento in rete delle notizie false). Così, in caso di azione simultanea in più contesti spaziali si potrebbe pensare ad un meccanismo simile a quello degli artt. 12-16 c.p.p.; tuttavia, risulta fonte di difficoltà applicative l'ineliminabile incertezza sul "dove" l'azione è stata "compiuta", anche in parte. Un ostacolo concreto alla verificazione del luogo geografico dell'individuo che compie azioni lesive è costituito dall'utilizzo di VPN (virtual private network) e di browser con crittografia stratificata. Dal punto di vista informatico, anche se fisicamente il soggetto agente opera da un computer situato in Italia, risulterà come connesso a server ubicati in altri paesi, creando così difficoltà anche dal punto di vista probatorio;

- un criterio diverso potrebbe essere quello che fa leva sull'evento finale (ossia sul "prodotto" o sugli effetti negativi delle azioni di I.A.), che se verificatosi in territorio nazionale potrebbe radicare la giurisdizione. Anche questo criterio non garantisce soluzioni semplici e immediate, poiché la sua applicazione potrebbe scontrarsi con la volontà di altri Stati di affermare la propria giurisdizione, con la mancanza di regole condivise a livello internazionale in tema di individuazione e acquisizione delle prove e con le difficoltà legate allo svolgimento di indagini su "entità" e soggetti responsabili dell'attacco anche in territori soggetti a sovranità e giurisdizione "altrui". Peraltro, in materia di ordine pubblico e di terrorismo in particolare, la casistica ci pone di fronte a "eventi diffusi" di cui non è semplice la localizzazione territoriale.

Dal punto di vista processuale, il carattere necessariamente transnazionale dei crimini in questione collegati all'estrema rapidità delle operazioni che possono essere condotte utilizzando

l'IA[71] rende davvero obsoleto il sistema della collaborazione internazionale basata sulla preventiva richiesta allo Stato nazionale nel cui territorio si svolga un segmento della condotta (ad esempio, per la presenza di strutture materiali utilizzate dall'agente umano) che si completa o ha riflessi nello Stato rogante, cui segue l'attesa della esecuzione. In dottrina si è invocata la possibilità di invocare i principi del diritto penale internazionale e della giurisdizione universale per consentire allo Stato "attaccato" di reagire in maniera più efficace attraverso l'esercizio diretto e immediato dei poteri reattivi, basata sul principio di territorialità della condotta (di suoi segmenti) o dei suoi effetti[72].

Tuttavia, tale posizione appare difficilmente compatibile con i principi del diritto internazionale e fonte di continui conflitti per l'affermazione di giurisdizioni concorrenti e con pretese di esclusività.

Sul punto la Convenzione di Budapest[73] si è mossa nella direzione di semplificare le procedure di cooperazione, in qualche caso ribaltandone le modalità sulla base del consenso anticipato.

---

[71] Il Rapporto congiunto Eurojust ed Europol su "Common Challenges in Cybercrime" pubblicato il 31 gennaio 2025, evidenzia la persistenza nell'anno 2024 di criticità nell'attività investigativa e di contrasto alla criminalità informatica, nonostante iniziative legislative anche recenti, come il *Cloud Act*, l'*AI Act* e il secondo protocollo addizionale della Convenzione di Budapest. Le criticità afferiscono, in particolare, alla gestione di grandi volumi di dati, alle sfide poste dai servizi di anonimizzazione e dalle tecnologie che oscurano le posizioni degli utenti, creando barriere sostanziali al tracciamento delle attività illecite ed ostacoli nella cooperazione internazionale. Si sottolinea la necessità di soluzioni adattabili alla natura dinamica delle minacce informatiche e che consentano il potenziamento delle capacità tecniche e operative delle forze dell'ordine per tenere il passo con i progressi tecnologici e contrastare preventivamente le minacce emergenti. Inoltre, le crescenti sfide relative alla conservazione dei dati, alle barriere giurisdizionali e alle complicazioni insite nei partenariati pubblico-privati richiedono un approccio che bilanci in maniera equilibrata misure di sicurezza rigorose con la salvaguardia della *privacy* e delle libertà civili.

[72] Su tali considerazioni cfr. AA.VV. raccolta degli atti del workshop della Fondazione Occorsio su *Intelligenza artificiale e giurisdizione penale.* pag 105

[73] L'Unione Europea sostiene la convenzione in questione e i suoi protocolli nell'ambito della sua strategia per la cibersicurezza; La convenzione è entrata in vigore il 1° luglio 2004; il primo protocollo addizionale è entrato in vigore il 1°marzo 2006; il secondo protocollo

Nel novembre del 2021 è stato aperto alla firma il Secondo Protocollo Addizionale della Convenzione di Budapest, promosso dal Consiglio d'Europa. Il Protocollo è volto essenzialmente al miglioramento della cooperazione internazionale tra i Paesi aderenti, sulla base della Mutua Assistenza Legale (AA). In particolare, il Protocollo affronta i temi emersi nella concreta attuazione della Convenzione riguardo; la diversità di previsioni legali negli Stati Parte; i casi e le condizioni per l'emissione degli ordini di accesso ai providers; le "interferenze" derivanti dal diverso regime della privacy tra UE e altri Paesi, tra cui gli USA. L'approccio della Convenzione e del Protocollo è dunque innanzitutto mirato al tema dell'acquisizione dei dati esterni e di quelli di traffico, nonché di contenuti, custoditi dai providers, in un ambiente non localizzato[74].

In ambito internazionale, in data 24 dicembre 2024, l'Assemblea Generale delle Nazioni Unite ha adottato la Convenzione Onu sulla criminalità informatica (Cybercrime Convention), che mira

addizionale non è ancora entrato in vigore; la decisione (UE) 2023/436 è in vigore dal 14 febbraio 2023.

[74] La tematica in questione riguarda il cloud computing e dunque tutte le problematiche derivanti dalla localizzazione dell'immagazzinamento dei dati (storage) in luoghi diversi e non sempre noti, nonché dal rapido trasferimento dei dati stessi, a discrezione del provider. Il Protocollo affronta poi il tema della Volontary Disclosure da parte dei providers come forma principale di esecuzione della MLA. Il cloud computing, tuttavia, non è la frontiera, ma solo uno dei sistemi di immagazzinamento e trattamento del dato. Per l'accertamento e la prevenzione dei reati è, invece, molto più impegnativa la prospettiva che viene dalla estrema rapidità delle operazioni e dalla loro anonimità e non-localizzazione quando le stesse operazioni vengono compiute utilizzando l'I.A., le capacità di calcolo e trasmissione enormemente sviluppati e, in un domani ormai attuale, il quantum computing. La MLA consistente nell'accesso al dato esterno o al contenuto della comunicazione sarà rilevante per i casi meno significativi dei cybercrimes e al contempo del tutto fuori tempo, visti i passaggi che – anche nella più avanzata delle applicazioni della Convenzione e del Protocollo – richiederà giorni, mentre attualmente i tempi di risposta medi sono nell'ordine di molti mesi. Si riproduce il paradossale effetto per cui la giurisdizione, cioè l'affermazione del potere tradizionalmente caratterizzante la sovranità degli Stati, deve bussare alla porta dell'operatore privato che gestisce come cosa propria un bene pubblico, in questo caso lo Spazio. Nei lavori preparatori per il Protocollo, viene sottolineato che la piena attuazione della Convenzione almeno per i profili attinenti alla MLA nel settore del cloud computing è condizione di ogni passaggio successivo.

a combattere la criminalità informatica in modo più efficiente, rafforzando la cooperazione internazionale e fornendo assistenza tecnica, in particolare, ai paesi in via di sviluppo[75]. La Convenzione, che sarà prossimamente aperta alla firma dei singoli Stati, si compone di otto capitoli, di cui i capitoli da due a cinque fissano gli obblighi di penalizzazione e le misure probatorie e processuali. Per quest'ultime è richiesto che debbano essere adottate per tutti i reati che sono commessi attraverso un sistema informatico o telematico e quando si tratta di raccogliere prove elettronicamente, e non solo per i reati previsti dalla Convenzione. Relativamente agli obblighi di penalizzazione, sono previsti 17 reati fra cui, specificamente, accanto ai reati informatici tradizionali, come l'accesso abusivo o la frode informatica, anche i reati di pornografia minorile e grooming mediante la rete, nonché quelli di diffusione non consentita di immagini intime e di riciclaggio. La Convenzione dedica particolare attenzione al contrasto della dimensione economica criminalità informatica, prevedendo disposizioni specifiche sulla cooperazione internazionale finalizzata alla confisca e sulla possibilità di concludere accordi o intese bilaterali o multilaterali per istituire organismi investigativi comuni. Per tutti i reati previsti è altresì richiesto che gli Stati, tenendo conto della loro gravità ed in conformità alla relativa legislazione interna, fissino un ampio termine di prescrizione, stabilendone uno più lungo disponendo la sospensione della prescrizione quando la persona, sospettata di aver commesso il reato, si sia sottratta all'amministrazione della giustizia.

§§§§§

[75] Cyber- and AI- crime: rassegna delle novità dicembre 24 www.sistemapenale.it/it/scheda/cyber-and-ai-crime-rassegna-delle-novita-

3.IA come strumento di prevenzione e di contrasto per i reati terroristici

Nella materia del terrorismo i sistemi di IA possono essere impiegati anche per finalità "benigne"[76].

Come visto in precedenza, l'utilizzo massiccio dei sistemi di IA oltre a costituire una attività di supporto alle decisioni giudiziarie può essere utilizzata per scopi di *law enforcement* nello svolgimento dell'attività del pubblico ministero e della polizia giudiziaria, nel cui ambito rientra la c.d. polizia predittiva.

In tale ultimo ambito, l'utilizzo dei sistemi di IA si sta sempre di più consolidando e, come visto in precedenza, si traduce in una rilevante opportunità per una più efficace lotta al crimine attraverso le seguenti modalità: cooperazione tra autorità; sinergia tra i sistemi di IA e la lotta al terrorismo; individuazione delle potenziali aree criminali e dei possibili autori di crimini efferati tra cui gli autori di reati terroristici.

§§§

3.1. Gli ambiti sinergici tra IA e la lotta al terrorismo: il riconoscimento facciale, i flussi finanziari, la radicalizzazione.

Una prima significativa applicazione dell'IA nella lotta al terrorismo internazionale riguarda l'utilizzo della tecnologia di riconoscimento facciale. Testata, seppur in via ancora sperimentale, da diverse autorità pubbliche, tale tecnologia consente l'estrazione e l'elaborazione di dati biometrici del volto delle persone riprese da telecamere installate in spazi pubblici, al fine di confrontarli con quelli archiviati in banche dati contenenti

[76] Con riferimento alla attività di polizia predittiva cfr. le considerazioni di C. ONORATI, *I.A., politica e reati contro la personalità dello stato,* in ATTI DEL WORKSHOP FONDAZIONE VITTORIO OCCORSIO (Intelligenza artificiale e giurisdizione penale), 2021.

le generalità di individui sospettati o già condannati per reati di terrorismo. In Italia, nell'ambito delle attività di pubblica sicurezza, sono oggi correntemente in dotazione alle forze dell'ordine sistemi di analisi dati, riconoscimento e allarme interamente automatizzati e utilizzabili tramite smartphone o tablet (O.D.I.N.O. in uso all'Arma dei Carabinieri e Mercurio adottato dalla Polizia di Stato). Un software apposito (S.A.R.I.) svolge funzioni di riconoscimento facciale consentendo di identificare un soggetto a partire da un fotogramma, confrontando quest'ultimo con banche dati che contengono dati biometrici e fotografie o con le immagini delle telecamere di sorveglianza di una determinata zona.

Con riferimento alla tecnica in questione, occorre preliminarmente stabilire cosa debba intendersi per "identificazione biometrica". Sul punto, l'art. 3, par. 34 del Regolamento AI ACT identifica come biometrici "i dati personali ottenuti da un trattamento tecnico specifico relativi alle caratteristiche fisiche, fisiologiche o comportamentali di una persona fisica, quali le immagini facciali o i dati dattiloscopici". Ai sensi del paragrafo 35 del medesimo articolo, l'identificazione di tali dati si realizza tramite "il riconoscimento automatizzato delle caratteristiche umane fisiche, fisiologiche, comportamentali o psicologiche allo scopo di determinare l'identità di una persona fisica confrontando i suoi dati biometrici con quelli di individui memorizzati in una banca dati".

Sul punto, le citate linee Guida della Commissione europea del 4 febbraio 2025 in applicazione del disposto dell'art. 5 dell'AI ACT vietano sistemi di IA destinati a creare *database* per riconoscimento facciale. In questi casi non è richiesto che l'unico scopo del *database* sia quello di essere utilizzato per il riconoscimento facciale, bastando che il *database* possa essere impiegato anche per tale finalità. Allo stesso modo lo *scraping* indiscriminato di immagini facciali è ritenuto sussistente quando la raccolta di dati o contenuti avvenga senza un focus specifico su

un singolo individuo o su un gruppo di individui, a prescindere dal rispetto dei protocolli di *opt-out* di internet, come robot.txt. Per la Commissione si è, però, comunque al cospetto di *scraping* vietato quando il risultato finale sia funzionalmente lo stesso di un'operazione di *scraping* indiscriminata fin dall'inizio. A fronte di ciò ed in generale per evitare impieghi che siano nei fatti elusivi dei divieti, la Commissione europea incoraggia l'adozione di misure di trasparenza e *audit* dei sistemi IA per garantire in itinere il rispetto delle norme.

Premesse tali considerazioni, nell'ambito della tecnologia del riconoscimento facciale è necessario distinguere tra sistemi di identificazione biometrica il cui impiego avviene in tempo reale e sistemi di identificazione biometrica il cui impiego avviene da remoto.

Con riferimento ai sistemi di identificazione biometrica il cui impiego avviene in tempo reale, nel Regolamento AI ACT viene stabilito un divieto generale di utilizzo negli spazi pubblici in ragione degli elevati rischi - diretti e indiretti - che simili tecnologie comportano per i diritti fondamentali (si pensi, ad esempio, alla libertà di riunione o di manifestazione del pensiero). Cionondimeno, l'art. 5 del Regolamento prevede tre eccezioni al rigido divieto in questione che rendono possibile l'utilizzo dell'identificazione biometrica in tempo reale anche nei luoghi pubblici quando si tratta:

- di una ricerca mirata di specifiche vittime di sottrazione, tratta di esseri umani o sfruttamento sessuale di esseri umani, nonché la ricerca di persone scomparse;
- di prevenire una minaccia specifica, sostanziale e imminente per la vita o l'incolumità fisica delle persone fisiche o una minaccia reale e attuale o reale e prevedibile di un attacco terroristico;
- di localizzare o identificare una persona sospettata di aver commesso un reato, ai fini dello svolgimento di un'indagine penale, o dell'esercizio di un'azione penale o

dell'esecuzione di una sanzione penale per i reati di cui all'allegato II, punibile nello Stato membro interessato con una pena o una misura di sicurezza privativa della libertà della durata massima di almeno quattro anni». I reati contenuti nell'allegato sono: il terrorismo, la tratta di esseri umani, lo sfruttamento sessuale di minori e la pornografia minorile, il traffico illecito di stupefacenti o sostanze psicotrope, il traffico illecito di armi, munizioni ed esplosivi, l' omicidio volontario e le lesioni gravi, il traffico illecito di organi e tessuti umani, il traffico illecito di materie nucleari e radioattive, il sequestro, la detenzione illegale e la presa di ostaggi, i reati che rientrano nella competenza giurisdizionale della Corte penale internazionale, l'illecita cattura di aeromobile o nave, la violenza sessuale, il reato ambientale, la rapina organizzata o a mano armata, il sabotaggio, la partecipazione ad una organizzazione criminale coinvolta in uno dei reati elencati sopra. ln tutte queste ipotesi, è fondamentale sottolineare come l'impiego della tecnologia in quanto in *real time* e conseguentemente maggiormente invasiva risulta subordinato all'adozione di "un'autorizzazione preventiva rilasciata da un'autorità giudiziaria o da un'autorità amministrativa indipendente e deve essere destinato esclusivamente a «confermare l'identità della persona specificamente interessata in presenza di comprovate ragioni d'urgenza"**.**

Con riferimento alla seconda modalità di utilizzo della tecnologia facciale costituita dai sistemi di identificazione biometrica il cui impiego avviene da remoto, l'Allegato III, par. I, l. a), b), c) include le tecnologie di identificazione biometrica "da remoto tra i sistemi di intelligenza artificiale ad alto rischio, il cui utilizzo è condizionato al rispetto dei presupposti cautelari previsti nel Regolamento. Più nel dettaglio, sono considerati ad alto rischio: i sistemi di identificazione biometrica remota», ad eccezione di

quelli destinati a essere utilizzati per la verifica biometrica la cui unica finalità è confermare che una determinata persona fisica è la persona che dice di essere; i sistemi di IA destinati a essere utilizzati per la categorizzazione biometrica in base ad attributi o caratteristiche sensibili protetti basati sulla deduzione di tali attributi o caratteristiche; i sistemi di IA destinati a essere utilizzati per il riconoscimento delle emozioni.

*

Un secondo ambito sinergico tra IA e antiterrorismo è rappresentato dalla possibilità di utilizzo di tali innovazioni tecnologiche per l'individuazione dei flussi finanziari illeciti riconducibili al finanziamento del terrorismo ed il loro congelamento, così come previsto dalla direttiva UE anti, nota anche come AML (Anti-Money Laundering), emanata dall'Unione Europea per prevenire e contrastare il riciclaggio di denaro e il finanziamento del terrorismo[77].

Con riferimento alla materia penale, la normativa richiede agli Stati membri di punire la fornitura e raccolta di capitali che venga operata "in qualsiasi modo", direttamente o indirettamente, con l'intenzione o quantomeno la consapevolezza (ammettendo quindi la configurabilità a titolo di dolo eventuale) che tali capitali saranno utilizzati per la commissione di un reato di terrorismo. Inoltre, la normativa in questione prevede che non sia necessario né stabilire che l'autore della condotta di finanziamento fosse a conoscenza dello specifico reato finanziato né, tantomeno,

[77] L'AML Package (Pacchetto antiriciclaggio UE) comprende i seguenti atti normativi: VI Direttiva Antiriciclaggio – Direttiva (UE) 2024/1640 (6AMLD), adottata il 31 maggio 2024. Modifica la Direttiva (UE) 2019/1937 (Whistleblower Protection) e abroga la precedente Direttiva (UE) 2015/849 (4AMLD); Regolamento Antiriciclaggio – Regolamento (UE) 2024/1624, anch'esso del 31 maggio 2024, funge da "single rulebook", definendo regole direttamente applicabili in tutti gli Stati membri; Regolamento (UE) 2024/1620, del 31 maggio 2024, che istituisce l'Autorità per la lotta al riciclaggio e al finanziamento del terrorismo (AMLA)

dell'effettivo utilizzo dei fondi per la commissione di eventuali reati.

L'Italia ha recepito le direttive UE antiriciclaggio attraverso il Decreto Legislativo n. 90/2017 (attuativo della quarta direttiva) e modificando il Decreto Legislativo n. 231/2007. Il recepimento della sesta direttiva è in corso e dovrà essere completato entro le scadenze previste (luglio 2027, con alcune eccezioni per specifici articoli).

A tale riguardo, la normativa italiana sul finanziamento si presenta già in linea con i citati obblighi europei, stante l'introduzione dell'art. 270-quinquies.1 c.p. avvenuta con la legge 153 del 2016.

In questo settore, così come già accaduto per l'antiriciclaggio, l'IA può risultare un utile strumento a supporto della individuazione dei capitali potenzialmente illeciti attraverso le seguenti modalità[78]:

- il miglioramento della segnalazione delle operazioni sospette. In particolare, l'IA, grazie all'analisi avanzata di grandi volumi di dati, consente di affinare la selezione delle anomalie, aumentare l'efficacia delle segnalazioni e ridurre i falsi allarmi;
- l'utilizzo di algoritmi predittivi e reti neurali avanzate Tecniche come Graph Neural Networks (GNN), reti neurali ibride (CNN-GRU), GAN (Generative Adversarial Networks) e deep reinforcement learning che

[78] Per ulteriori approfondimenti in materia F.DI VIZIO, *Prevenzione e investigazione: l'uso di IA, big data e soluzioni tecnologiche in ambito finanziario e nel contrasto al riciclaggio (AML) e al finanziamento del terrorismo (CFT),* in DISCRIMEN, 2024; F.DI VIZIO, *Prevenzione e investigazioni: l'uso di IA, Big Data e soluzioni tecnologiche in ambito finanziario e nel contrasto al riciglaggio e al finanziamento del terrorismo*. Relazione al corso "*L*a digital transformation*: evoluzione del contesto e profili di impatto di diritto penale sostanziale e processuale*", organizzato dalla Scuola Superiore della Magistratura in collaborazione con la Scuola di Polizia economico-finanziaria della Guardia di finanza, Cod. FPFP23015, 16 novembre 2023, Lido di Ostia (Roma).

consentono di identificare modelli nascosti, adattarsi a nuove strategie criminali e anticipare rischi;

- l'utilizzo di database e analisi della blockchain. In particolare, un modello AI sviluppato con MIT e IBM utilizza un dataset di 200 milioni di transazioni Bitcoin per identificare patterns tipici di riciclaggio e aumentare l'efficienza nel monitoraggio dei depositi verso exchange;
- la predisposizione di piattaforme a supporto delle istituzioni. In particolare Google Cloud ha lanciato una piattaforma AI per l'AML che riduce i falsi allarmi e migliora l'identificazione di casi reali; HSBC ha ottenuto una riduzione del 60 % negli alert, con un incremento di due-quattro volte nel tasso di rilevazioni reali; ThetaRay, startup israeliana, offre sistemi di monitoraggio delle transazioni basati su AI, utilizzati da varie banche globali per scoprire "unknown unknowns" (pattern criminali nuovi o sconosciuti); ComplyAdvantage sfrutta l'IA, il machine learning e il NLP per gestire onboarding, monitoraggio di transazioni, screening di persone politicamente esposte (PEP) e sanzioni.

*

Infine, l'IA può giocare un ruolo strategico principale per prevenire il fenomeno della radicalizzazione che attraverso il già descritto utilizzo delle piattaforme digitali e l'amplificazione dei discorsi di odio riesce a diffondere una visione integralista dell'Islam, idonea a spingere l'individuo verso una adesione totalizzante e ideologicamente estremista.

Come meglio verrà approfondito nel capitolo quarto a proposito del ruolo delle piattaforme digitali, gli algoritmi che operano on line possono fornire un fondamentale aiuto alla sua prevenzione. Ciò può avvenire tramite il c.d. *redirect method,* un sistema che utilizza algoritmi di IA per intercettare gli utenti ritenuti più

vulnerabili al discorso *jihadista.* In tali evenienze l'IA potrebbe essere addestrata in modo da riconoscere parole ed espressioni caratteristiche della propaganda estremista e reindirizzare l'utente verso contenuti di contro-narrativa, riducendo così il rischio di adesione a ideologie violente[79].

Inoltre, l'IA può essere utilizzata per la identificazione e la rimozione dei contenuti radicalizzanti, tramite tecniche di:

- *image matching,* mediante il quale il sistema di IA impara a riconoscere contenuti ritenuti pericolosi;
- *terrorist clustering,* mediante il quale il sistema di IA segnala profili collegati a soggetti precedentemente esclusi dalle piattaforme social per propaganda terroristica;
- *repeat offenders algorithms,* mediante il quale il sistema di IA individua nuovi *account* riconducibili a utenti già rimossi dalle piattaforme social per contenuti illeciti, attraverso l'analisi di parametri di identificazione indiretti.

§§§

3.2. La prevenzione predittiva dei reati (c.d. *predictive policing)*: i sistemi *place-based* e quelli *person-based.*

Passando ora al versante prevalentemente preventivo con l'espressione *predictive policing* si fa riferimento a un insieme concatenato di strumenti analitici che mediante l'elaborazione dei dati, mira a prevedere la commissione di futuri reati e a garantire un miglior presidio da parte delle forze dell'ordine delle zone di rischio.

Le predizioni si fondono su modelli statistici che integrano variabili di diversa natura, quali la localizzazione di reati

[79] A. VEDASCHI, *Intelligenza artificiale e misure antiterrorismo alla prova del diritto costituzionale,* in *Consulta Online* 17 febbraio 2020, 5

pregressi, i movimenti sospetti, la stagionalità o le condizioni meteorologiche, ma talvolta anche dati altamente sensibili e controversi, come l'origine etnica, il livello di istruzione o il profilo socioeconomico degli individui.

Secondo i principali teorici della materia, la finalità ultima della *predictive policing* consisterebbe nella riduzione dell'incertezza, al fine di ottimizzare l'allocazione delle agenzie di law enforcement.

L'importanza strategica dell'impiego di sistemi di IA nelle attività di *predictive policing* ed i preziosi risultati grazie ad essi raggiungibili, sono ben messi in evidenza nel famoso caso di Umar Farouk Abdulmutallab, noto anche come il "terrorista delle mutande".

Si tratta di un episodio riferito da G. Italiano in un suo recente saggio[80] nel quale l'autore afferma: "già nel recente passato si sono verificati casi in cui l'utilizzo di opportune (anche semplici) analisi algoritmiche avrebbe potuto prevenire il verificarsi di pericolosi eventi terroristici. Ad esempio, è famoso il caso di Umar Farouk Abdulmutallab, noto anche come il "terrorista delle mutande" (*Underwear Bomber*), che è riuscito a imbarcarsi sul volo Amsterdam-Detroit nel giorno di Natale 2009, con dell'esplosivo cucito all'interno della biancheria intima che indossava, e che ha cercato di farsi esplodere durante il volo. Per una fortunata coincidenza [l'intervento di alcuni passeggeri insospettitisi], l'attacco terroristico non è andato a buon fine [...]. L'*intelligence* aveva dati e informazioni a sufficienza per valutare il grado di pericolosità del terrorista e aveva anche elementi sufficienti per inserirlo nella *black list*, così da negargli la possibilità di imbarco su voli diretti negli Stati Uniti. Ma in quel caso l'*intelligence* non è semplicemente riuscita a connettere le molteplici informazioni provenienti da varie fonti, che erano a sua disposizione. Come dire, nel patrimonio informativo c'erano tutti

[80] G.F. ITALIANO *Intelligenza artificiale: passato, presente, futuro*, cit., p. 222

i dati necessari, ma è semplicemente mancata l'utilizzazione di un buon algoritmo per mettere in correlazione tutti questi dati".

Riflettendo sul caso in questione, l'autore[81] rileva, quindi, che: "sicuramente, e non soltanto in questa circostanza, tecniche di IA sono e possono essere impiegate con successo nell'analisi delle informazioni disponibili, delle transazioni, dei file di *log*, del traffico sulla rete, e di tutte le "impronte" che ogni individuo lascia in rete e nei sistemi digitali, allo scopo di identificare possibili anomalie e attività sospette, o semplicemente per comporre in una visione coerente le informazioni provenienti da sorgenti multiple ed eterogenee, ed estrarne conoscenza, in modo tale da prendere in maniera automatica decisioni oppure fornire il supporto a decisori umani, che devono essere in grado di reagire sempre più velocemente agli stimoli esterni".

La disponibilità dei sistemi di intelligenza artificiale ha determinato un'evoluzione significativa di queste tecniche grazie alla loro capacità di processare grandi quantità di informazioni individuando correlazioni spesso non percepibili all'operatore umano e colmando al contempo il deficit di risorse delle forze di polizia.

Al riguardo deve segnalarsi che i settori di maggiore sviluppo hanno riguardato di due diverse tipologie di software di polizia preventiva: i sistemi *place-based* ed i sistemi *person-based.*

*

I sistemi *place-based* hanno come peculiarità quella di individuare i cosiddetti hotspot ossia le aree a maggior rischio di commissioni di reati. L'obiettivo di tali sistemi è quello di concentrare la presenza delle forze dell'ordine in determinate zone, ritenute statisticamente più esposte, così al fine di dissuadere la commissione di reati ovvero favorire l'arresto in flagranza, sulla base dell'assunto che alcuni reati generano effetti

[81] G.F. ITALIANO *op.cit.*

di emulazione o fenomeni a cascata delle aree contigue. In Italia, un esempio paradigmatico è rappresentato dal *software X-LAW,* sviluppato dalla Questura di Napoli, che consente di mappare il territorio in funzione del rischio criminale, individuando orari e zone particolarmente critici per la predisposizione di interventi mirati. In particolare, il software in questione, attraverso l'approccio del machine learning, attinge ai dati provenienti dall'archivio denunce e li compara con quelli sulle caratteristiche socioeconomiche e demografiche locali e sugli eventi in programma, per fornire precisi avvisi geolocalizzati circa l'alta probabilità di verificazione di un crimine in una data area (specie per i reati predatori urbani).

*

I sistemi *person-based* si caratterizzano per essere orientati alla previsione del rischio individuale talora con finalità di profilazione. Tali sistemi si fondano sull'idea secondo cui l'individuazione preventiva delle persone solitamente coinvolte in attività illecite - sulla base di reti sociali, precedenti relazioni con autori o vittime di reati - rappresenti una strategia promettente per individuare in anticipo potenziali autori e vittime di reato. Una sottocategoria dei *person-based software* è costituita dai sistemi *suspect-based,* nei quali rientra *Keycrime,* il *software* sviluppato dalla Questura di Milano. In questo caso, l'algoritmo si concentra sull'analisi di condotte seriali, elaborando *pattern* riconoscibili nei *modus operandi* degli autori dei reati, rispetto a una serie di fattispecie già realizzate, per anticiparne le mosse future. Tali sistemi si basano su un processo costante di raccolta, catalogazione e confronto dei dati - provenienti dalle dichiarazioni delle vittime, da rilevazioni ambientali o da riscontri oggettivi- che vengono classificati e correlati in funzione delle caratteristiche comportamentali degli autori di reati, al fine di riconoscere e prevenire la reiterazione degli stessi.

§§§

3.3. Le problematiche applicative ed etiche dei sistemi di IA nell'ambito dell'attività di prevenzione dei reati.

Pur a fronte delle indubbie potenzialità che tali strumenti di polizia predittiva sono in grado di offrire in termini di prevenzione, ottimizzazione delle risorse e reattività operativa, in dottrina si evidenziano profili problematici non secondari rispetto alla tutela dei diritti fondamentali.

In particolare, un primo ordine di criticità attiene alla tutela della privacy conseguente alla gran mole di dati che queste applicazioni (fornite, ad esempio, di sensori e telecamere avanzate) possono acquisire in relazione alla vita, anche privata, dei cittadini, e che potrebbero essere manipolati abusivamente, sottratti, deformati, con grave pregiudizio per le persone cui essi si riferiscono. In assenza di adeguate garanzie di proporzionalità e trasparenza tali strumenti rischiano di porsi in potenziale conflitto con l'art. 8 CEDU. In questo senso sia la Corte EDU che la CGUE hanno sottolineato la necessità di introdurre misure tecniche e giuridiche idonee a garantire l'anonimizzazione dei dati, la documentazione degli accessi e il rispetto dei termini di conservazione.

Un secondo profilo di criticità attiene al potenziale effetto discriminatorio. Infatti, la predizione si basa fondamentalmente su una rielaborazione attuariale di diversi tipi di dati: tra i dati inseriti talora compaiono anche informazioni relative all'origine etnica, al livello di scolarizzazione, alle condizioni economiche, alle caratteristiche somatiche riconducibili a soggetti appartenenti a determinate categorie criminologiche (come ad esempio nel caso di potenziali terroristi) che potrebbero portare ad effetti finali di natura discriminatoria o di rafforzamento di pregiudizi di carattere culturale e sociale. Di contro, la deliberata omissione di

alcuni dati, specie in rapporto agli altri, può minare l'effettiva validità predittiva (accuracy) e all'imparzialità (fairness) di questi algoritmi, i quali potrebbero produrre risultati poco affidabili o comunque discriminatori. Sul punto deve essere rilevato che è ormai noto che la struttura di un algoritmo non è mai neutra poiché il programmatore fa delle scelte che, necessariamente, influenzano il risultato dell'operazione computazionale; il programmatore può fare degli errori di progettazione; un algoritmo la cui struttura sia protetta da diritti di proprietà intellettuale e non open source è sottratto alla possibilità di controllo, verifica e confutazione da parte della parte processuale. Infine, un terzo fattore di criticità è rappresentato dalla circolarità autoreferenziale dei dati *(c.d.feedback loop)*. Al riguardo, l'intensificazione dei controlli in una determinata area, a seguito dell'etichettatura come "zona calda", produce inevitabilmente un aumento dei reati rilevati, rafforzando la percezione di pericolosità del contesto e perpetuando la concentrazione dell'attenzione investigativa su di esso, a discapito di altre aree potenzialmente più critiche, ma meno sorvegliate. Inoltre, l'impiego sistematico di tali strumenti rischia di accentuare una logica di tipo militarizzato dell'azione preventiva, concentrata sull'individuazione e sorveglianza di persone o aree sensibili, piuttosto che sull'elaborazione di strategie sociali e strutturali orientate alla rimozione dei fattori criminogeni.

§

3.3.1. Gli interventi correttivi del legislatore europeo in materia di applicazioni predittive.

L'emersione di tali criticità ha indotto il legislatore europeo a ripensare le scelte iniziali introducendo dei divieti originariamente non previsti.

Come visto nel capitolo primo, nella originaria previsione dell'*AI Act,* le applicazioni predittive rientravano, infatti, tra i sistemi di IA ad alto rischio, e dunque erano tutte ammesse purché sottoposti alle misure di contenimento dei rischi prescritte dal Regolamento europeo. Tuttavia, a seguito delle modifiche apportate al documento da Parlamento e Consiglio, nella versione definitiva del Regolamento alcune pratiche di predizione *person-based* sono state qualificate come vietate, in quanto costitutive di un pericolo intollerabile per i diritti fondamentali.

In particolare, l'art.5 par. I, lett. d) del Regolamento europeo vuole evitare che le tecnologie predittive si pongano in contrasto con la presunzione di innocenza e il rispetto dell'antropocentrismo che permea l'intero Regolamento, esplicitamente vietando «l'immissione sul mercato, la messa in servizio per tale finalità specifica o l'uso di un sistema di IA per effettuare valutazioni del rischio relative a persone fisiche al fine di valutare o prevedere il rischio che una persona fisica commetta un reato, unicamente sulla base della profilazione di una persona fisica o della valutazione dei tratti e delle caratteristiche della personalità».

Il divieto, tuttavia, presenta alcune rilevanti eccezioni. Infatti, la seconda parte dell'art.5, par. I, lett.d) in questione stabilisce che il divieto "non si applica ai sistemi di IA utilizzati a sostegno della valutazione umana del coinvolgimento di una persona in un'attività criminosa, che si basa già su fatti oggettivi e verificabili direttamente connessi a un'attività criminosa".

Come condivisibilmente evidenziato in dottrina[82], la logica sottesa a tale scelta legislativa è quella di ammettere pratiche di efficientamento dell'azione preventiva che poggino sulla valutazione umana di dati oggettivi già in possesso dell'autorità perché in questo caso le tecnologie predittive non si pongono in contrasto con la presunzione di innocenza e il rispetto dell'antropocentrismo che, come detto, permea l'intero Regolamento.

[82] F. COPPOLA, *op.cit*. pag. 89 ss.

Se ne ricava che, in ambito europeo, non sono ammesse soluzioni predittive fondate su valutazioni esclusivamente *person based* essendo invece possibile un loro utilizzo a supporto di un fondamento probatorio preesistente di natura oggettiva riguardante il coinvolgimento della persona nell'attività criminosa che si ritiene in esecuzione.

L'Allegato III, par.6 lett.d, considera ad alto rischio "i sistemi di IA destinati ad essere utilizzati dalle autorità di contrasto per loro conto oppure delle istituzioni, organi, organismi dell'unione a sostegno delle autorità di contrasto per determinare il rischio di commissione del reato di recidiva in relazione ad una persona fisica non solo sulla base della profilazione delle persone fisiche di cui all'articolo 3 paragrafo 4 della direttiva 2016/680 o per valutare i tratti e le caratteristiche della personalità del comportamento criminale pregresso di persone fisiche o gruppi". In tale ambito finiscono per essere inevitabilmente inseriti anche i richiamati software di *X-Law* attesi i potenziali effetti discriminatori laddove la scelta del posto in cui rafforzare i controlli di polizia da parte del *software* avvenisse in modo randomico oppure incontrollabile, finendo per pregiudicare determinati gruppi di persone coinvolte.

Infine, l'Allegato III, par. 6, lett. e) considera ad alto rischio "i sistemi di IA destinati a essere utilizzati dalle autorità di contrasto o per loro conto, oppure da istituzioni, organi e organismi dell'Unione a sostegno delle autorità di contrasto, per effettuare la profilazione delle persone fisiche di cui all'articolo 3, paragrafo 4, della direttiva (UE) 2016/680 nel corso dell'indagine dell'accertamento e del perseguimento di reati».

Riferendosi espressamente alla profilazione di persone (almeno) indiziate da reato, vi rientrano i *suspect-based systems*. Analoghe problematiche pongono poi ulteriori sistemi in fase di sviluppo e sperimentazione. Tra questi sono sicuramente di rilievo i sistemi c.d. Virtual HUMINT, sviluppo tecnologico della tradizionale attività di human intelligence (basata sull'interazione

interpersonale svolta da persone fisiche) che offre maggiori garanzie di sicurezza, anonimato e rapidità di accesso. La tecnologia in questione consente di creare avatar per monitorare o avvicinare profili-bersaglio mantenendo la copertura grazie ad algoritmi progettati per imitare il comportamento umano (ed evitare di essere così identificati come robot) e per evitare la tracciabilità dell'operatore e della sua organizzazione di appartenenza.

Ciò significa che anche gli strumenti di *Keycrime* nonché quelli ad essi assimilabili possono essere utilizzati dalle forze di polizia a condizione che l'attività di predizione avvenga sulla base di dati di natura oggettiva (come precedenti attività delittuose e/o indagini in corso) e purché detti sistemi di I.A. siano utilizzati nel rispetto degli obblighi di supervisione, trasparenza e controllo umano delineati dal Regolamento.

§§§§§

4. Conclusioni

Nel primo capitolo abbiamo soffermato l'attenzione su come l'IA possa costituire una opportunità in termini di progresso economico e sociale per l'intera umanità, ma allo stesso tempo un fattore di rischio.

Nel capitolo in questione, specificamente dedicato alla materia del terrorismo, sono emerse le medesime problematiche perché da un lato i sistemi di IA possono diventare una opportunità costituendo dei formidabili strumenti per contrastare il terrorismo; dall'altro le nuove forme di IA in mano alle organizzazioni terroristiche possono diventare un pericolosissimo strumento per commettere reati di terrorismo.

Sotto il profilo delle opportunità i sistemi di IA, nel doveroso rispetto dei diritti fondamentali dell'individuo e delle garanzie del diritto di difesa, assumono fondamentale importanza nell'ambito dell'attività c.d. di polizia predittiva grazie alla presenza di tecnologie che consentono, ad esempio, di identificare un soggetto a partire da un fotogramma, confrontando quest'ultimo con banche dati che contengono dati biometrici e fotografie o con le immagini delle telecamere di sorveglianza di una determinata zona alla possibilità di utilizzare tecniche di riconoscimento facciale. Inoltre, l'utilizzo massiccio di IA per scopi di *law enforcement* oramai rappresenta una realtà in diversi Stati (soprattutto negli USA).

Tuttavia, come accennato, l'IA può diventare un fattore di rischio contribuendo, all'uso distorto che di essa possono farne le stesse organizzazioni criminali.

Dal tragico 11 settembre del 2001, che con l'attentato alle torri gemelle da parte dell'organizzazione terroristica al Qaida sconvolse l'intera umanità, anche le modalità di aggressione del terrorismo internazionale si sono, infatti, evolute.

L'avvento della tecnologia e, più in generale, della transizione digitale ha spostato l'attenzione delle organizzazioni criminali sui nuovi sistemi di IA.

In questo modo tali sistemi diventano degli altrettanto formidabili strumenti per commettere reati prevalentemente in tre distinte macroaree:

- *cyber threats*;
- *psysical threats*;
- *political threats.*

Ognuna di queste macroaree presenta delle specificità dal punto di vista delle interferenze con il diritto penale, andando ad incidere su istituti che riguardano la teoria generale del reato.

Nel caso dei *political threats* l'utilizzo dei deepfake per amplificare i discorsi di odio a fini di propaganda terroristica impone di trovare un valido punto di equilibrio tra la libertà di

espressione, che può coinvolgere anche la materia religiosa, ed i limiti impliciti all'esercizio dei diritti fondamentali della persona. Con riferimento ai *cyber threats* ed ai *psysical threats*, si pongono diverse problematiche che interessano il diritto penale. Infatti, in questi casi l'autore fisico del reato sempre di più assume una posizione marginale nella realizzazione del fatto, mentre la sfida attuale, ancora solo in parte esplorata dalla dottrina, ruota intorno alla possibilità di ravvisare profili di responsabilità direttamente in capo alla macchina (*machina delinquere non potest*).

In conclusione, la spasmodica attenzione delle organizzazioni criminali per tali sistemi intelligenza artificiale impone la necessità di spostare l'attenzione sulle modalità con cui a livello di regolazione normativa le varie potenze mondiali stanno affrontando il tema dello sviluppo dei sistemi di intelligenza artificiale e fino a che punto la stessa possa diventare un fattore di rischio per l'intera umanità.

Situazione quella da ultimo descritta che impone una domanda: fino a che punto può spingersi la tecnologia? Esiste un limite? Ed i nostri decisori politici cosa possono fare? Si tratta di quesiti che meglio verranno affrontati nel capitolo successivo, allorquando verranno messi a confronto tre diversi approcci alla disciplina dei sistemi di IA: quello regolatorio europeo; quello liberista, autoregolatorio americano; quello dirigista cinese.

CAPITOLO TERZO: MODELLI DI REGOLAZIONE DELL' INTELLIGENZA ARTIFICIALE: IMPLICAZIONI PENALI E PRIME DECISIONI GIUDIZIARIE SULL'IA GENERATIVA

1.L'IA tra modelli regolatori, auto-regolatori e dirigisti: prospettive a confronto

L'attuale disciplina dell'IA è governata da due diverse tipologie di approcci normativi:

- uno di carattere generale, volto a predisporre un quadro normativo unitario e trasversale della materia;
- l'altro di carattere settoriale, volto a predisporre un quadro normativo su un singolo aspetto della materia quello dell'IA generativa[83] e dei potenziali effetti distorsivi legati alla stessa.

Nel primo ambito si inscrive l'Unione europea che, attraverso il citato regolamento AI ACT, ha voluto introdurre un quadro regolatorio generale al fine di garantire il buon funzionamento del mercato interno uniformando lo sviluppo, l'immissione sul mercato, la messa in servizio e l'utilizzo di sistemi di IA sul presupposto della prevenzione dei rischi sistemici e della tutela dei diritti fondamentali[84].

[83] Sulla IA forte e la c.d. strong AI cfr. *supra,* capitolo primo, par.1.2.1.

[84] Sulla portata ed i limiti del regolamento AI ACT vi è ampia letteratura in dottrina. Cfr sul punto: Maria Vittoria CATANZARITI *L'AI Act alla prova delle sfide globali: potenzialità e limiti di un modello regolatorio*, in Giustizia Insieme maggio 2025; G. LAZCOZ MORATINOS, *Human oversight (article 14), in The EU regulation on artificial intelligence: a commentary* a cura di A. Huergo Lora, Milano, 2025, 243 e ss..; M. INGLESE, *Il regolamento sull'intelligenza artificiale come atto per il*

In dottrina dubbi e perplessità sono stati manifestati in merito all'appropriatezza della scelta del legislatore unionale di regolare la materia in questione attraverso lo strumento normativo del regolamento anziché attraverso quello della direttiva. Invero l'utilizzo del regolamento - che costituisce la base giuridica più ampia tra quelle disponibili nei Trattati - appare confliggere con il disposto dell'art. 114 T.F.U.E. che su una materia di competenza concorrente, quale deve essere considerato il mercato interno ai sensi dell'art. 4, par. 2, lett a) T.F.U.E., e nella prospettiva della protezione dei diritti fondamentali, che non costituirebbe da sé una base giuridica autonoma, predilige l'impiego della direttiva, se del caso dettagliata[85].

A ciò si aggiunga che il tenore del regolamento non è eccessivamente prescrittivo e svariati spazi d'azione sono lasciati agli Stati membri, con ciò potendosi affermare che il grado di armonizzazione, se non propriamente minimo, non sia in senso stretto nemmeno massimo.

Il descritto modello regolatorio europeo si differenzia sia da quello USA, caratterizzato dall'autoregolazione in quanto improntato al pragmatismo e al principio di non ostacolare l'innovazione, che da quello cinese, settoriale e di più marcata impronta dirigista. Sul punto, in dottrina è stato correttamente sottolineato come il regolamento AI ACT abbia trasformato l'Unione europea in un "gigante" dal punto di vista della regolazione relegandolo, tuttavia, in una posizione di secondo piano dal punto di vista della tecnologia al cospetto di Stati Uniti

completamento e il buon funzionamento del mercato interno? in Rivista Quaderni AISDUE - *L'Unione europea e la nuova disciplina sull'intelligenza artificiale: questioni e prospettive*, a cura di F. FERRI, Napoli, 2024, 71 e ss.; B. CAPPIELLO, *The EU and the AI ACT. Was it worthwhile to be the first?*, in CERIDAP, 4, 2024, 235 e ss., ;C. IURILLI, *Il diritto naturale come limite e contenuto dell'intelligenza artificiale. Prime riflessioni sul nuovo Regolamento Europeo "AI Act"*, in Judicium, 24 giugno 2024.

[85] *L'Unione europea e la nuova disciplina sull'intelligenza artificiale: questioni e prospettive*, a cura di F. FERRI, Napoli, 2024

e Cina, che proprio per le difficoltà di inquadrare una materia in continua evoluzione rifuggono da tale approccio regolatorio [86].
Tali diverse impostazioni normative si riflettono anche in ambito giudiziario dividendosi tra impostazioni più libertarie ed altre decisamente proibizioniste. Pur nella richiamata diversità, gli orientamenti giurisprudenziali che via via si stanno sedimentando appaiono convergere su alcuni punti irrinunciabili: trasparenza obbligatoria, consenso per l'uso dell'immagine altrui, sanzioni severe per chi causa danni reali.

§§§

1.1. Diversità di approcci regolatori tra UE, USA e CINA nell'era dell'intelligenza artificiale

I diversi approcci normativi di UE e USA nella materia in questione rischiano di essere ancor di più accentuati a seguito del mutato quadro politico che, con l'attuale amministrazione di Donald Trump, sin dall'inizio si è mostrato piuttosto refrattario agli eccessi regolatori europei.
Infatti, inizialmente l'executive order n.14110 del 30 ottobre 2023 - emanato dal Presidente Joe Biden per sollecitare il Congresso a dare una risposta rapida all'assenza di una legge federale condivisa - aveva fatto registrare significative convergenze tra l'amministrazione americana e quella UE al fine di elaborare una strategia tendenzialmente unitaria in materia di IA.
In particolare, il citato provvedimento n.14110 dal titolo "Safe, Secure, and Trustworthy Development and Use of Artificial

[86] L. TORCHIA, *Pubblica amministrazione e transizione digitale*, in Giorn. dir. amm., 6, 2024, 729).

Intelligence" richiamando le "linee guida etiche" si poneva in continuità con le azioni iniziate in seno alle istituzioni europee indicando come strada maestra la necessità di una cooperazione con gli alleati e i partner internazionali al fine di gestire i rischi dell'IA, sbloccare il potenziale dell'IA e promuovere approcci comuni a sfide condivise. Nelle intenzioni di Joe Biden la cooperazione doveva riguardare non solo le potenze amiche ma anche le nazioni nemiche al fine di guidare conversazioni e collaborazioni chiave a livello globale per assicurarsi che l'IA potesse beneficiare il mondo intero, invece di esacerbare le iniquità, minacciare i diritti umani e causare ulteriore danno.

Tale scenario sembra ora lasciare spazio ad una inversione di tendenza conseguente all'emanazione in data 23 gennaio 2025 dall'executive order emanato da Donald Trump dal titolo "Removing Barriers to American Leadership in Artificial Intelligence" che ha rimosso quello sin qui descritto della precedente amministrazione guidata da Joe Biden.

Sul punto, l'articolo 2 del citato executive order del 23 gennaio 2025 ha affermato che la nuova linea politica è quella di sostenere e implementare la posizione dominante dell'America a livello globale per promuovere la crescita umana, la competitività economica e la sicurezza nazionale.

Tale rinnovata impostazione politica nasce anche dalla necessità di sollecitare un ripensamento dell'intero sistema normativo euro unitario - già oggetto di contestazioni ben prima dell'insediamento di Trump a causa della regolamentazione estremamente prolissa, ripetitiva e composita[87]- e più in generale di ridefinire i rapporti con l'Europa in un settore nel quale gli Usa vogliono rivendicare l'assoluta leadership.

A proposito di rivendicate leadership, deve essere evidenziato che anche in Cina non esiste un intervento di carattere regolatorio organico e gli interventi normativi assumono carattere settoriale

[87] A. SANTOSUOSSO, B. MARONE, *Regole per l'IA: cosa può imparare l'Italia dalle strategie Usa e UK*, in Agenda Digitale, 21 novembre 2023.

comprendendo ambiti corrispondenti a quelli dell'AI ACT con particolare riferimento alla disciplina sui deep synthesis e alle misure ad interim per la gestione dei servizi di intelligenza artificiale generativa. Tali disposizioni settoriali si ispirano ad una impronta decisamente dirigista soprattutto nell'ambito delle censure preventive in merito ai contenuti diffusi sulle piattaforme digitali.

Sullo sfondo del descritto scenario sociopolitico che caratterizza i modelli regolatori in questione, si innesta l'evoluzione tecnologica del terrorismo ed il sempre più frequente utilizzo dei sistemi di IA da parte delle stesse organizzazioni criminali, il cui operare può essere favorito dall'assenza di una risposta globale unitaria.

Il quadro è, peraltro, in continua evoluzione e si lega al complessivo scenario geopolitico che caratterizza la materia in questione[88].

§§§

1.2. Geopolitica e sviluppo industriale dell'IA.

Lo sviluppo industriale dell'intelligenza artificiale (IA) ha una data simbolica di partenza che può essere fissata al 2006, anno in cui il matematico neozelandese Shane Legg[89] Legg pubblicò un contributo accademico di grande rilievo, presentato alla 17ª Conferenza Internazionale sull'Apprendimento Automatico (ALT 2006).

In quel lavoro, Legg rilanciava il dibattito sulla predizione universale in contesti di machine learning e IA, gettando le basi

[88] Su questa dimensione di confronto tra "potenze" si veda il volume di Limes, *L'intelligenza non è artificiale*, 12, 2022

[89] S. LEGG *Is There an Elegant Universal Theory of Prediction?* Proceedings of the 17th International Conference on Algorithmic Learning Theory (ALT 2006) Serie: *Lecture Notes in Computer Science* (LNCS), vol. 4264 Editore: Springer, Berlin, Heidelberg Pagine: 274–287

teoriche per riflessioni che nei decenni successivi avrebbero trovato concretezza industriale.

È tuttavia dal 2010 che si può parlare del "decennio di svolta" per l'IA, quando nasce DeepMind, il primo grande laboratorio europeo dedicato in maniera esclusiva allo studio e allo sviluppo dell'intelligenza artificiale. Tra i soci fondatori figurano lo stesso Shane Legg e Demis Hassabis, ex campione di scacchi e scienziato cognitivo. I due avevano l'ambizione di costruire un polo europeo indipendente, ma ben presto si resero conto che le opportunità di finanziamento si concentravano prevalentemente nella Silicon Valley. Fu proprio lì che riuscirono a convincere alcuni investitori visionari, tra cui Peter Thiel, attratto dalle potenzialità espresse da Hassabis attraverso le sue analogie con il gioco degli scacchi, tradizionalmente considerato un banco di prova per l'intelligenza artificiale.

Nel 2014, l'attenzione globale verso l'IA si consolidò con la mossa decisiva di Google, che decise di investire massicciamente nel settore acquisendo DeepMind. Questo passaggio sancì la fine del sogno di un'IA puramente europea e spostò il baricentro dell'innovazione tecnologica oltreoceano.

L'anno successivo, nel 2015, arrivò una contromossa importante: Elon Musk, consapevole dei rischi e delle sfide etiche poste dall'IA, promosse la nascita di OpenAI, coinvolgendo figure chiave come Sam Altman, Greg Brockman e Ilya Sutskever. L'obiettivo era ambizioso: creare un'organizzazione no profit capace di sviluppare intelligenza artificiale in modo sicuro, trasparente e a beneficio dell'intera umanità.

Parallelamente, la Cina entrava in scena in modo deciso: sempre nel 2015 annunciava il piano industriale **"**Made in China 2025**"**, volto a trasformare il Paese nel leader mondiale delle tecnologie avanzate, con una forte attenzione a semiconduttori, robotica e IA.

Gli equilibri geopolitici si fecero più tesi: nel 2016 l'amministrazione Obama emanò un ordine esecutivo volto a

contrastare la cessione di tecnologie strategiche, aprendo quella che sarebbe diventata la guerra dei semiconduttori tra Stati Uniti e Cina. Lo scontro si intensificò nel 2018, quando a Vancouver venne arrestata Meng Wanzhou, direttrice finanziaria e figura di spicco di Huawei, accusata di violazioni alle sanzioni statunitensi. Sempre nel 2018, Ilya Sutskever fornì una delle prime definizioni operative di intelligenza artificiale generale (AGI), descrivendola come un sistema autonomo in grado di superare le prestazioni umane nella maggior parte delle attività di valore economico. A fine anno, Elon Musk abbandonò OpenAI: i suoi interessi con Tesla e SpaceX, sempre più intrecciati con applicazioni avanzate di IA, erano ormai incompatibili con la missione originaria dell'organizzazione.

Il 2019 segnò un punto di svolta: OpenAI strinse un accordo strategico con Microsoft, trasformandosi da organizzazione no profit in una società a scopo di lucro limitato. Questa partnership accelerò enormemente lo sviluppo di modelli linguistici avanzati, aprendo la strada al lancio, nel 2022, di ChatGPT, che con la sua diffusione planetaria consolidò la collaborazione tra OpenAI e Microsoft, cambiando radicalmente la percezione pubblica dell'IA.

Il fermento però non si fermò lì. Nel 2025, in un'inaspettata mossa industriale, Meta annunciò l'acquisizione di Google con l'obiettivo di rafforzare la propria infrastruttura di calcolo e ampliare la capacità di gestione dati su scala globale. Questo evento sancì un nuovo capitolo nella competizione tra i colossi tecnologici.

Secondo Shane Legg, l'industria dell'IA viaggia a un ritmo molto più rapido di quanto le generazioni precedenti avrebbero immaginato. Egli stesso ha previsto che il 2028 sarà l'anno della vera affermazione dell'intelligenza artificiale generale, momento in cui i sistemi non solo eguaglieranno, ma supereranno le capacità cognitive e produttive dell'essere umano in maniera strutturale. Una prospettiva che, tra entusiasmi e timori,

rappresenta forse il traguardo più ambito e controverso della nostra epoca tecnologica.

§§§

1.3. Immediatezza e gradualità operativa delle norme del regolamento AI ACT tra buon funzionamento del mercato interno e tutela dei diritti fondamentali.

Come anticipato nel capitolo primo, per far fronte alla irresistibile ascesa dei sistemi di IA, l'Unione Europea ha adottato un intervento regolatorio di carattere generale: il Regolamento EU 2024/1689 del Parlamento europeo e del Consiglio europeo del 13 giugno 2024, noto come AI ACT, pubblicato sulla Gazzetta Ufficiale dell'Unione Europea Serie L del 12 luglio 2024. Il regolamento è entrato in vigore in alcune sue parti il 2 agosto del 2025 e mira a stabilire regole armonizzate sull'intelligenza artificiale modifica numerosi regolamenti e direttive europee già esistenti [90].

Nelle considerazioni introduttive il legislatore europeo chiarisce innanzitutto che la finalità primaria è quella di garantire il buon funzionamento del mercato interno mediante un quadro giuridico uniforme in materia di sviluppo, immissione sul mercato, messa in servizio e utilizzo di sistemi di IA. In tale prospettiva viene assicurata la libera circolazione transfrontaliera di beni e servizi basati sull'IA, limitando la possibilità per gli Stati membri di

[90] Il riferimento è ai regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e alle direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828.

introdurre restrizioni allo sviluppo, salvo quanto previsto dallo stesso regolamento.

Tuttavia, oltre a questo riferimento obbligato al mercato interno, emerge subito quello che rappresenta il vero obiettivo politico e valoriale dell'AI ACT: promuovere la diffusione di una IA antropocentrica e affidabile, capace di garantire un livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea. Il regolamento si propone quindi di prevenire gli effetti nocivi dei sistemi di IA, tutelare democrazia, Stato di diritto e ambiente, e al contempo stimolare l'innovazione tecnologica.

Con particolare riferimento alla tematica dei diritti fondamentali deve evidenziarsi, che nelle considerazioni introduttive i diritti fondamentali vengono indicati come interessi pubblici da proteggere assieme alla salute e alla sicurezza, all'interno del più ampio e generale contesto della tutela della sicurezza dei prodotti. Tale affermazione di principio discende dalla circostanza che, in generale, il legislatore europeo nell'esercizio delle competenze che gli sono conferite dai Trattati è tenuto al rispetto dei diritti fondamentali.

Sul punto in dottrina è stato evidenziato come la caratteristica che contraddistingue il regolamento in questione rispetto ad altri interventi normativi eurounitari è data dal fatto che la tutela dei diritti fondamentali costituisce, assieme alla previsione di standard tecnici, lo scheletro dell'intero impianto normativo e non, invece, uno dei vari requisiti da rispettare. La conseguenza di tale impostazione è che il rischio per i diritti fondamentali diventa il principale criterio per la previsione di obblighi più stringenti tanto che, come detto in precedenza, l'impatto negativo di un sistema di intelligenza artificiale sui diritti fondamentali garantiti dalla Carta costituisce un co-criterio per la classificazione di una applicazione come di rischio elevato ai sensi del citato art. 6, par. 3 del regolamento in questione.

Allo stesso modo l'impatto sui diritti fondamentali è, peraltro, criterio da seguire nella valutazione preventiva all'impiego dei sistemi di IA ad alto rischio cui sono chiamati i deployer, per tali intendendosi qualsiasi persona fisica o giuridica - compresi un'autorità pubblica, un'agenzia o altro organismo - che utilizza un sistema di IA sotto la sua autorità, salvo nel caso in cui il sistema di IA stesso sia utilizzato nel corso di un'attività personale non professionale.

Al riguardo l'art. 27 del Regolamento[91] stabilisce che prima di utilizzare un sistema di IA ad alto rischio di cui all'articolo 6, paragrafo 2, ad eccezione dei sistemi di IA ad alto rischio destinati a essere usati nei settore elencati nell'allegato III, punto 2 comprensivi dell'istruzione e della formazione professionale, i deployer che rivestono la qualifica di organismi di diritto pubblico o che si configurano come enti privati che forniscono servizi pubblici e i deployer di sistemi di IA ad alto rischio di cui all'allegato III, punto 5, lettere b) e c) come tali comprensivi della profilazione predittiva, hanno l'obbligo di effettuare una valutazione dell'impatto sui diritti fondamentali che l'uso di tale sistema può produrre. In particolare, tale valutazione deve comprendere i seguenti elementi:

- una descrizione dei processi del deployer in cui il sistema di IA ad alto rischio sarà utilizzato in linea con la sua finalità prevista;
- una descrizione del periodo di tempo entro il quale ciascun sistema di IA ad alto rischio è destinato a essere utilizzato e con che frequenza;
- le categorie di persone fisiche e gruppi verosimilmente interessati dal suo uso nel contesto specifico;

[91] E. CIRONE, *L'AI Act e l'obiettivo (mancato?) di promuovere uno standard globale per la tutela dei diritti fondamentali*, in Quaderni AISDUE, 1, 2025, 12 e ss.; M. INGLESE, *op. cit.* in Rivista Quaderni AISDUE -

- i rischi specifici di danno che possono incidere sulle categorie di persone fisiche o sui gruppi di persone sopraindicate tenendo conto delle informazioni trasmesse dal fornitore a norma dell'articolo 13;
- una descrizione dell'attuazione delle misure di sorveglianza umana, secondo le istruzioni per l'uso;
- le misure da adottare qualora tali rischi si concretizzino, comprese le disposizioni relative alla governance interna e ai meccanismi di reclamo.

Il descritto obbligo di valutazione di impatto si applica al primo uso del sistema di IA ad alto rischio. Il deployer può, in casi analoghi, basarsi su valutazioni d'impatto sui diritti fondamentali effettuate in precedenza o su valutazioni d'impatto esistenti effettuate da un fornitore. Se, durante l'uso del sistema di IA ad alto rischio, ritiene che uno qualsiasi degli elementi sopra elencati sia cambiato o non sia più aggiornato, il deployer adotta le misure necessarie per aggiornare le informazioni.

Una volta effettuata la valutazione di impatto, il deployer deve notificare all'autorità di vigilanza del mercato i suoi risultati, presentando l'apposito modello compilato. A tal fine per agevolare i deployer nell'adempimento dei loro obblighi l'apposito ufficio per l'IA elabora un modello di questionario, anche attraverso uno strumento automatizzato.

La normativa delineata nell'art.27 in questione deve coordinarsi con quella contenuta nell'art. 46 del regolamento AI ACT, relativa ai sistemi ad alto rischio per i quali sono previsti specifici sistemi di notifica, nonché con quella contenuta nell'art.35 del regolamento UE 2016/679 sul trattamento dei dati personali. A quest'ultimo riguardo se uno qualsiasi degli obblighi sopraindicati è già stato adempiuto mediante la valutazione d'impatto sulla protezione dei dati effettuata a norma del citato articolo 35, la valutazione d'impatto sui diritti fondamentali in caso di utilizzo dei sistemi di IA integra la valutazione d'impatto sulla protezione dei dati.

Una volta evidenziato lo scheletro normativo dell'AI ACT imperniato sul buon funzionamento del mercato interno e sulla tutela dei diritti fondamentali, occorre osservare il regolamento in questione si compone di 113 articoli, suddivisi in XIII capi. L'art. 113, quale disposizione di chiusura, fissa come regola generale la data del 2 agosto 2026 per l'applicazione del regolamento.

A tale previsione, tuttavia, si accompagnano alcune deroghe, poiché il legislatore ha previsto l'entrata in vigore anticipata di specifiche disposizioni, che si applicano già a partire da date precedenti ed in particolare:

- le norme dei capi I e II si applicano a decorrere dal 2 febbraio 2025. In particolare, il capo I contiene negli artt. da 1 a 4 le disposizioni di carattere generale. Il capo II nel cui ambito trova collocazione l'art.5 introduce la disciplina sulle pratiche di IA vietate e come tali considerate inaccettabili;
- le norme del capo III, sezione 4, il capo V, il capo VII, il capo XII e l'articolo 78 si applicano a decorrere dal 2 agosto 2025, ad eccezione dell'articolo 101. Al riguardo le norme del capo III, sezione 4 riguardano gli artt. da 28 a 39 relativamente agli organismi di notifica; le norme del capo V riguardano gli artt. da 51 a 56 e dettano la disciplina per i sistemi di IA per finalità generali; le norme del capo VII riguardano gli artt. da 64 a 71 relativamente alla governance della IA; il capo XII comprende gli artt. da 99 a 101 relativamente agli aspetti sanzionatori. L'art.101che prevede le sanzioni per i fornitori di modelli di IA per finalità generali entra in vigore il 2 agosto 2026;
- l'articolo 6, paragrafo 1, e i corrispondenti obblighi di cui al presente regolamento si applicano a decorrere dal 2 agosto 2027".

La ragione di questa distinzione temporale di entrata in vigore delle norme è quella di consentire alle imprese di intraprendere i percorsi organizzativi interni per una effettiva adesione alle

norme al fine di garantire il buon funzionamento del mercato interno.

§

1.3.1. Il codice di condotta UE del 10 luglio 2025 e la cooperazione con le imprese del settore.

Nella descritta logica di aiutare le imprese a adeguarsi al reticolato normativo appena descritto, il 10 luglio 2025, la Commissione europea ha approvato un codice di condotta volontario finalizzato a ridurre gli oneri amministrativi e contestualmente fornire una maggiore certezza giuridica.
Progettato congiuntamente con gli attori del settore, il codice di condotta in questione è stato approvato proprio quando un gruppo nutrito di multinazionali europee del settore ha manifestato forti criticità nei confronti della legislazione comunitaria, perché ritenuta troppo vincolante rispetto ai differenti sistemi normativi previsti nel mondo. In particolare, un gruppo di 46 imprese tra cui Hair Bass, Lufthansa, BB Paribas e Mistral ha chiesto la sospensione temporanea delle nuove regole. Queste aziende accusano le norme comunitarie di mettere a repentaglio l'ambizione europee in materia di intelligenza artificiale perché a loro dire compromettono non solo lo sviluppo di campioni europei, ma anche la capacità di tutti i settori di utilizzare l'intelligenza artificiale sulla scala richiesta dalla concorrenza globale. In risposta alle critiche, la commissione europea ha ricordato di essere al lavoro su un progetto di semplificazione della legislazione relativa al mondo digitale ribadendo che in ogni caso, indagini e sanzioni sulle imprese non potranno partire prima del 2 agosto 2026 secondo il disposto dell'art.113 del regolamento AI ACT.
In tale contesto si colloca la recente adozione del codice di condotta. Tra i principali destinatari del codice in questione ci

sono le Big Tech. Tra loro, solo Google e OpenAI stanno revisionando il testo in questione, senza aver espresso al momento nessun commento. Microsoft, Amazon e la più importante azienda di AI europea - Mistral - non hanno ancora reagito. Così come Meta, che è in passato è stata critica sull'AI Act.

Il codice è stato realizzato da 13 esperti indipendenti dopo aver sentito oltre 1000 attori del settore.

In particolare, il documento di quasi 60 pagine affronta tre distinti temi: quello della trasparenza e del diritto di autore; quello della sicurezza; quello della protezione.

Con riferimento alla trasparenza e al diritto di autore viene prevista la fornitura di un modulo attraverso cui le aziende, nello specifico i provider di IA generativa, tra cui rientrano ChatGpt, Gemini, Copilot, Meta AI, possono condividere le informazioni sul funzionamento dei loro modelli. Sul punto le previsioni del codice sono stringenti: infatti dovranno essere indicare non solo le modalità con cui i modelli sono addestrati, ma anche i loro utilizzi, i limiti, le risorse computazionali utilizzate e il consumo energetico. Ogni cambiamento e aggiornamento dovrà essere segnalato. Relativamente al copyright e all'origine dei dati che utilizzati dai modelli, i provider devono esplicitare la provenienza per conformarsi alla legge sul diritto d'autore. Si prevede anche la creazione di strumenti attraverso cui i detentori di proprietà intellettuale possono fare ricorso per l'utilizzo illecito delle loro opere per l'addestramento e chiedere di rimuoverle dai database. In particolare, il codice in questione raccomanda di escludere dei modelli di ricerca i siti noti per ripetuti attacchi di pirateria informatica. I sistemi in questione dovranno impegnarsi inoltre a verificare che le loro conversazioni non contengano linguaggio offensivo o violento.

Infine, con riferimento al distinto tema della sicurezza le società si devono impegnare a condurre test periodici per identificare e mitigare i "rischi sistemici" dei modelli, la cui disciplina è contenuta negli artt.51 e ss del regolamento AI ACT, quali ad

esempio possono essere lo sfruttamento dell'IA per creare armi chimiche e biologiche o per organizzare cyber attacchi, la manipolazione del comportamento umano e dunque le capacità di persuasione e influenza sull'utente.
Infine, relativamente al tema della protezione, si richiede di creare un processo definito da adottare per evitare che l'IA possa essere sfruttata per fare danni alla sicurezza pubblica. Più complesso e fumoso è il tema della mitigazione dei rischi di diffusione di disinformazione e contenuti dannosi.
In conclusione, l'obiettivo della legislazione è di garantire che i modelli generici di intelligenza artificiale sul mercato europeo, compresi quelli più potenti, siano sicuri e trasparenti.

§§§

1.4.Gli interventi regolatori di carattere generale in Italia: il ddl 1146-B.

In ambito interno, il legislatore si propone di disciplinare la materia attraverso il citato intervento di carattere generale rappresentato dal DDL 1146-B sull'IA[92].
Al riguardo i primi commentatori della disciplina in materia hanno avanzato più di un dubbio sulla compatibilità di tale intervento normativo con quello adottato in ambito europeo che, attraverso lo strumento del regolamento, idoneo a disciplinare direttamente e immediatamente tutti i rapporti giuridici ad esso sottoposti, pare restringere siffatta possibilità.
Sul punto in dottrina è stato evidenziato come l'AI Act, appaia lasciare, su più di un fronte, un indiscusso potere discrezionale a favore degli Stati membri, che, da questo punto di vista, hanno

[92] C.BURELLI, *Prime brevi considerazioni sul "ddl intelligenza artificiale": incompatibilità o inopportunità?* in Rivista Quaderni AISDUE - *L'Unione europea e la nuova disciplina sull'intelligenza artificiale: questioni e prospettive*, a cura di F. FERRI, Napoli, 2024

dinnanzi a sé un atto che, per certi versi, somiglia più a una direttiva che non a un regolamento in senso stretto, per sua natura come detto direttamente e immediatamente applicabile a tutti i rapporti giudici ad esso sottoposti. In particolare, tale orientamento dottrinale ha inteso rimarcare come le norme del regolamento AI ACT lungi dall'essere autosufficienti si avvicinano, talvolta, a dei programmi di legislazione[93].

Tali considerazioni appaiono condivisibili allo scrivente in quanto confortate dal fatto che la Commissione europea sempre di più si affida ad atti delegati e di esecuzione nonché all'ulteriore intervento legislativo o amministrativo da parte degli Stati membri.

A ribadire la compatibilità tra la normativa interna e quella dettata in ambito europeo soccorre, altresì, il disposto dell'art. 1, comma 2, secondo cui "le disposizioni della presente legge si interpretano e si applicano conformemente al diritto dell'Unione europea" che in qualche modo ripropone l'idea di un percorso di armonizzazione che tuttavia meglio si addice ai rapporti tra normativa interna e direttiva.

Fugati i dubbi sulla compatibilità della normativa interna con quella unionale, occorre osserva come il DDL 1146-B–risulta essere composto da 27 articoli suddivisi in 6 capi, all'interno dei quali vengono affrontati una serie di temi cruciali per la regolamentazione dell'IA in Italia. Tra questi, si trovano i principi e le finalità dell'IA, le disposizioni di settore, la strategia nazionale, la tutela degli utenti, il diritto d'autore, le disposizioni penali e quelle finanziarie.

Il DDL è stato approvato con modifiche alle Camera dei deputati il 25 giugno del 2025 e attende ora la discussione alla Camera dei deputati.

Un concetto chiave del disegno di legge è l'autonomia: l'IA è vista come uno strumento che coadiuva le decisioni umane senza sostituirle, promuovendo lo sviluppo di sistemi comprensibili e

[93] C.BURELLI, *op.cit.* in Rivista Quaderni AISDUE

tecnologicamente avanzati. Questo approccio antropocentrico vuole garantire che le decisioni automatizzate siano sempre controllate dall'autodeterminazione umana.

Gli articoli 3, 4 e 5 stabiliscono le prescrizioni etiche e operative per l'IA in Italia, concentrandosi su dignità umana, sicurezza e trasparenza, e promuovendo lo sviluppo economico.

Il Capo II del DDL riguarda l'uso dell'IA in sanità, lavoro e giustizia, migliorando efficienza e la trasparenza.

L'articolo 7 disciplina l'uso dell'IA nel settore sanitario, mentre l'articolo 10 regola l'uso dell'IA nel settore lavorativo.

Nell'ambito della difesa e della sicurezza nazionale il DDL prevede l'uso dell'IA per monitorare minacce, proteggere dati e gestire emergenze informatiche includendo strumenti per il disaster recovery e il miglioramento della cybersicurezza. L'articolo 6 esclude le attività di IA legate alla sicurezza nazionale dalla normativa generale. I sistemi di IA destinati all'uso pubblico devono essere installati su server ubicati in Italia per garantire la sicurezza dei dati sensibili.

Per quanto riguarda la governance dell'IA, il ddl prevede l'assegnazione di tali compiti ad AgID e ACN, una decisione contestata da alcune associazioni per i diritti digitali che preferivano un'autorità indipendente. Il Garante della Privacy ha evidenziato la mancanza di un ente autorizzato per i sistemi di identificazione biometrica in tempo reale e si è candidato per questo ruolo. Le opposizioni propongono la creazione di vari osservatori e commissioni, tra cui un Osservatorio sui Diritti Digitali a Palazzo Chigi, una commissione per l'uso dell'IA in ambito giudiziario, una commissione dati, analisi e la ricerca clinica presso il Ministero della Salute.

L'articolo 21 delinea gli investimenti nei settori dell'IA, cybersicurezza e calcolo quantistico. Il governo ha previsto un fondo da 1 miliardo di euro, gestito da Cdp Venture Capital Sgr, per sostenere lo sviluppo dell'IA in Italia. Questo fondo è destinato sia alle PMI che alle grandi aziende per favorire ricerca

e innovazione. Tuttavia, l'apertura del fondo a investitori stranieri ha suscitato dibattiti sulla tutela dell'industria nazionale e il controllo strategico delle tecnologie emergenti.
Con riferimento al sistema sanzionatorio, l'articolo 25 del DDL apporta modifiche al Codice penale, inasprendo le pene per reati commessi mediante l'uso dell'IA. Le aggravanti sono previste quando l'IA costituisce un mezzo insidioso, ostacola la difesa pubblica e privata, o peggiora le conseguenze del reato. Viene, altresì, introdotto il reato di "Illecita diffusione di contenuti generati o alterati con sistemi di IA" (articolo 612-quater), sul quale ci soffermeremo nel prosieguo a proposito dei deep-fake.

§§§

1.5. Le interconnessioni in ambito europeo tra normativa AI ACT, GDPR e DSA.

Compiuto l'excursus sul piano normativo interno, va rimarcato come, a livello europeo, l'AI Act non possa essere considerato un corpus autonomo, ma debba essere letto in connessione con la disciplina in materia di protezione dei dati personali e con quella relativa alle piattaforme digitali, a conferma della crescente integrazione tra i diversi strumenti di regolazione dell'economia digitale.
Invero una delle peculiarità dei sistemi di IA è quella di "nutrirsi" di dati personali con riferimento ai quali, l'articolo 8, paragrafo 1, della Carta dei diritti fondamentali dell'Unione europea (di seguito "la Carta") e dell'articolo 16, paragrafo 1, del trattato sul funzionamento dell'Unione europea (TFUE), stabilisce che ogni persona ha diritto alla protezione dei dati di carattere personale configurando in tal modo il trattamento dei dati personali come un diritto fondamentale dell'uomo. Ulteriormente disciplinando

la materia, il 27 aprile 2016 il Parlamento europeo ed il Consiglio hanno approvato il Regolamento generale sulla protezione dei dati (UE) 2016/679 meglio noto come GDPR, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali e alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE. A completare il pacchetto in materia di protezione di dati personali devono essere menzionati anche il regolamento EU 2022/868 del 30 maggio 2022 - Digital governance act (DGA)- che riguarda la promozione dell'utilizzo dei dati nel settore pubblico, l'interoperabilità dei dati tra enti pubblici e settore pubblico e privato e del loro riutilizzo ed il regolamento n.2854/2023 -Data Act- del 13 dicembre 2023 che si propone di regolamentare il flusso di dati all'interno dell'UE e tra questa e altri Paesi.

Quando i sistemi di IA si "nutrono" o meglio utilizzano l'immagine e la voce di un individuo determinato, occorre sempre il consenso dell'interessato fatte salve le eccezioni legate all'interesse pubblico, a necessità di giustizia o polizia, oppure a scopi scientifici, didattici, culturali. Sul punto, il GDPR inserisce i dati relativi all'immagine e alla voce delle persone all'interno della categoria dei dati biometrici imponendo il rispetto di tali prescrizioni non solo alle grandi piattaforme o aziende ma, anche, ai privati che trattano dati personali al di fuori dell'uso domestico. Al riguardo, un recente caso controverso è stato rappresentato dalle azioni sanzionatorie intraprese in ambito europeo contro Clearview AI (software di riconoscimento facciale): il Garante della privacy italiano ed altri omologhi degli stati membri della UE, sul presupposto della violazione del principio di liceità del trattamento hanno multato la già menzionata società americana per raccolta non autorizzata di immagini dal web.

Problematiche più specifiche rispetto a quelle sin qui descritte, si pongono nella distinta ipotesi dei deepfake allorché la diffusione di tali dati personali avvenga distorcendo la realtà per il tramite

della riproduzione di immagini o voci artefatte sulle piattaforme digitali.

Si tratta di temi che meglio verranno approfonditi nel prosieguo, ma che in questa sede devono essere evidenziati per l'ulteriore momento di collegamento tra l'AI ACT e la normativa contenuta nel regolamento (UE) 2022/2065, meglio nota come Digital Services Act-DSA, che stabilisce gli obblighi che le piattaforme digitali devono osservare in materia rimuovendo i contenuti illeciti.

Al riguardo, il sempre più frequente utilizzo di tali tecniche di IA generativa per finalità illecite, anche da parte delle organizzazioni terroristiche quando ad esempio alimentano e diffondono i discorsi di odio[94] sulle piattaforme, e la oramai acquisita consapevolezza dei potenziali effetti distorsivi e manipolatori che ne conseguono, ha indotto diversi Stati[95] ad introdurre specifiche fattispecie incriminatrici per meglio disciplinare questo specifico aspetto.

§§§§§

2. IA generativa e diritto penale: profili di responsabilità e prospettive di regolamentazione.

Come rilevato in precedenza[96], l'intelligenza artificiale generativa si configura come una tipica espressione dei sistemi c.d. "sapienti", propri della cosiddetta *Strong AI,* i quali non si limitano a emulare le capacità cognitive umane, ma si

[94] Sull'utilizzo dei deepfake da parte delle organizzazioni terroristiche cfr. *infra*-capitolo par.2.3.

[95] T. RUOCCO *L'avanzata delle leggi contro i deepfake* in www.agendadigitale.eu; M. MARTORANA *Lotta al deepfake: stato dell'arte nell'UE* in www.altalex.com

[96] Cfr. *infra*-capitolo primo, par. 1.2. e 3.1.3.

contraddistinguono per la loro attitudine a svilupparle autonomamente. I sistemi in questione si fondano sul principio del *self- learning,* ossia sull'apprendimento automatico derivante da dati ed esperienze pregresse, che consente loro di migliorare progressivamente attraverso l'analisi di nuove fonti informative e l'osservazione degli *output* generati.

In tale ambito si colloca la questione dei deepfake che secondo la definizione contenuta nell'art. 3, comma 60, del regolamento AI ACT, consiste in "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".

Al riguardo, l'AI Act approccia il tema dei deepfake limitatamente a due aspetti: quello definitorio appena descritto e quello relativo alla imposizione ai produttori e agli utilizzatori degli obblighi di trasparenza[97]. Tali obblighi sono prevalentemente tarati su un uso di detti sistemi di IA nell'ambito dell'arte, della parodia e della satira; per tali ragioni e, a titolo meramente esemplificativo, se nel corso di una rappresentazione teatrale viene artificialmente alterata la voce di un politico, tale attività è consentita, ma chi utilizza i già menzionati sistemi ha l'obbligo di dichiararlo per evitare di ingannare il pubblico presente nella già richiamata ottica di tutele dei diritti fondamentali.

L'arte, la parodia e la satira costituiscono l'unico ambito nel quale lecitamente si può sovrapporre il falso al vero.

Al di fuori di questo ambito le alterazioni non sono consentite ed il giurista si trova a dover inquadrare la tematica dei deepfake tra fattispecie esistenti e vuoti normativi. In tal modo, il tema in questione ci riporta nella logica del metaverso e delle problematiche allo stesso collegate sotto il profilo dell'impatto che l'evoluzione tecnologica può avere con il diritto penale sostanziale.

[97] Cfr. *infra*-capitolo primo, par. 3.1.3

A tal fine per colmare i possibili vuoti normativi, devono essere registrati, sia in ambito europeo che in quello extra-europeo, numerosi interventi che a diverso titolo hanno disciplinato la materia in questione.

§§§

2.1. La rilevanza penale dei deepfake in Italia tra vuoti normativi e prospettive di nuove incriminazioni.

In Italia non esiste al momento una specifica fattispecie incriminatrice in materia di deep fake.
Nell'attuale vuoto normativo, la creazione e la diffusione di deepfake che danneggiano la reputazione di una persona può configurare il reato di diffamazione aggravata punito e previsto dall'articolo 595, comma terzo, del Codice penale che prevede pene fino a tre anni di reclusione per chi diffonde contenuti falsi lesivi della dignità di un individuo. In tali casi se vengono utilizzate immagini o dati biometrici senza il consenso dell'interessato potrà trovare applicazione anche l'art.167 del codice della privacy come conseguenza dell'illecita diffusione.
De iure condito, la citata fattispecie dell'art.595 c.p. è destinata ad essere applicata anche ai casi di manipolazione politica allorquando vengano distorti i discorsi introducendo nel dibattito politico elementi di falsità con audio o immagini false di esponenti parlamentari.
Invece, nel caso in cui il deepfake abbia un contenuto sessuale lo stesso a legislazione vigente potrà trovare applicazione la fattispecie dell'art. 612- ter, c.d. legge sul revenge porn, che punisce la diffusione illecita di materiale sessualmente esplicito, se riguarda una persona non pubblica ed è divulgato senza consenso.

Infine, potranno trovare applicazione l'art. 494 del Codice penale che punisce la sostituzione di persona qualora il deepfake sia usato per ingannare qualcuno sulla propria identità nonché le fattispecie di istigazione previste nel codice penale nell'ipotesi in cui alla pubblicazione di una immagine segua l'incitazione alla violenza.
In tale scenario, il legislatore italiano si sta muovendo per colmare il vuoto normativo in materia di deepfake.

§

2.1.1. La proposta di introduzione dell'art.612 quater c.p. e i limitati ambiti di operatività della fattispecie.

Sul punto, all'interno del già richiamato intervento di carattere generale, l'art.26 del DDL sull'intelligenza artificiale n.1146-B prova a contrastare detto fenomeno tramite la proposta di introduzione dell'art. *612-quater* c.p. - *Illecita diffusione di contenuti generati o manipolati artificialmente,* che prevede: "Chiunque cagiona un danno ingiusto ad una persona cedendo, pubblicando o altrimenti diffondendo, senza il suo consenso, immagini, video o voci falsificati o alterati mediante l'impiego di sistemi di intelligenza artificiale e idonei a indurre in inganno sulla loro genuinità, è punito con la reclusione da uno a cinque anni. Il delitto è punibile a querela della persona offesa. Si procede tuttavia d'ufficio se il fatto è connesso con altro delitto per il quale si deve procedere d'ufficio ovvero se è commesso nei confronti di persona incapace, per età o per infermità, o di una pubblica autorità a causa delle funzioni esercitate".
Dal punto di vista della collocazione sistematica, il legislatore inserisce la norma in questione nel Codice penale all'interno dei delitti contro la libertà morale ed immediatamente dopo l'art. 612 *ter* c.p. che punisce la *diffusione illecita di immagini o video sessualmente espliciti* c.d.revenge porn.

Quanto al bene giuridico tutelato dalla fattispecie, lo stesso deve essere individuato nella libertà morale, nel decoro e nell'onore della vittima, allorquando risulti leso dalla circolazione abusiva di contenuti dannosi, generati o manipolati tramite gli strumenti di intelligenza artificiale e capaci di indurre in errore circa la loro genuinità.

Quanto al soggetto attivo, la fattispecie in questione si configura come un reato comune non richiedendo alcuna particolare qualifica da parte dell'autore del reato.

Quanto alla condotta incriminata, il legislatore punisce l'azione di chi invia, consegna, cede, pubblica, o comunque diffonde immagini o video di persone o di cose ovvero di voci o suoni in tutto o in parte falsi, generati o manipolati mediante l'impiego di sistemi di intelligenza artificiale, atti a indurre in inganno sulla loro genuinità o provenienza.

Quanto all'evento, la descritta azione deve aver cagionato ad altri un danno ingiusto come tale comprensivo di diverse tipologie di pregiudizio, tra cui quello morale, materiale, alla dignità, ecc.. In tal modo la formulazione della norma come fattispecie di danno copre tanto i deepfake lesivi dell'onore o della reputazione, quanto quelli che integrano cyberbullismo o attacchi a figure istituzionali.

Quanto all'elemento soggettivo è richiesto il dolo generico e quindi la coscienza e volontà di cagionare un danno ingiusto non essendo previsto nella norma un ulteriore elemento di carattere finalistico o funzionale.

Il delitto è punibile a querela di parte, salvo che il reato sia connesso con uno procedibile d'ufficio o sia commesso ai danni di persona incapace per età o per infermità, o di una pubblica autorità a causa delle funzioni esercitate.

Per il reato in questione è prevista una pena da l a 5 anni di reclusione che consente l'accesso all'art. 131 bis c.p. non essendo il reato, a differenza della fattispecie delineata nell' art. 612 ter

c.p., incluso tra i reati ostativi una concessione del beneficio della particolare tenuità del fatto.

Una volta esaurita l'analisi degli elementi costitutivi del reato, occorre osservare come l'entrata in vigore dell'art. 612-quate comporterebbe un duplice vantaggio: da un lato dare alle forze dell'ordine uno strumento diretto per perseguire chi diffonde deepfake malevoli (senza dover "far rientrare" forzatamente la condotta in fattispecie preesistenti);dall'altro lanciare un chiaro segnale deterrente in un momento in cui questi fenomeni stanno emergendo anche in Italia (specialmente tra i giovani, con implicazioni di bullismo e violenza online).

Tuttavia, in dottrina sono state sollevate delle considerazioni critiche relativamente al limitato raggio di azione della fattispecie in questione[98].

In particolare, è stata sottolineato come il legislatore abbia costruito una fattispecie strutturata con livelli progressivi e non alternativi dei requisiti che devono produrre l'evento in quanto la falsità totale o parziale del contenuto non esaurisce il terreno di indagine dell'interprete che, invece, dovrà accertare che lo stesso oltre ad essere falso sia stato generato manipolato da un sistema di IA che abbia una idonea capacità decettiva.

Secondo tale orientamento dottrinale, tali considerazioni unitamente al confinamento della fattispecie all'interno di spazi già penalmente presidiati potrebbe sensibilmente ridurre l'ambito operativo della norma in questione.

A quest'ultimo riguardo si evidenzia la prevalenza della più grave fattispecie di stalking nei casi in cui l'autore del reato dovesse inviare ripetutamente alla vittima dei video generati con l'IA in cui, tramite il c.d. *deep fake*, la persona offesa appaia in atteggiamenti ridicoli e poco decorosi che in realtà non ha mai assunto.

[98] F. COPPOLA, *Note in prima lettura sulle disposizioni penali del c.d. DDL "Intelligenza Artificiale"* in Iura and Legal Systems, 2024, 102

In questi casi se per l'abitualità della condotta molesta si provocasse nella persona offesa uno specifico danno ingiusto, consistente nel "perdurante e grave stato di ansia o di paura", ovvero "un fondato timore per l'incolumità propria o di un prossimo congiunto o di persona al medesimo legata da relazione affettiva", oppure lo si costringesse "ad alterare le proprie abitudini di vita", l'autore del reato risponderebbe del più grave reato di *stalking* con la conseguenza che l'offesa sarebbe già ampiamente presidiata, ancorché realizzata tramite strumenti di intelligenza artificiale.

Tale limitato raggio di azione dell'art. 612 quater c.p. impone di svolgere ulteriori riflessioni.

Se l'obiettivo del legislatore è quello di arginare la diffusione di contenuti falsi e pregiudizievoli per la generalità dei consociati, appare del tutto condivisibile l'opinione di chi ritiene più coerente concentrare l'attenzione sulla capacità ingannatoria del contenuto a prescindere dallo strumento tecnico impiegato per generare l'inganno stesso. In questa prospettiva risulterebbe probabilmente più efficace la costruzione di un modello delittuoso di pericolo concreto teso a salvaguardare l'utente medio secondo i parametri costituzionali.

Infine, il ddl in questione, oltre al nuovo reato, annuncia iniziative di sensibilizzazione (un vademecum sui deepfake è allegato al progetto) e strumenti di supporto alle vittime. È da notare che questa proposta si ispira in parte a precedenti stranieri (come vedremo, Francia e Germania) ma costruisce una fattispecie autonoma focalizzata sull'inganno mediante IA.

§

2.1.2. La proposta di introdurre l'art.9-novies della legge n. 212/1956 per reprimere l'utilizzo dei deepfake in materia di propaganda elettorale.

La descritta iniziativa legislativa evidenzia la volontà di legiferare sul fenomeno deepfake in modo dedicato, trovando un punto di equilibrio tra libertà di creazione digitale e tutela delle vittime. Il legislatore italiano dovrà tenere conto sia dei vincoli costituzionali (art. 21 libertà di espressione – da contemperare con altri diritti) sia del contesto europeo: la norma nazionale dovrà integrarsi con l'AI ACT e con le garanzie del GDPR. Interessante notare è che, parallelamente, in Parlamento erano stati depositati anche disegni di legge di iniziativa parlamentare sulla falsificazione di immagini (ad es. uno per estendere il reato di diffusione di immagini sessualmente esplicite anche alle falsificazioni digitali), confluiti poi nel progetto unico del Governo. L'auspicio di molti giuristi è che l'Italia diventi uno dei primi Paesi UE a dotarsi di una legge anti-deepfake, fungendo da modello.

Un discorso a parte merita la c.d. manipolazione politica che può avvenire distruggendo l'immagine pubblica del candidato, o del partito o del movimento di riferimento, distorcendone i discorsi ed introducendo nel dibattito politico elementi di falsità con audio o immagini false. In questi casi nessuna contronarrazione, correzione, tentativo di eliminazione è in grado di sanare completamente l'effetto del falso. Queste situazioni normalmente sono destinate ad amplificarsi durante le competizioni elettorali. Per tali ragioni in Italia in data 23 gennaio 2025, con Atto della Camera dei deputati, n. 2212, è stata presentata la proposta di legge (ddl C.2212) volta alla modifica della legge n. 212/1956, che disciplina il periodo del c.d. "silenzio elettorale". Il divieto di propaganda elettorale, ad oggi, non può essere applicato ai social network e ad altri strumenti telematici che non sono espressamente menzionati in suddetta legge. Si è pertanto avvertita l'esigenza, soprattutto negli ultimi anni, data la diffusione della categoria di intelligenza artificiale generativa in questione di introdurre un'apposita disciplina che punisca la

diffusione di contenuti ingannevoli, in grado di manipolare la veridicità dell'informazione durante il periodo elettorale e, dunque, di alterare le sorti della votazione. L'Atto della Camera propone di introdurre all'interno della legge n. 212/1956, l'art. 9-novies, volto a sanzionare penalmente "chiunque, al fine di alterare il libro svolgimento delle campagne elettorali o referendarie o di manipolarne il risultato cede pubblica o altrimenti diffonde contenuti ingannevoli o manipolati generati in tutto o in parte con intelligenza artificiale[99].

§§§

2.2. Evoluzione delle proposte di legge contro i deepfake negli altri Paesi Europei.

Nel 2023, il Regno Unito tramite l'Online Safety Act[100], ha introdotto il reato di condivisione di deepfake pornografici non consensuali, equiparandolo al revenge porn. Si prevede anche di criminalizzare presto la creazione stessa di tali immagini (il governo britannico ha annunciato un emendamento nel prossimo Victims and Prisoners Bill per punire chi "fa deepfake intimi" anche se non li diffonde). Il Regno Unito ha scelto quindi un approccio mirato: punire i deepfake più lesivi (sessuali) come reati specifici, mentre per altre tipologie si affida al diritto esistente. In ambito politico-elettorale, ad oggi non c'è una legge dedicata ai deepfake: se ne è discusso durante l'iter dell'Online Safety Act ma si è optato per lasciare alla normativa sulle false communications e alla vigilanza della Commissione Elettorale il

[99] Sulla tematica dei deepfake in concomitanza con le competizioni elettorali, come visto nel capitolo primo, la Commissione UE ha sollecitato le maggiori piattaforme a rendere riconoscibile l'origine IA dei contenuti in vista delle elezioni europee 2024.
[100] www.legislation.gov.uk

compito di affrontare eventuali video fake su candidati. Il Regno Unito sta anche investendo in tecnologie di autenticazione: il governo ha finanziato progetti per certificare media autentici con impronte digitali (ad es. nell'ambito della Coalition for Content Provenance and Authenticity – C2PA). Inoltre, l'Online Safety Act conferirà poteri al regolatore Ofcom per chiedere alle piattaforme maggiori misure contro i "contenuti manipolati" che causano danni (anche se il confine con la disinformazione politica è delicato, ragion per cui il governo ha limitato l'obbligo solo ai contenuti illegali e a quelli legati ad autolesionismo e simili, escludendo la disinformazione politica dal regime obbligatorio). In sintesi, l'approccio UK è intermedio: punire penalmente i deepfake più odiosi, ma per il resto affidarsi alla coregolamentazione (Ofcom + piattaforme) e agli strumenti generali.

*

In Francia a fine 2024 è stato presentato all'Assemblée Nationale un disegno di legge per imporre la segnalazione obbligatoria dei contenuti generati da IA sui social network. La proposta (projet de loi n.675)[101] mira a combattere la disinformazione stabilendo che ogni utente che posti un'immagine (o video) alterata/generata da IA debba inserirvi un avviso chiaro e visibile che indichi l'uso di IA. Inoltre, la proposta in questione obbliga le piattaforme a dotarsi di strumenti tecnici di rilevazione automatica di contenuti IA e di meccanismi di segnalazione per gli utenti e prevede sanzioni pecuniarie sia per i singoli (fino a 3.750 € se omettono l'avviso) sia per i gestori (fino a 50.000 € se non rispettano gli obblighi di controllo). Questa iniziativa si affianca a normative francesi già esistenti: dal giugno 2023 infatti, la Francia obbliga gli "influencer" a dichiarare se le foto pubblicitarie postate sono ritoccate o filtrate con IA (Legge 9/6/2023 sulla regolamentazione degli influencer), e ha nel Codice penale l'art.

[101] Proposta di legge n.675 depositata il 3 dicembre 2024 in www.assemblee-nationale.fr

226-8 una fattispecie normativa che già punisce i deepfake sessuali non consensuali (assimilati alle violazioni dell'intimità della vita privata). La nuova proposta estenderebbe tali obblighi a tutti gli utenti e a tutte le immagini IA, non solo fini commerciali. Da notare che il disegno di legge francese prevede anch'esso eccezioni per opere artistiche, satiriche o di finzione, per evitare di colpire creatività e libertà di creazione. In sintesi, la Francia spinge per un approccio molto pratico: etichettatura obbligatoria generalizzata e responsabilizzazione delle piattaforme subito, senza attendere il 2026 dell'AI Act.

*

In Germania nel luglio 2024 il Bundesrat (Camera delle Regioni) ha avanzato un disegno di legge per introdurre un reato di "violazione della personalità mediante falsificazione digitale" (nuovo §201b StGB). La proposta tedesca criminalizza creazione e diffusione di contenuti audio-visivi realistici raffiguranti una persona, ottenuti con tecniche di IA senza il suo consenso, prevedendo fino a 2 anni di reclusione o multa. Il testo tutela anche la memoria dei defunti (estendendo la protezione post mortem) e prevede aggravanti in caso di reati connessi o scopi di lucro, nonché esenzioni per condotte socialmente adeguate (es. parodie evidenti). L'obiettivo, dichiarato, è proteggere i diritti della personalità e il processo democratico dalle manipolazioni via deepfake, specie pensando a scenari elettorali (richiamando casi di disinformazione avvenuti all'estero durante campagne elettorali). Tuttavia, il governo federale tedesco ha reagito con cautela: nell'agosto 2024 il Ministero della Giustizia ha espresso parere contrario ritenendo che le leggi esistenti coprono già molte fattispecie (diffamazione, violazione di intimità tramite immagini §201a, ecc.) e che la nuova norma rischierebbe sovrapposizioni e problemi di determinatezza. In particolare si è criticato il requisito nel draft Bundesrat di dover provare una violazione effettiva di diritti: nel diritto vigente tedesco, per perseguire la diffamazione basta la potenzialità lesiva, mentre la proposta richiederebbe un

danno concreto, paradossalmente restringendo la tutela. Il disegno di legge è ora all'esame del Bundestag, ma appare probabile che subirà modifiche sostanziali o ritardi, vista la resistenza governativa. È interessante notare come Germania e Italia stiano procedendo in parallelo su binari simili (reato ad hoc) ma con esiti diversi: l'Italia appare più propensa a introdurlo subito, la Germania temporeggia ritenendo sufficiente l'arsenale esistente evidenziando approcci differenti nel bilanciamento innovazione/diritti.

*

In altri Paesi UE come Spagna e Paesi Bassi si stanno valutando ipotesi di riforme. In particolare, in Spagna si discute di includere i deepfake nelle fattispecie di calunnia e ingiuria aggravate o nella legge sulla violenza di genere (se usati per diffamare donne); nei Paesi Bassi è in corso un dibattito su possibili obblighi di watermark per i media dopo un noto caso di fake news finanziaria generata da AI. La Polonia ha annunciato nel 2023 un piano per dotarsi di linee guida sulla certificazione dei video autentici, da usare in contesto elettorale contro eventuali deepfake russi. L'Estonia ha già nel 2021 modificato il suo Codice Penale per chiarire che la "diffusione di informazioni false con conseguenze per la sicurezza pubblica" copre anche i media sintetici. Molte di queste iniziative sono influenzate dal contesto locale (ad es. paesi baltici timorosi di campagne di disinformazione ostili).

§§§

2.3. La disciplina normativa dei deepfake negli USA.

In ambito federale, negli USA[102]non esiste una legge omnibus sui deepfake, analogamente a quanto già visto sulla disciplina generale dell'IA.
Tuttavia, negli ultimi tempi si è riacceso il dibattito sulla necessità di introdurre interventi per disciplinare i potenziali effetti distorsivi della IA generativa con particolare riferimento al tema delle fake-news e dei deepfake in relazione ai quali la tendenza è vietare e punire specificamente due macro-ambiti: da un lato i deepfake pornografici non consensuali (equiparandoli a reati sessuali o di cyber-harassment) e dall'altro deepfake in ambito elettorale (vedendoli come minaccia all'integrità del voto).

§

2.3.1. Il No fakes act e l'approvazione del Take it down act.

In ambito federale, la mancanza di una legge omnibus sui deepfake è dovuta ai forti vincoli imposti dal Primo Emendamento che rendono difficile proibire o limitare in astratto la libertà di espressione.
Tuttavia, anche con l'amministrazione Trump sta crescendo la consapevolezza politica del problema ed è stato riproposto il NO FAKES Act (Nurture Originals, Foster Art, and Keep Entertainment Safe Act). Questo disegno di legge introdurrebbe la prima tutela federale del diritto di immagine ("right of publicity") estesa ai digital replica, dando ad artisti, personaggi pubblici e cittadini il diritto esclusivo di autorizzare l'uso della propria voce o sembianze in contenuti generati da IA. In pratica il NO FAKES Act creerebbe uno standard nazionale per

[102]https://www.medialaws.eu/rivista/laregolamentazione del deepfake in europa stati uniti e cina

proteggere le persone dall'uso non consensuale della propria immagine in deepfake, consentendo di agire legalmente contro individui, aziende o piattaforme che diffondano repliche digitali senza permesso. Questo tentativo è sostenuto sia dall'industria dell'intrattenimento (Recording Academy, attori di Hollywood preoccupati da "cloni" digitali) sia da politici di entrambi i partiti, ed è visto come un notevole progresso in un paese dove il diritto di immagine è finora disciplinato solo dalle leggi dei singoli Stati. Accanto a ciò, un altro disegno di legge federale, proposto nel 2019 e denominato Deepfakes Accountability Act puntava a introdurre obblighi di watermark e divulgazione per contenuti IA, ma non è mai stato approvato. Più recentemente, nel 2023, è stato discusso il Rel Accountability Act per vietare certi deepfake in ambito elettorale e di sicurezza nazionale, ma anche questo è in stallo.

Recentemente, il 19 maggio 2025 è stato approvato il "Take It Down Act", una legge federale ufficialmente intitolata Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks (TAKE IT DOWN Act) che incrimina due distinti comportamenti: la pubblicazione non consensuale di immagini intime, comprese le deepfake generate dall'intelligenza artificiale (IA), di adulti e minori e la minaccia di tale pubblicazione.

Come vedremo nel prosieguo, negli Stati Uniti molti stati già sanzionavano la distribuzione non consensuale di immagini intime; tuttavia, mancava una norma federale che includesse anche contenuti IA generati ("deepfake"). Il testo ha ricevuto supporto bipartisan: approvato al Senato per consenso unanime a febbraio 2025 e con votazione di 409-2 alla Camera il 28 aprile 2025

Tra i promotori spiccano i Senatori Ted Cruz (R-TX) e Amy Klobuchar (D-MN), insieme a rappresentanti della Camera come Maria Elvira Salazar (R-FL) e Madeleine Dean (D-PA); la legge ha ricevuto il sostegno anche della first Lady Melania Trump che

ha fatto campagna a favore della legge come parte della sua iniziativa *Be Best*, sottolineando la protezione verso bambini e ragazze vittime di sfruttamento digitale e aggiungendo simbolicamente anche la sua firma, un evento raro nella storia delle leggi federali.

Come meglio vedremo nel capitolo successivo, nella materia in questione il ruolo delle piattaforme digitali diventa fondamentale. Infatti, i siti e le piattaforme devono implementare un sistema di "*notice and takedown*": se ricevono una segnalazione da parte della vittima, devono rimuovere i contenuti in questione entro 48 ore e cercare anche versioni duplicate. La legge prevede sanzioni penali e obbligo di risarcimento per chi pubblica immagini intime senza consenso o minaccia di farlo. Il Federal Trade Commission (FTC) è l'ente incaricato dell'applicazione della legge e può multare le piattaforme per il mancato rispetto delle disposizioni. Tale decisione è inappellabile.

Dal punto di vista sanzionatorio, il TAKE IT DOWN Act (Public Law 119-12) nella forma di emendamento alla Section 223 del Communication Act del 1934 vengono previste pene fino a due anni di reclusione per la pubblicazione reale o alterata di immagini; analoga pena è prevista per la minaccia di pubblicazione. La legge introduce altresì la possibilità per i tribunali di ordinare restituzione dei danni (compresi quelli emotivi o materiali), confisca dei profitti ottenuti con l'illecito e sequestro degli strumenti impiegati per la sua realizzazione.

Le organizzazioni per i diritti digitali e per la libertà di stampa hanno espresso orientamenti critici sul provvedimento normativo in questione sottolineandone i possibili abusi da parte di star dello spettacolo in malafede che potrebbero sfruttare il meccanismo di rimozione anche in assenza di una verifica rigorosa sui potenziali effetti distorsivi nonché la compressione del diritto di difesa in mancanza di un processo di appello per chi ritiene che una rimozione sia errata.

Con specifico riferimento ai deepfake in materia elettorale non esiste una specifica disciplina in ambito federale. Il Deepfakes in Federal Elections Prohibition Act (H.R. 6088), presentato alla Camera nel 2020, si proponeva di vietare la diffusione di video o audio profondamente alterati (deepfake) in prossimità (entro 60 giorni) di un'elezione federale, con intento di danneggiare un candidato o ingannare gli elettori. Prevedeva sanzioni penali fino a 5 anni di carcere, multe o entrambe. Tuttavia, questa proposta non è mai diventata legge.
Viceversa, in ambito di legislazione statale molti stati hanno approvato leggi specifiche per contrastare i deepfake politici, soprattutto in prossimità delle elezioni.

§§§

2.4. La disciplina normativa dei deepfake in Cina e le tecnologie di sintesi.

Come negli USA, anche in Cina non esiste un intervento di carattere regolatorio generale sui sistemi di intelligenza artificiale viceversa abbondano regolazioni settoriali, già in vigore, riguardanti aspetti settoriali, tra cui si evidenziano:

- le misure ad interim per la gestione dei servizi di intelligenza artificiale generativa;
- le disposizioni per la governance dei servizi di informazione con sintesi profonda" meglio nota come regolamento sui deep synthesis o IA generativa;
- le misure sperimentali per una revisione etica delle attività artistiche e tecnologiche;

- le misure amministrative per la regolazione di servizi di raccomandazione algoritmica (conosciute con l'acronimo inglese IISARM)

Come osservato in dottrina[103], tali regolazioni settoriali coprono alcuni settori di grande importanza nello sviluppo dell'IA, e che in parte si sovrappongono con materie che in Europa sono regolate dal AI Act. Nel 2023, il Consiglio di Stato (o Guowuyuan che corrisponde alla nostra Presidenza del consniglio), ha inserito la redazione una legge in tema di IA nel calendario dei lavori. Sullo stato di avanzamento dei lavori annunciati dal Guowuyuan non sono attualmente disponibili informazioni né risultano pubblicate bozze.

In parallelo a tali iniziative la CASS (Chinese Academy of Social Sciences), un'associazione di accademici indipendenti, che ha un peso molto rilevante nell'attività legislativa in Cina, sta sostenendo l'approvazione di una legge in tema di IA cinese.

La proposta di legge sull'IA redatta dal CASS, che già era stato protagonista in occasione dell'approvazione del PIPL (corrispondente del GDPR europeo) si caratterizza, come l'AI Act europeo, per la centralità di una lista negativa, equivalente all'elenco tassativo di sistemi di IA ad alto rischio, previsti nell'AI Act. Tutti i sistemi di IA descritti nella lista negativa dovrebbero essere approvati dalle autorità, prima di poter essere immessi nel mercato.

La proposta non include invece usi vietati dell'IA. Si deve tuttavia aggiungere che numerosi degli usi dell'IA che l'art. 5 dell'AI Act vieta ricadrebbero nelle maglie dei regolamenti cinesi che già vietano alcune pratiche di IA (ad esempio quelle discriminatorie, che nell'AI Act possono essere vietate).

Ciò premesso, soffermeremo la nostra analisi alle disposizioni contenute nelle misure ad interim e a quelle del regolamento sui

[103] G. SANTONI, *Il futuro della regolamentazione IA in CINA ecco i possibili scenari.* in Agenda Digitale del 1° aprile 2025

deep synthesis, riservando l'approfondimento dell'IISARM al capitolo quarto per l'evidente attinenza con il DSA.

§

2.4.1. Le misure ad interim promulgate il 13 luglio 2023, da parte della Cyberspace Administration of China (CAC)

Con specifico riferimento alle misure ad interim le stesse sono state promulgate il 13 luglio 2023, da parte della Cyberspace Administration of China (CAC) insieme ad altri sei ministeri cinesi, e sono entrate in vigore il 15 agosto 2023. Si tratta del primo regolamento amministrativo specifico per i servizi di IA generativa rivolti al pubblico in Cina.
Le misure in questione mirano a regolamentare lo sviluppo e l'uso dell'intelligenza artificiale generativa, cercando di bilanciare innovazione tecnologica e controllo normativo. Il testo stabilisce che i fornitori di servizi di IA generativa devono garantire che i contenuti prodotti siano accurati, legali e privi di informazioni dannose o diffamatorie e, ai sensi dell'art. 5 delle misure ad interim, sono responsabili come se fossero i creatori del contenuto generato attraverso l'IA. A differenza dell'AI Act che prevede espressamente delle ipotesi in cui la violazione di un divieto può essere imputata a chi usi un sistema di IA, la responsabilità del soggetto che inserisce input nel sistema di IA generativa non è affrontata. In effetti, le misure ad interim cinesi prendono in considerazione solo la figura del provider, trascurando quindi di prendere in considerazione la responsabilità del soggetto che contribuisca alla generazione di contenuti illeciti attraverso l'addestramento di un sistema di IA. Inoltre, l'art.16 del regolamento in questione impone il monitoraggio dei contenuti generati per evitare la diffusione di fake news, materiale violento o illegale. Gli utenti devono essere informati quando interagiscono con un sistema di IA generativa, attraverso delle

etichette che segnalano che il contenuto è generato da IA, e devono essere messi in condizione di comprendere i limiti e le funzionalità del sistema. Allo stesso tempo, l'art.22 sempre del regolamento in questione stabilisce che i fornitori di tali servizi devono fornire una funzionalità che consenta agli utenti di segnalare i contenuti generati dall'IA che siano contrari alla legge, sulla falsariga di quanto in Europa il Digital Services Act prevede con riferimento ai contenuti disseminati dalle piattaforme. La differenza tra il meccanismo di notice and takedown previsto dal DSA e quello di report previsto dalle misure ad interim cinesi è che queste ultime consentono l'invio del report direttamente alle autorità nazionali di controllo. Infine, il regolamento contiene quattro distinti elenchi di contenuti illeciti per evitare che vengano generati contenuti contrari alla legge: il primo (art. 4, c. 1), indica contenuti che sono vietati per motivi politici. Il contenuto generato dall'IA deve rispettare i valori cardine del socialismo, non incitare alla rivolta contro il sistema socialista, non incitare al separatismo (si pensi alle problematiche legate a Hong Kong, al Tibet, allo Xinjiang, ma anche a Taiwan, che, facendo uecialmente parte della Repubblica Popolare cinese, dev'essere rappresentata come una provincia cinese, pena la violazione della norma). Anche le informazioni false o violente rientrano in questa categoria; il secondo elenco ci riporta a obiettivi di regolazione comuni anche al AI Act. Secondo l'art. 4, c. 2, il processo di addestramento del IA deve garantire attraverso specifiche misure che l'IA generativa non operi discriminazioni basate su etnia, fede religiosa, nazionalità, regione (la discriminazione tra regioni esiste anche in Cina), sesso, età o professione. Si noti che la norma menziona la discriminazione basata sulle preferenze sessuali; il terzo elenco riguarda illeciti che producono un danno concorrenziale, ovvero che rischiano di compromettere l'integrità del mercato. Così, l'IA deve rispettare la proprietà intellettuale, l'etica commerciale, e gli algoritmi non possono essere usati per distorcere la concorrenza;

d) infine, l'ultimo elenco riguarda i danni ai diritti delle persone fisiche. Quindi l'IA deve essere progettata in modo da impedire danni alla salute fisica e mentale, ai diritti d'immagine, alla reputazione, alla riservatezza.

§

2.4.2. Il regolamento sui deep synthesis o IA generativa emanato dalla Cyberspace Administration of China (CAC)

Le Provisions on the Administration of Deep Synthesis in Internet-Based Information Services, meglio conosciute come regolamento sui deep synthesis, sono state approvate il 25 novembre 2022 dalla Cyberspace Administration of China (CAC), dal Ministero dell'Industria e dell'Information Technology (MIIT) e dal Ministero della Pubblica Sicurezza (MPS), e sono entrate in vigore il 10 gennaio 2023 e hanno come obiettivo quello di disciplinare le tecnologie di sintesi[104] viste come potenzialmente destabilizzanti per l'ordine pubblico.
Tale regolamento impone obblighi stringenti a fornitori e utilizzatori di tecnologia deepfake, con obiettivi di sicurezza, trasparenza e tracciabilità ed in particolare:

- ogni contenuto generato con IA deve essere chiaramente etichettato come tale, ad esempio con un watermark digitale inserito nel file
- i fornitori di servizi di sintesi, dalle app di face-swap ai tool vocali, devono far sì che la loro tecnologia non violi leggi o metta in pericolo la sicurezza nazionale, l'ordine pubblico o la morale;

[104] Le tecnologie di sintesi vocale, note anche come text-to-speech (TTS), convertono il testo scritto in linguaggio parlato, creando un audio dal suono naturale. Queste tecnologie sono ampiamente utilizzate per rendere i contenuti digitali più accessibili, supportare persone con difficoltà di lettura e migliorare l'esperienza utente in diverse applicazioni

- i fornitori di servizi di sintesi devono anche implementare meccanismi per verificare l'identità reale degli utenti (real-name registration) al fine di prevenire l'anonimato nell'uso di deepfake. In pratica, per usare un software di deepfake in Cina bisogna registrarsi con il proprio documento, e i file prodotti devono avere un'identificazione incorporata.

Con specifico riferimento alle piattaforme online cinesi, da WeChat a Weibo, il cui ruolo verrà meglio approfondito nel capitolo quarto, sono obbligate a rimuovere prontamente deepfake che diffondano notizie false o contenuti illeciti, e a segnalarli alle autorità.

Il regolamento enfatizza, altresì, il contrasto alla disinformazione: la creazione e diffusione di fake news tramite deepfake è espressamente vietata e può portare a conseguenze penali. Ad esempio, un'azienda che non etichetta un video deepfake rischia pesanti sanzioni amministrative e la sospensione del servizio, mentre l'utente che lo ha caricato può essere arrestato per "diffusione di voci" (reato già esistente).

Inoltre, la Cina tutela anche la privacy biometrica: la legge sulla protezione delle informazioni personali (PIPL, simile per certi versi al GDPR) classifica le caratteristiche biometriche come dati sensibili che richiedono il consenso esplicito per essere usati. Ciò significa che creare un deepfake del volto di qualcuno senza autorizzazione viola anche la PIPL, aggiungendo un ulteriore illecito. Un caso concreto: nel 2021 l'app cinese Zao (che permetteva di fare face-swap con celebrità) fu costretta a modificare i propri termini dopo l'intervento dell'autorità, perché raccoglieva in modo invasivo le immagini degli utenti. L'episodio di Zao anticipò l'attuale stretta normativa.

§

2.4.3. Considerazioni conclusive sulla normativa cinese in materia di IA generativa.

Conclusivamente, la Cina adotta un approccio quasi opposto a quello americano, molto più preventivo e centralizzato nel controllo dei deepfake sul presupposto che gli stessi non devono minare la sicurezza pubblica, la privacy individuale e la fiducia. Pertanto, si obbliga a marchiare ogni contenuto artificiale, a identificare chi lo produce e a bloccarlo se pericoloso. Questa filosofia si riallaccia anche alla censura di Stato: la Cina vuole controllare la narrativa online e i deepfake rappresentano un rischio (pensiamo a video satirici su leader politici: sarebbero subito censurati come "fake harmful to public order"). Di recente, le autorità cinesi hanno persino lanciato campagne per educare il pubblico a riconoscere i deepfake e a segnalarli. L'effetto pratico è che in Cina i grandi social come TikTok (versione locale Douyin) e Tencent usano algoritmi di rilevamento automatico: ad esempio Douyin etichetta i video sospetti con avvisi e ha "bannato milioni" di clip IA non conformi. Per le aziende che sviluppano AI generativa (tipo Baidu con ERNIE, equivalente di ChatGPT) è obbligatorio implementare funzioni di watermarking integrate nei loro modelli.

Il confronto con l'UE è netto: l'UE con l'AI Act chiede etichette ma non impone identità reale né filtri preventivi universali; la Cina invece sì, a scapito magari della libertà individuale ma con maggiore efficacia immediata nel contenere il fenomeno. Va detto che la Cina può permettersi tale approccio perché non ha le stesse garanzie di free speech: la costituzione cinese subordina la libertà di espressione all'interesse dello Stato; dunque, obblighi che in Occidente sarebbero incostituzionali (come la registrazione obbligatoria per postare contenuti) lì sono la norma. In conclusione, la Cina vede i deepfake soprattutto come una minaccia sociale e li affronta con leggi severe e controlli tecnici: è uno dei paesi più avanzati nel dotarsi di una cornice regolatoria

per l'IA generativa, e probabilmente influenzerà altre nazioni autoritarie a seguire l'esempio.

§§§§§

3.Dal revenge porn ai reati terroristici: l'impatto dei deep fake sulle decisioni giudiziarie.

Nel capitolo secondo abbiamo già esaminato come in ambito accademico ed istituzionale, recenti studi abbiano evidenziato la tendenza di alcuni gruppi terroristici[105], tra cui ISIS, Al-Qaeda, Hamas, Hezbollah ad esplorare l'uso dell'IA generativa per scopi propagandistici, contenuti in linguaggi diversi, meme e immagini manipolate per intensificare l'impatto emotivo e attrarre nuovi adepti.

Tuttavia, l'analisi casistica non offre decisioni giudiziarie riferibili all'utilizzazione dei deepfake per alimentare discorsi di odio a fine di radicalizzazione nella materia del terrorismo[106].

Invero, l'analisi condotta sui casi giudiziari esistenti, non solo in Italia ma anche all'estero, e sulle relative azioni legali intraprese, consente di concentrare la interferenza dei deepfake in tre principali aree:

- revenge porn;
- condotte diffamatorie;
- manipolazione politica.

Sui temi in questione la risposta dei tribunali è in continua evoluzione muovendosi tra vuoti di tutela, applicazioni di norme esistenti ed evoluzioni normative.

[105] UN News *Terror threat posed by ISIL 'remains volatile and complex,' Security Council hears* in news.un.org

[106] B.G. SCHARFFS, *A Commitment to Religious Freedom as the Bond that Makes Us Free*, in The Review of Faith & International Affairs, n. 4/22, p. 24.

Peraltro, in molte delle situazioni che verranno illustrate alla rilevanza mediatica dei casi non è poi seguita la trasposizione degli stessi nell'ambito giudiziario sicché stenta ancora a sedimentarsi un consolidato orientamento sul punto.
Le scarne decisioni giudiziarie che via via si stanno sedimentando appaiono comunque convergere su alcuni punti irrinunciabili: trasparenza obbligatoria, consenso per l'uso dell'immagine altrui, sanzioni severe per chi causa danni reali.
Procediamo ora ad esaminare i casi giudiziari più significativi che hanno portato la materia in questione all'attenzione dei Tribunali variamente interessati.

§§§

3.1. Deepfake e diffusione non consensuale di immagini intime: prospettive giurisprudenziali comparate.

In Italia, non esistono decisioni giudiziarie in materia. Nel 2021 il Garante della privacy italiano[107] aprendo un'istruttoria è intervenuto con riferimento al caso DeepNude, un software che utilizzando sistemi di IA era in grado di generare falsi nudi da foto reali. La decisione assunta è stata quella di bloccare il software a livello internazionale, evidenziando possibili violazioni della dignità umana.
Sempre in Italia, un episodio emblematico rispetto al quale non siamo ancora in presenza di una decisione giudiziaria è costituito dalla vicenda che coinvolge l'attuale Presidente del Consiglio Giorgia Meloni. Nel 2023 è emerso che, alcuni anni prima, due uomini origini di Sassari avevano diffuso online dei video pornografici deepfake con il volto della Meloni sovrapposto a quello di un'attrice. Tali video, caricati su un sito pornografico

[107] Vademecum garante della privacy sul fenomeno dei deep fakes Cfr. https://www.garanteprivacy.it/home/docweb/-/docweb display/docweb/9512278

estero, hanno totalizzato milioni di visualizzazioni, causando clamore mediatico. La Meloni, divenuta Primo Ministro, ha agito in sede civile nei loro confronti lamentando a tal fine la lesione dei diritti della personalità (immagine e reputazione) e la violazione della privacy e chiedendo un risarcimento 100.000 da devolvere a fondi per vittime di violenza domestica. Il processo è attualmente in corso a Sassari ed è destinato a costituire un precedente importante: infatti è la prima volta che una figura istituzionale italiana intraprende un'azione legale contro autori di deepfake, e il suo esito fisserà criteri risarcitori e interpretativi rilevanti[108].

Nel Regno Unito, nel 2025 si è concluso uno dei primi processi: un ventiseienne dell'Essex, Brandon Tyler, è stato condannato a 5 anni di reclusione per aver creato e diffuso online centinaia di immagini e video pornografici falsi aventi come vittime circa 20 donne a lui conosciute (amiche, ex colleghe). L'uomo aveva usato foto ordinarie delle vittime (prese dai social) per manipolarle con i sistemi di IA, inserendole in scenari sessualmente espliciti, e poi le aveva pubblicate su forum incitanti allo stupro. Il giudice, definendo la condotta come "la peggiore forma di mascolinità tossica", lo ha ritenuto colpevole sia per molestie (harassment) sia per condivisione di immagini intime senza consenso. Questa sentenza ha segnato la prima applicazione concreta del divieto sui deepfake sessuali in ambito anglosassone e merita di essere segnalata in quanto ha affermato che la condivisione di deepfake sessualmente espliciti deve essere considerato un fatto costituente reato[109].

Sempre nel Regno Unito, un altro caso nel 2024 ha destato orrore: un uomo di Bolton, Hugh Nelson, è stato condannato alla pena di 18 anni di reclusione per aver utilizzato l'IA al fine di creare materiale pedopornografico estremamente realistico. Durante lo svolgimento del processo è emerso che Nelson prendeva foto

[108] M.CARTISANO, *Deepfake porno con Giorgia Meloni, ecco tutti i reati commessi,* in agendadigitale.eu, 25 ottobre 2023

[109] Tale decisione può essere consultata su essex.police.uk

normali di bambini (anche fornitegli da pedofili conosciuti online) e le trasformava, tramite un software che utilizzava sistemi di IA in immagini di abusi sessuali su minori. La gravità del caso, approdato per la prima volta innanzi ai tribunali britannici, e la severità della pena detentiva dimostrano come i giudici considerino i deepfake in ambito sessuale altrettanto dannosi degli abusi reali, punendoli di conseguenza[110].

In Spagna, nel 2022, un falso video hard attribuito alla star internet Sfera Ebbasta generato con volti di attori ha determinato la vittima a sporgere denuncia per diffamazione.

Negli USA, nell'ottobre 2023, a 14 anni, Elliston Berry, studentessa della Aledo High School in Texas, scopre che un compagno di classe aveva preso foto innocenti da Instagram per generare, tramite IA, immagini pornografiche false ritraenti lei e alcune compagne nude. Tali immagini sono state poi diffuse su Snapchat e in altri canali social per nove mesi, senza che le piattaforme intervenissero tempestivamente[111]. La rilevanza mediatica della vicenda in questione nonostante l'assenza di una decisione giudiziaria ha portato le autorità americane ad approvare il citato "Take it down act".

In Cina, nell'aprile del 2023 una nota influencer cinese, conosciuta come "CaroLailai", ha scoperto che il suo volto era stato sostituito con quello di un'attrice pornografica in un video virale e successivamente ha denunciato la catena illegale alla polizia, sollevando attenzione sull'uso illecito della tecnologia deepfake a fini diffamatori e sessuali[112].

Infine, sempre in Cina un caso esplosivo, noto come "MaskPark", ha coinvolto una rete Telegram con oltre 100.000 membri, dove video sessualmente espliciti sintetici - creati con deepfake a partire da immagini riprese con videocamere nascoste - venivano

[110] Per informazioni sul caso in esame cfr.cps.gov.uk
[111] cfr. spectrumlocalnews.com
[112] L'unica menzione significativa di questa vicenda si trova in un articolo dello studio legale HSF Kramer che riporta la notizia della denuncia di tale vicenda

condivisi e venduti. Una donna, "Ming", ha scoperto di essere stata filmata di nascosto in contesti privati dal suo partner che successivamente generava immagini pornografiche tramite IA condividendole sulla indicata piattaforma. Il caso ha sollevato forte indignazione, mettendo in luce le difficoltà legali per le vittime, tra cui l'onere della prova gravoso, scarsa formazione delle forze dell'ordine, e forte riprovazione sociale[113].

§§§

3.2. La giurisprudenza sui deepfake diffamatori: tra libertà di espressione e tutela della reputazione

Come abbiamo detto non esiste una casistica giudiziaria relativamente all'utilizzo dei deepfake in materia di terrorismo.
Tuttavia, l'area di intervento in questione sollecita una preliminare riflessione sulla difficoltà, anche da parte dei giudici, di individuare un punto di equilibrio tra diritti fondamentali ed esigenze di sicurezza nelle democrazie contemporanee. Dopo l'11 settembre 2001, il terrorismo ha spinto gli Stati a introdurre norme che spesso incidono in senso restrittivo sulla libertà di espressione[114]. La questione centrale è individuare il confine tra libera manifestazione del pensiero e messaggi che istigano alla violenza.
Tali difficoltà ovviamente si ripercuotono anche con riferimento alle situazioni in cui la libertà di espressione si traduce in offese all'onore e alla reputazione delle persone.

[113] Le agenzie di stampa ABC News e Reuters confermano la notizia riportando che non sono emersi annunci o azioni pubbliche da parte delle autorità statali in Cina legate a questo caso
[114] DE CARIA, *Il dibattito sull'online speech: alla ricerca di un fondamento ideologico solido e coerente, tra liberta` di espressione, ''economic speech'' e liberta` di iniziativa economica*, in Dir. econ., 2023, 315.

In Georgia, nel maggio 2023, il conduttore radiofonico Mark Walters intentava una azione legale nei confronti di OpenAI, dopo aver scoperto che ChatGPT lo aveva falsamente accusato di appropriazione indebita (embezzlement) e frode dalla Second Amendment Foundation. Nel gennaio 2024, il giudice respingeva la richiesta di archiviazione presentata da OpenAI, permettendo in tal modo che la questione venisse portata davanti ad un giudice. Nel maggio 2025, il tribunale della Georgia ha emesso un summary judgment in favore di OpenAI, chiudendo il caso definitivamente. Le argomentazioni principali del summary judgment vertevano sul fatto che: il testo di ChatGPT non poteva ragionevolmente essere interpretato come vero (soprattutto grazie ai disclaimer integrati nel sistema); Walters non aveva dimostrato né negligenza né malizia consapevole da parte di OpenAI; Walters ammetteva di non aver subito danni e di non aver richiesto una ritrattazione preventiva, precludendo così risarcimenti; le misure adottate da parte di OpenAI (disclaimer, sistemi di sicurezza, trasparenza) fossero adeguate e come tali impedissero la qualificazione della diffamazione[115].

In Australia, all'inizio di aprile 2023 ChatGPT è stata accusata di diffamazione, allorquando il sindaco Brian Hood ha intentato una azione legale dopo aver scoperto che la chatbot lo aveva falsamente descritto come un corruttore attinto da un provvedimento custodiale[116]. Brian Hood inviava ad OpenAI un "concerns notice" il 21 marzo 2023, disponendo un termine di 28 giorni per correggere le affermazioni false diffuse da ChatGPT. Se non avesse risposto, sarebbe partita una causa per diffamazione. Nel febbraio 2024, Hood ha ritirato l'azione legale intrapresa dopo che OpenAI aveva corretto le affermazioni false tramite aggiornamenti al modello. Nel dicembre 2024, è emerso che ChatGPT non risponde più alle richieste contenenti il nome

[115] La decisione è consultabile sul sito Loeb & Loeb

[116] *Se l'intelligenza artificiale finisce in tribunale ChatGpt diffama un sindaco australiano che ora vuole denunciarla*, la Repubblica, 9 aprile 2023

"Brian Hood". In pratica, il nome è stato bloccato ("hard-coded"), impedendo qualsiasi risposta – anche neutra o positiva – su di lui. Tale atteggiamento da parte di ChatGpt è stato interpretato come una forma di "censura digitale" o digital suppression, suscitando a sua volta critiche per l'assenza di un quadro normativo adeguato finalizzato a proteggere le persone da questo tipo di esclusione automatica.

Negli USA, nell'estate del 2024 l'attivista conservatore Robby Starbuck ha intrapreso una causa per diffamazione da oltre 5 milioni di dollari contro Meta AI, evidenziando responsabilità per affermazioni false generate dall'IA, anche se protette da disclaimer poichè una chatbot di Meta affermava falsamente che fosse coinvolto nell'assalto del 6 gennaio, negazionista dell'Olocausto e colpevole di crimini inesistenti.[117]

Sempre negli Usa la nota star della NBA LeBron James nell'agosto del 2025 ha intrapreso una azione legale a seguito della creazione di video deepfake (tipo "brainrot") che ritraevano LeBron James in scenari surreali e offensivi, come essere incinto, apparire senza tetto, inginocchiarsi con la lingua fuori, o assistere a violenze non consensuali con altri personaggi famosi come Diddy e Stephen Curry. Questi video sono diventati virali su Instagram, con milioni di visualizzazioni. La sua squadra legale ha inviato lettere di diffida (cease-and-desist) a Interlink AI (strumento basato su FlickUp) e ai creatori di tali video, che utilizzavano la sua immagine senza permesso. Subito dopo le diffide Interlink AI ha rimosso tutti i modelli realistici di persone dalla propria piattaforma, per evitare ulteriori controversie legali. Almeno tre account Instagram che condividevano questi video sono stati cancellati[118].

Infine, tra i casi emergenti deve essere menzionata l'azione legale intentata da un utente contro Microsoft falsamente accusato da

[117] Lo studio legale Dhillon Law Group ha dato notizia dell'atto di citazione depositato da Robby Starbuck presso la Delaware Superior Group

[118] L'azione legale in questione è stata ufficialmente documentata dallo studio legale Grubman Shire Meiselas & Sacks

Bing AI di essere un detenuto terrorista. La causa non ha avuto sviluppi essendo poiché è stato richiesto un arbitrato. In generale, questi casi appaiono segnare l'inizio di una nuova era di contenziosi su *hallucinations* da IA e responsabilità legale – con dibattiti su elementi chiave come gli avvisi di disclaimer, la negligenza reale, e la distinzione tra utenti "informati" o "ragionevolmente consapevoli" di limitazioni intrinseche presenti nei modelli[119].

§§§

3.3. Deepfake e manipolazione politica: analisi dei casi giudiziari.

Come detto in precedenza, la c.d. manipolazione politica può realizzarsi attraverso la distruzione dell'immagine pubblica di un candidato, di un partito o di un movimento politico, distorcendo discorsi o introducendo nel dibattito elementi di falsità mediante audio o immagini artificiosamente generati. In simili circostanze, nessuna contronarrazione, rettifica o tentativo di rimozione appare in grado di neutralizzare integralmente l'effetto del falso, che tende comunque a produrre un impatto duraturo e difficilmente reversibile sull'opinione pubblica. Tali dinamiche trovano la loro massima espressione nei periodi di competizione elettorale, nei quali la capacità manipolativa dei contenuti sintetici è destinata ad amplificarsi.

In tale ambito le decisioni giudiziarie su casi di manipolazione politica sono estremamente limitate sicché l'indagine rimane confinata ad un'osservazione dei casi che sono stati portati all'attenzione dell'opinione pubblica.

[119] Zhang, T., et al. (2023). Language Models Are Multilingual But Might Hallucinate When Translating.
Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On Faithfulness and Factuality in Abstractive Summarization. ACL.

In Italia, a livello mediatico un episodio emblematico è costituito dagli attacchi via Twitter nei confronti del Presidente della Repubblica italiana in occasione della formazione del Governo dopo le elezioni politiche del 2018. È ancora in corso una indagine per chiarire le dinamiche dietro la creazione, nella notte tra il 27 e il 28 maggio 2018, di circa 400 account Twitter utilizzati per condizionare l'opinione pubblica e favorire le dimissioni del Presidente della Repubblica Sergio Mattarella a seguita del suo rifiuto della proposta del Presidente del Consiglio incaricato, Giuseppe Conte, di nominare Paolo Savona Ministro dell'Economia. Le indagini si sono concentrate sull' degli account che hanno rilanciato tweet con hashtag diventati virali – #mattarelladimettiti, #impeachment e #impeachmentmattarella – per individuare eventuali bot. In particolare, sono stati svolti accertamenti per individuare: quali criteri siano stati utilizzati per la composizione dello username e le informazioni pubbliche di base del profilo; il rapporto tra numero di following e followers; i collegamenti tra following e followers; i temi trattati nei tweet. Secondo fonti giornalistiche gli accertamenti compiuti hanno consentito di individuare 360 account "anomali" di cui si può ritenere plausibile che, per la loro creazione e contestualizzazione nel brevissimo tempo in cui sono emersi, siano stati utilizzati particolari algoritmi basati sull'intelligenza artificiale[120].

In Germania a febbraio 2024 il Landgericht di Berlino ha emesso un'ingiunzione urgente contro la diffusione di un video deepfake riguardante il Cancelliere Olaf Scholz. Nel video (realizzato da un collettivo artistico provocatorio) Scholz sembrava tenere un discorso ufficiale proponendo di mettere al bando un partito politico tedesco (AfD), con tanto di voce simulata perfettamente somigliante alla sua e immagini autentiche del Cancelliere. Il governo federale ha agito legalmente invocando la tutela del diritto al nome e all'immagine del Cancelliere (diritti della

[120] *La fabbrica dei troll contro Mattarella: attacco su twitter dopo il no a Savona*, www.ilfattoquotidiano.it, 3 agosto 2018

personalità riconosciuti dall'ordinamento tedesco, §12 BGB) e sostenendo che il video ingannevole minava la fiducia pubblica. Il Tribunale, in sede cautelare, ha disposto il divieto di pubblicazione del deepfake, ritenendo che non fosse riconoscibile come satira e che in un'ottica di bilanciamento la tutela dell'identità del Cancelliere prevalesse sulla libertà artistica ed espressiva del collettivo. Nella motivazione i giudici hanno affermato che l'elevata verosimiglianza della voce e l'uso di simboli ufficiali (es. l'aquila federale nel video) potevano indurre lo spettatore medio a credere che si trattasse di un autentico discorso governativo, configurando una grave fake news idonea a "minare la fiducia nella comunicazione pubblica del governo". Pur lasciando aperto la questione relativa al carattere satirico dell'opera, il provvedimento inibitorio in questione che costituisce la prima decisione giudiziaria in Germania su un deepfake politico ha stabilito che la diffusione di deepfake ingannevoli su figure pubbliche può essere fermata per via civilistica dovendo ritenersi prevalente rispetto alla libertà di espressione la tutela dell'interesse pubblico a non essere disinformato. Inoltre, la decisione in questione indica un criterio: quando il contenuto fake replica troppo fedelmente l'originale e non contiene indizi chiari della sua falsità, non può essere invocata la protezione della satira. Il caso Scholz evidenzia dunque il delicato bilanciamento fra libertà di espressione politica e tutela dell'identità/reputazione di fronte alle manipolazioni digitali[121].

In Francia, un militante è stato indagato per aver diffuso nel 2019 un deepfake satirico del Presidente Macron. L'autorità giudiziaria francese ha poi archiviato il caso per irrilevanza penale, data la chiara natura parodistica della vicenda in questione.

In Ucraina, durante l'invasione russa del 2022 ha fatto scalpore un video deepfake del Presidente ucraino Zelenskyj che sembrava

[121] La sentenza integrale è identificabile con il numero di procedura 15 O 579/23 presso il Landgericht Berlin II, datata 13.02.2024

invitare le truppe alla resa; il video, benché rapidamente smentito, ha evidenziato il potenziale bellico-propagandistico dei deepfake. Negli Stati Uniti d'America, nel marzo del 2019 è stato pubblicato in rete un video, poi risultato manipolato con la tecnica dello "shallow fake" della voce di Nancy Pelosi. Nel video, che è stato condiviso il 23 maggio 2019, il discorso di Pelosi era stato rallentato, facendo sembrare che stesse biascicando le sue parole. La versione modificata del video è diventata virale sui social media ed è stata ritwittata dall'account Twitter ufficiale del presidente degli Stati Uniti Donald Trump, ricevendo oltre 6,3 milioni di visualizzazioni al 31 luglio 2019. Su una popolare pagina Facebook, il video ha ricevuto oltre 2,2 milioni di visualizzazioni nelle 48 ore successive al suo caricamento iniziale, con i commentatori che hanno definito Pelosi "ubriaca" e un "pasticcio balbettante"[122].

Sempre negli Usa un altro caso emblematico è costituito dalla manipolazione di movimenti del corpo per la revoca del pass alla Casa Bianca del giornalista CNN Jim Acosta. Un video shallow fake è stato anche citato come prova per giustificare azioni politiche controverse. In questo caso, il corrispondente della CNN Jim Acosta si è visto revocare il suo pass stampa alla Casa Bianca il 7 novembre 2018, a seguito della diffusione di un video in cui si mostra un membro dello staff della Casa Bianca che tentato di togliergli il microfono dopo uno scambio teso con il Presidente Trump.

Ancora negli USA, il 20 luglio 2025 è comparso sulla piattaforma digitale Tik Tok un video generato dall'IA che manipola le immagini dell'incontro del 2016 nello studio ovale quando Obama aveva ospitato Trump vincitore delle elezioni per il passaggio di consegne. Solo che ad un certo punto la realtà viene

[122] S. MERVOSH, *Distorted Videos of Nancy Pelosi Spread on Facebook and Twitter, Helped by Trump*, The New York Times, 24 maggio 2019; C. MONJE JR., *Twitter letter to Chairman Schiff*, Twitter, 31 luglio 2019; D. HARWELL, *'Sexist' videos edited to make Nancy Pelosi look drunk go viral, with Trump's help*, The Independent, 24 maggio 2019.

alterata perché entrano in campo agenti dell'FBI che costringono Barack Obama ad inginocchiarsi per ammanettarlo. Vicino si vede Trump che ride e in sottofondo si sente la canzone YMCA dei Village people colonna sonora della campagna elettorale di Trump. Poco dopo il video viene rilanciato diventando virale.

In Malesia nel giugno 2019 è emerso uno scandalo politico intorno a un sex tape che coinvolgeva presumibilmente il ministro malese degli affari economici Azmin Ali e l'assistente maschile di un ministro rivale. Mentre l'assistente ha affermato che il video era reale ed è stato successivamente arrestato, Ali e i suoi sostenitori, incluso il primo ministro malese, hanno sostenuto che il video era un deep fake realistico fatto per sabotare la sua carriera politica. Tuttavia, gli esperti internazionali non sono riusciti a trovare alcun segno che il video fosse stato manipolato. A metà agosto 2019, non era ancora chiaro se il video fosse stato diffuso nella sua versione originale oppure in una adulterata.

In Gabon nel 2018 è stato postato in rete dal Governo locale un video del Presidente Bongo che rilasciava i tradizionali auguri per il nuovo anno. Il Presidente, che era stato assente per molti mesi dalla vita politica, era oggetto di una intensa speculazione politica sul suo reale stato di salute. Tuttavia, l'insolita apparizione di Bongo nel video ha portato molti sui social media, incluso il politico gabonese Bruno Moubamba, a dichiarare che il video era un deep fake, confermando il loro sospetto che il governo stesse coprendo la cattiva salute o la morte di Bongo. Una settimana dopo la pubblicazione del video tra crescenti disordini, i membri dell'esercito del Gabon hanno lanciato un tentativo di colpo di stato, richiamando proprio la pubblicazione del video. Da analisi forensi non sono però emerse tracce che provino che fosse deep fake ed è emerso che da agosto 2018 il presidente aveva subito un grave ictus[123].

[123] A. BRELAND, *The Bizarre and Terrifying Case of the "Deepfake" Video that Helped Bring an African Nation to the Brink*, Mother Jones, 15 marzo 2019; J. BLAKKARLY, *A gay sex tape is threatening to end the political careers of two men in Malaysia*, SBS News, 17 giugno 2019.

§§§§§

4.Conclusioni

Lo sviluppo industriale della IA - che ha come epicentro l'Europa gli Usa e la Cina – ha determinato diversi modelli regolatori.

L'UE, attraverso il regolamento AI ACT, ha scelto di regolare la materia attraverso un intervento di carattere generale che ha dato luogo ad una normativa eccessivamente dettagliata e strettamente collegata a quella dettata in materia di dati personali (GDPR) e alla circolazione delle informazioni sulle piattaforme digitali (DSA).

Gli USA, con l'amministrazione di Donald Trump volendo rivendicare la leadership in materia, stanno adottando la logica della deregulation, rimandando al libero mercato l'individuazione del punto di equilibrio tra sviluppo economico e tutela dei singoli

La CINA in una logica dirigista ed autoritaria ha scelto di disciplinare singoli aspetti della materia, tra cui in particolare quella della deep synthesis.

Sullo sfondo di tali modelli regolamentari, si innesta l'evoluzione tecnologica del terrorismo ed il sempre più frequente utilizzo dei sistemi di IA da parte delle stesse organizzazioni criminali che, come detto in apertura, potrebbero beneficiare di tale frammentazione normativa tra un'Europa a trazione fortemente regolatoria e gli USA che sembrano avviarsi sulla strada opposta della auto-regolamentazione del mercato.

Tuttavia, c'è uno specifico versante sul quale le diverse legislazioni si stanno avvicinando: quello relativo alla disciplina dei deep fake, in pieno sviluppo a livello globale, e che più da vicino riguarda ed interessa il diritto penale.

Con riferimento alla materia in questione, sia in ambito europeo che al di fuori, i diversi Stati stanno colmando i vuoti normativi attraverso la previsione di fattispecie incriminatrici volte a punire gli effetti distorsivi e manipolatori che conseguono a tale forma di tecnologia.

L'UE, con la sua tradizione di tutela multilivello dei diritti, sta cercando di costruire una risposta completa ma che rispetti i principi democratici: ciò significa permettere l'uso creativo e positivo dell'IA generativa nei film, nei meme, nell'arte e allo stesso tempo predisporre argini normativi contro gli usi perversi, dalla distruzione di reputazione tramite fake porn, alla propaganda ostile.

Infine, con riferimento all'ambito giudiziario, la casisitica giudiziaria ha individuato tre distinte aree di intervento: quella sessuale; quella diffamatorio; quella della manipolazione politica. Non esistono decisioni giudiziarie strettamente attinenti alla materia del terrorismo; tuttavia, in ambito ONU sempre più viene rimarcato l'utilizzo dell'IA generativa da parte delle organizzazioni terroristiche per l'amplificazione dei discorsi di odio.

In tale ambito il bilanciamento tra libertà di espressione e tutela dei diritti individuali resta il filo conduttore: la sfida è garantire che la verità fattuale e la dignità personale non siano travolte dall'onda della manipolazione digitale, senza per questo imbrigliare indebitamente il potenziale creativo dell'IA.

CAPITOLO QUARTO: FORME DI RESPONSABILITA' DELLE PIATTAFORME DIGITALI.

1.Gli obblighi normativi delle piattaforme digitali nel contrasto al terrorismo

Nell'era della transizione digitale un ruolo fondamentale lo svolgono anche le piattaforme digitali[124], luoghi che permettono a persone, aziende o organizzazioni di interagire e, soprattutto, di condividere informazioni per i più disparati fini, sia leciti che illeciti.

Così può accadere che all'interno delle piattaforme ogni individuo possa esprimere liberamente le proprie idee ed opinioni in qualunque ambito e persino "influenzare" per finalità del tutto lecite altri individui, ad esempio, quando vengono formulati suggerimenti per l'acquisto di un prodotto commerciale, o per la pubblicizzazione di un evento oppure per lanciare una iniziativa benefica.

Si tratta di un fenomeno che ebbe inizio all'incirca nel 1990, quando internet iniziò ad essere impiegato come mezzo pubblicitario, affiancandosi a strumenti classici quali la televisione, la stampa e la radio, e quando nacque la figura del testimonial sostituita poi da quella dell'*influencer*.

[124] Cfr. due voci enciclopediche di POLLICINO, *Potere digitale*, e RESTA, *Poteri privati e regolazione*, entrambe in Enc. del dir., I tematici, V, Potere e Costituzione, Giuffre` Francis Lefebvre, 2023.

Il successivo incontenibile sviluppo dei social network ha poi rapidamente ridefinito il modo in cui gli individui si relazionano tra di loro. Questo ha creato un nuovo concetto di comunicazione, che da unidirezionale è diventata interattiva: l'utente finale, infatti, non è più passivo, ma coinvolto direttamente.

Accanto al descritto, uso lecito, tuttavia, esistono situazioni in cui l'utilizzo delle piattaforme digitali avviene, per la commissione di finalità illecite finendo, in tal modo, per interferire con il diritto penale.

In tale ambito si collocano le organizzazioni terroristiche che utilizzano le piattaforme online a fini di propaganda terroristica alimentando con le nuove tecniche di IA generativa i discorsi che incitano all'odio religioso ed istigano al terrorismo.

Sul punto deve essere evidenziato come la propaganda online abbia nel tempo rappresentato uno dei più importanti fattori di successo del terrorismo c.d. islamico; allo stesso tempo le piattaforme digitali e più in generale la rete si sono rilevate il "luogo" d'eccellenza, come anche le carceri, per l'efficace funzionamento dei meccanismi di radicalizzazione[125]. Basti pensare che ai tempi delle Torri gemelle, e durante l'organizzazione delle stragi in Europa, il sistema per la chiamata all'azione, per il consolidamento di scelte radicali, per la adesione all'idea del martirio viaggiava ancora prevalentemente su carta o su audiocassette contenenti le immagini dei martiri, gli inni al martirio, i canti di battaglia, che venivano trasportate fisicamente e clandestinamente, con assunzione di gravi rischi, nei luoghi dell'ascolto in comune e della radicalizzazione. È sufficiente ricordare qui la "capacità" di radicalizzazione che hanno avuto da sempre le foto – ritenute pacificamente vere - delle varie forme di

[125] GALLI, *Prevenzione del terrorismo nell'Unione Europea: un nuovo ruolo e responsabilità per le piattaforme informatiche?* in Rassegna di diritto pubblico europeo, 2019, XVIII; GRAZIANI, *Terrorismo internazionale, radicalizzazione e tecnologia,* in federalismi.it, 2023.
HESS S., KALB M., The media and the war on terrorism, Brookings Institution Press, Washington 2003.

tortura che sono state utilizzate ad *Abu Ghraib* o le immagini degli "arancioni" di Guantanamo, che poi sono state riproposte nelle tuniche fatte indossare agli ostaggi giustiziati dal nascente Stato Islamico. Lo stesso effetto si è cercato di ottenere con le immagini terribili delle vittime dei bombardamenti, soprattutto bambini, accompagnate da frasi "classiche" di rivendicazione di attentati (*"il vostro sangue è forse diverso dal nostro?*", o frasi di contenuto analogo).

Oggi la possibilità di utilizzo delle nuove tecniche di IA ed in particolar modo di quelle generative consente ancor di più di amplificare tali discorsi e di raggiungere un numero indeterminato di persone[126].

Tale illecito utilizzo impone la necessità di individuare l'esistenza di obblighi normativi a carico delle piattaforme digitali finalizzate a rimuovere tali contenuti illeciti allorquando i limiti della libertà di espressione[127] siano stati oltrepassati.

§§§

1.1. Policy aziendali delle piattaforme digitali e orientamenti algoritmici verso fonti affidabili.

Come visto nel capitolo secondo[128] l'utilizzo benevolo dell'IA può giocare un ruolo strategico principale per prevenire il fenomeno della radicalizzazione intesa come il processo attraverso cui si diffonde una visione integralista dell'Islam

[126] CASIERE, *A brave new (digital) world. Tra terrorismo e autoritarismo digitale,* in COSCIENZA E LIBERTÀ, 2024, 67.
CASIERE, *Intelligenza artificiale, terrorismo e responsabilità delle piattaforme digitali. Tra Stati Uniti e Unione Europea,* in COSCIENZA E LIBERTÀ, 2024, 68. Council of Europe, Cyberterrorism: the use of Internet for terrorist purposes, Strasburg 2007

[127] C. BASSU *Istigazione all'odio, terrorismo e sicurezza nell'era digitale: c'è un limite alla libertà di espressione?* in Diritto Pubblico Europeo, 2019

[128] Cfr. *supra* cap. secondo, par.3.1.

spingendo l'individuo verso una adesione totalizzante ed ideologicamente estremista.

Ancor prima di esaminare gli obblighi normativi, deve essere evidenziato come le piattaforme digitali nelle loro policy aziendali stiano sempre di più prevedendo l'utilizzo di algoritmi che operano on line al fine di prevenire e contrastare il fenomeno del terrorismo sotto due rilevanti profili:

- intercettando gli utenti ritenuti più vulnerabili al discorso *jihadista;*
- identificando e rimovendo i contenuti radicalizzanti.

Nel primo caso, ciò avviene tramite il già descritto *redirect method,* un sistema che consente di addestrare l'IA in modo da riconoscere parole ed espressioni caratteristiche della propaganda estremista e poi reindirizzare l'utente verso contenuti di contro-narrativa, riducendo così il rischio di adesione a ideologie violente. Si tratta di una metodologia nata da Jigsaw (Google) insieme a Moonshot e che utilizza annunci mirati per mostrare contenuti alternativi e critici a chi cerca online materiale legato a estremismo/odio, così da "reindirizzare" l'attenzione verso messaggi che smontano le narrazioni violente. È stata integrata anche su YouTube e catalogata tra le pratiche ispiratrici dalla rete europea RAN[129] di esperti e pratici che collaborano per ridurre i rischi di radicalizzazione e scambiarsi strumenti utili, mantenendo un approccio centrato sui diritti fondamentali. Allo stato si tratta di una tecnica sperimentale che si è dimostrata efficace verso chi cercava contenuti jihadisti, ma che necessita di interventi più rigorosi per misurare cambiamenti di atteggiamenti o comportamenti.

[129] La rete RAN una community di professionisti ("practitioners") che lavorano direttamente sul campo nella prevenzione e contrasto della radicalizzazione violenta. Riunisce operatori che vanno oltre il livello politico-istituzionale: insegnanti, poliziotti, operatori sociali, ricercatori, psicologi, leader religiosi, ONG. Funziona come piattaforma di scambio di esperienze e buone pratiche tra gli Stati membri.

Con riferimento al secondo aspetto, l'IA può essere utilizzata per la identificazione e la rimozione dei contenuti radicalizzanti, tramite tecniche di:

- *image matching,* mediante il quale il sistema di IA impara a riconoscere contenuti ritenuti pericolosi;
- *terrorist clustering,* uno strumento di IA che segnala profili collegati a soggetti precedentemente esclusi dalle piattaforme social per propaganda terroristica;
- *repeat offenders algorithms,* ossia sistemi di IA progettati per individuare nuovi *account* riconducibili a utenti già rimossi dalle piattaforme social per contenuti illeciti, attraverso l'analisi di parametri di identificazione indiretti.

§§§

1.2. Il contrasto della diffusione di contenuti terroristici *online* in ambito normativo unionale e interno.

In ambito unionale, la disciplina in materia ha conosciuto un'evoluzione significativa attraverso l'adozione di due atti normativi di rilievo.

Da un lato, la direttiva (UE) 2017/541 ha imposto agli Stati membri l'obbligo di garantire la tempestiva rimozione dei contenuti *online* che istigano alla commissione dei reati di terrorismo, rafforzando così la dimensione preventiva del contrasto al fenomeno[130].

Dall'altro lato, il regolamento (UE) 2021/784, riguardante la lotta alla diffusione di contenuti terroristici on line, nella sua formulazione originaria, prevedeva la possibilità di procedere alla rimozione proattiva dei contenuti tramite strumenti automatizzati di intelligenza artificiale, anche in assenza di un previo ordine di

[130] SANTINI," *Una prima lettura della direttiva 2017/541 sulla lotta contro il terrorismo che sostituisce la decisione quadro Gai 2002",* in Diritto penale contemporaneo 2017, fasc.4 luglio 2017

rimozione da parte di un'autorità pubblica e senza la necessaria supervisione umana.

Una simile impostazione ha sollevato diffuse e motivate perplessità in merito alla compatibilità del sistema con le garanzie fondamentali della persona, in particolare con il diritto alla libertà di espressione.

In risposta a tali rilievi critici, nel corso dell'iter di approvazione il testo del regolamento è stato modificato prevedendo l'obbligo di un controllo umano effettivo sull'operato dell'algoritmo e subordinando la rimozione dei contenuti all'adozione di un ordine formale da parte di una autorità competente, cui spetta la valutazione della legittimità e proporzionalità dell'intervento censorio.

L'Italia con il d.lgs. 24 luglio 2023, n. l07, ha adeguato l'ordinamento giuridico nazionale alle disposizioni del Regolamento (UE) 2021/784 con una disciplina dedicata al contrasto della diffusione di contenuti terroristici *online.*

L'ordinamento nazionale affida all'ufficio del pubblico ministero competente il potere di emanare, con decreto motivato, un ordine di rimozione dei contenuti ritenuti terroristici nei confronti dei servizi di *hosting.* Gli hosting provider - nel cui ambito rientrano i social network, le piattaforme video e più in generale ogni servizio on line nel quali gli utenti possono caricare e condividere contenuti - possono essere equiparati a degli intermediari centrali che non producono i contenuti ma li ospitano e li mettono a disposizione. In questa posizione di centralità gli hosting provider hanno sempre di più mutato il loro originario ruolo neutrale subiscono a causa di una duplice pressione: da un lato quella degli utenti, che caricano e fruiscono dei contenuti con rischi di abuso, hate speech e disinformazione; dall'altro quella delle autorità pubbliche, che chiedono alle piattaforme di vigilare, segnalare o rimuovere i contenuti illeciti (es. terroristici, violenti, illegali). In tal modo, la normativa in materia sempre di più ha trasformato il loro originario ruolo da soggetti tecnici a figure con un ruolo

quasi regolatorio avendo il potere di decidere cosa può restare online e cosa deve essere cancellato[131].

In caso di ordine di rimozione transfrontaliero, l'ordine di rimozione è emanato dal giudice per le indagini preliminari presso il tribunale del capoluogo del distretto in cui il prestatore di servizi di *hosting* ha lo stabilimento principale o in cui il rappresentante legale del prestatore di servizi di *hosting* risiede o è stabilito.

Il provvedimento è portato a conoscenza dei destinatari. I prestatori di servizi di *hosting* che hanno ricevuto l'ordine di rimozione e i fornitori dei contenuti che, in conseguenza dell'ordine, sono stati rimossi o resi inaccessibili, nei dieci giorni successivi alla conoscenza del provvedimento, possono presentare opposizione innanzi al giudice per le indagini preliminari, che provvede con ordinanza in camera di consiglio.

Infine, gli artt. 6 e 7 del Decreto in questione contengono la disciplina sanzionatoria, sia amministrativa che penale, al fine di rendere effettiva l'osservanza delle norme che presidiano gli adempimenti e che prescrivono la doverosità (e la rilevanza) delle *policies* antiterrorismo in capo ai prestatori di servizi di *hosting.*

In particolare, dal punto di vista amministrativo, si prevedono tre distinte tipologie di sanzioni ed in particolare:

- salvo che il fatto costituisca reato, è soggetto alla sanzione amministrativa pecuniaria da 25.000 a 100.000 euro, il prestatore di servizi di *hosting* che: non informa tempestivamente, (...), l'autorità che ha emesso l'ordine di rimozione dell'avvenuta esecuzione dell'ordine, indicandone in particolare la data e l'ora; rimuove i contenuti terroristici o disabilita l'accesso ai contenuti

[131] A. VICINANZA, *La responsabilità delle piattaforme digitali nei confronti dei contenuti illegali: dal caso Telegram al Digital Services Act,* in Quaderni aisdue; PIRAINO, *Spunti per una rilettura della disciplina giuridica degli internet service providers*, in AIDA, 2017, 498 s; ORTOLANI, *The Resolution of Content Moderation Disputes under the Digital Services Act*, in Giust. consensuale,
2022, 539

terroristici ai sensi dell'articolo 3, paragrafo 3, del regolamento, omettendo di adottare le misure necessarie per ripristinare i contenuti o riabilitare l'accesso agli stessi, in conformità dell'articolo 4, paragrafo 7, del regolamento; dopo aver ricevuto una decisione emessa dall'autorità competente (...), omette di ripristinare immediatamente i contenuti o l'accesso agli stessi, fatta salva la possibilità di applicare le proprie condizioni contrattuali conformemente al diritto del- l'Unione e nazionale; nella conservazione dei contenuti terroristici rimossi o il cui accesso è stato disabilitato, ovvero nella conservazione dei relativi dati, non osserva le disposizioni di cui all'articolo 6 del regolamento; non rispetta gli obblighi di trasparenza di cui all'articolo 7 del regolamento; non predispone il meccanismo di reclamo di cui all'articolo 10 del regolamento o, nell'esame, nella decisione e nella gestione dei reclami, non rispetta le disposizioni di cui al paragrafo 2 del medesimo articolo 10; fuori dei casi di cui all'articolo 11, paragrafo 3, del regolamento, omette di comunicare al fornitore di contenuti le informazioni di cui ai paragrafi 1 e 2 del medesimo articolo 11; omette di informare l'organo del Ministero dell'interno per la sicurezza e la regolarità dei servizi di telecomunicazione e la competente Direzione Generale del Ministero delle imprese e del *Made in ltaly* della designazione del rappresentante legale, comunicando la relativa accettazione, o di rendere pubbliche le informazioni relative al rappresentante legale designato;

- salvo che il fatto costituisca reato è soggetto alla sanzione amministrativa pecuniaria da 50.000 a 200.000 euro, il prestatore di servizi di *hosting* esposto a contenuti terroristici che: non include nelle sue condizioni contrattuali o non applica disposizioni volte a contrastare

l'uso improprio dei suoi servizi per la diffusione al pubblico di contenuti terroristici; fuori dei casi di cui alla lettera a), non osserva taluno de- gli obblighi di condotta di cui all'articolo 5, paragrafo 1, del regolamento; adotta misure specifiche prive di taluno dei requisiti di cui all'articolo 5, paragrafo 3, del regolamento; dopo aver ricevuto una decisione di cui all'articolo 5, paragrafo 4 o 6, del regolamento, omette di comunicare all'organo del Ministero dell'interno per la sicurezza e la regolarità del servizio di telecomunicazione nei tre mesi successivi al ricevimento della decisione o ad una delle successive cadenze annuali, le misure specifiche che ha adottato e che intende adottare per conformarsi alle disposizioni di cui ai paragrafi due e tre del medesimo articolo 5;

- salvo che il fatto costituisca reato, è soggetto alla sanzione amministrativa pecuniaria da 75.000 a 300.000 curo, il prestatore di servizi di *hosting* esposto a contenuti terroristici che: omette di adottare misure specifiche per proteggere i propri servizi dalla diffusione al pubblico di contenuti terroristici ai sensi dell'articolo 5, paragrafo 2, del regolamento; dopo aver ricevuto una decisione di cui all'articolo 5, paragrafo 6, omette di adottare le misure imposte dalla decisione per garantire il rispetto delle disposizioni di cui ai paragrafi 2 e 3 del medesimo articolo 5».

Quanto alle sanzioni penali, il Decreto introduce le tre seguenti fattispecie e una causa di non punibilità per il rappresentante legale che segnali tempestivamente le carenze *dell'hosting* che rappresenta in Europa. In particolare, l'art. 7 stabilisce che:

- salvo che il fatto costituisca più grave reato, è punito con l'arresto fino a sei mesi o con l'ammenda da 100.000 a 400.000 euro il prestatore di servizi di *hosting* che: (...), omette di designare o istituire un punto di contatto per la ricezione degli ordini di rimozione in via telematica e per

l'immediata esecuzione dei medesimi (...), oppure omette di rendere disponibili al pubblico le informazioni relative al punto di contatto designato o istituito; non avendo lo stabilimento principale nell'Unione europea, omette di designare, per iscritto, una persona fisica o giuridica quale suo rappresentante legale nell'Unione ai fini del ricevimento, dell'attuazione e dell'esecuzione degli ordini di rimozione e delle decisioni emesse dalle autorità competenti, oppure designa un rappresentante legale che non risiede o non è stabilito in uno degli Stati membri in cui il prestatore di servizi di hosting offre i propri servizi, oppure omette di conferire al rappresentante legale i poteri e le risorse necessari per ottemperare agli ordini di esecuzione e per cooperare con le autorità competenti;

- salvo che il fatto costituisca più grave reato sono puniti con l'arresto fino a sei mesi e con l'ammenda da 100.000 a 400.000 curo il prestatore di servizi di *hosting* e il rappresentante legale designato (...) che: omettono di rimuovere *i* contenuti terroristici entro un'o- ra dal ricevimento dell'ordine di rimozione o di disabilitare l'accesso ad essi entro il medesimo termine; nel caso di cui all'articolo 11, paragrafo 3, del regola- mento, forniscono informazioni riguardanti la rimozione o la disabilitazione dell'accesso a contenuti terroristici; nel caso di cui all'articolo 14, paragrafo 5*,* del regolamento, non informano immediatamente della presenza dei con- tenuti terroristici l'autorità giudiziaria o altra autorità che a quella abbia l'obbligo di riferire;
- salvo che il fatto costituisca più grave reato, quando l'omissione di cui al comma 2, lettera a), è sistematica o persistente, il prestatore di servizi di *hosting* e il rappresentante legale di cui all'articolo 17 del regolamento sono puniti con l'arresto fino a un anno e con l'ammenda da euro 250.000 sino ad euro 1.000.000 o,

laddove superiore, sino ad un importo pari al 4 per cento del fatturato realizzato a livello mondiale dal prestatore di servizi di hosting nell'ultimo esercizio chiuso anteriormente all'accertamento della violazione.

Nei casi in questione può essere disposto il sequestro preventivo, quando il prestatore di servizi di *hosting,* nei quindici giorni successivi all'accertamento e alla contestazione delle violazioni, non provvede agli adempimenti omessi. Infatti, in questi casi l'autorità giudiziaria può, disporre l'interdizione dell'accesso al dominio internet nelle forme e con le modalità di cui all'articolo 321 del codice di procedura penale.

Infine, quanto alla causa di non punibilità, l'art. 7 in questione stabilisce che le descritte sanzioni penali non si applicano al rappresentante legale di cui all'articolo 17 del regolamento che, entro quindici giorni dalla sua designazione, comunica all'organo del Ministero dell'interno per la sicurezza e la regolarità dei servizi di telecomunicazione e alla competente Direzione Generale del Ministero delle imprese e del *Made in Jtaly* di non disporre dei poteri e delle risorse necessari al corretto e integrale adempi- mento dei suoi compiti ai sensi del medesimo articolo 17.

Anche altre legislazioni in ambito europeo, come ad esempio quella spagnola, sulla specifica materia del terrorismo hanno introdotti disposizioni analoghe a quella italina.

§

1.3. Il contrasto della diffusione di contenuti terroristici *online* negli USA: executive pressure e voluntary cooperation

Negli USA, non esiste una legge federale unica che obblighi le piattaforme a rimuovere i contenuti terroristici entro tempi certi. L'approccio è più frammentato e si basa fondamentalmente su "executive pressure e voluntary cooperation "in virtù del quale il

governo USA ha spesso spinto le big tech a collaborare (es. con il Global Internet Forum to Counter Terrorism – GIFCT), ma senza un obbligo normativo come quello europeo. In virtù di tale meccanismo i contenuti pericolosi sono stati originariamente espunti dalle principali piattaforme *online* tramite spontanee pratiche cooperative avviate nel 2015 da Meta, Microsoft, X e YouTube. Dette piattaforme, in sinergia con le autorità pubbliche, hanno enucleato nei "termini di servizio" le primordiali *policies* antiterrorismo e le modalità di collaborazione con le autorità.

Ovviamente tanto in ambito penale, con il Material Support Statutes (18 U.S.C. §§ 2339A e 2339B) che vieta di fornire "supporto materiale" a organizzazioni terroristiche, anche tramite servizi digitali, che in sede civile con l'Anti-Terrorism Act (ATA) (18 U.S.C. § 2333) le vittime di atti terroristici internazionali possono citare in giudizio, davanti ai tribunali federali USA, chiunque abbia fornito consapevolmente assistenza materiale a gruppi terroristici.

Gli atti in questione non impongono obblighi diretti di monitoraggio o censura preventiva ma consentono di ipotizzare a carico delle piattaforme (come Facebook, Twitter/X, YouTube) ipotesi di responsabilità se chi agisce è in grado di dimostrare che abbiano "saputo o dovuto sapere" che i loro servizi fornivano un sostegno materiale a un'organizzazione terroristica designata. Si tratta di temi che meglio verranno esplicitati nei paragrafi successivi esaminando i casi giudiziari affrontati dalla giurisprudenza statunitense che costituiscono i *leading case* in materia.

§§§

2. Limiti alla pubblicazione ed individuazione di un (difficile) punto di equilibrio tra tutela della sicurezza pubblica e salvaguardia dei diritti individuali

Strettamente collegata all'attività di rimozione e di contrasto dei contenuti terroristici dalle piattaforme digitali, è la tematica inerente alla individuazione di limiti alla libertà di espressione e alla relativa disciplina nell'era digitale.

Sul punto, tradizionalmente il mondo del costituzionalismo democratico si divide tra il modello europeo, che ammette limitazioni al diritto di manifestazione del pensiero legittimate dalla salvaguardia dei valori democratici cui si ispira l'ordinamento e l'approccio statunitense, che fa perno sulla centralità della *freedom of speech* nell'impianto costituzionale, sancita chiaramente dal primo emendamento della Costituzione[132].

Con particolare riferimento all'ambito europeo, la CEDU ammette espressamente limitazioni alla libertà di espressione finalizzate a salvaguardare i principi della democrazia, prevedendo in generale che l'esercizio di una libertà trovi un limite invalicabile nella garanzia del diritto altrui che non può subire una eccessiva, indebita compressione. Tali affermazioni di principio costituiscono il presupposto legittimante di quegli interventi normativi che prevedono obblighi per le piattaforme digitali e più in generale sanzioni per chi si renda responsabile di incitamento all'odio o alla violenza[133].

---

[132] DE GREGORIO, *Digital Constitutionalism in Europe. Reframing Rights and Powers in the Algorithmic Society*, Oxford University Press, 2022; GELLERT–JANSSEN, *The Impact of Artificial Intelligence on European Tort Law: Taking Stock of an Ongoing Process*, in The Future of European Private Law, a cura di Janssen–Lehmann– Schulze, Nomos/Hart, 2023, 175 e 193

[133] Cfr. ex multis, O. POLLICINO e G. DE GREGORIO, *Hate speech, una prospettiva di diritto costituzionale comparato*, in Giornale di Diritto amministrativo, 4, 2019, pp. 421-436; I. SPIGNO, Discorsi d'odio: modelli costituzionali a confronto, Milano, Giuffré, 2018; O. POLLICINO, G. PITRUZZELLA e S. QUINTARELLI, *Parole e potere: libertà di espressione, hate speech e fake news*, Milano, Egea, 2017; F. DI TANO, *Hate speech online: scenari, prospettive e criticità giuridiche del fenomeno*, in Ciberspazio e diritto, 51 (2/3), 2014, pp.

Al riguardo la materia in questione risulta disciplinata da un intervento regolatorio di carattere generale costituito dal citato Digital service Act meglio conosciuto come DSA regolamento UE 2022/ 2065 che si concentra sulla regolamentazione delle piattaforme online[134], stabilendo regole chiare per garantire la sicurezza e la trasparenza delle attività online. A tale normativa deve aggiungersi il Digital market Act DMA Regolamento del 14 settembre 2022 che stabilisce regole per prevenire comportamenti anticoncorrenziali da parte delle grandi piatteforme che svolgono un ruolo cruciale nei mercati digitali.

Come abbiamo già potuto constatare si tratta di una normativa che risulta interconnessa con quella dell'AI ACT e del GDPR sul trattamento dei dati personali.

Invece, negli Stati Uniti, la libertà di manifestazione del pensiero è piena, intangibile e anche i contenuti più radicali, provocatori, scabrosi sono ammessi a circolare nel libero mercato delle idee, senza limitazioni imposte all'oggetto dei messaggi o alle modalità in cui sono espressi. Per tali ragioni non esiste una normativa di carattere generale relativa alla responsabilizzazione degli intermediari digitali, perché si potrebbe tradurre in una attività di censura collaterale. La normativa primaria federale di riferimento è il *Communication Decency Act* che, alla *section* 230 esclude che chi fornisce servizi interattivi digitali possa essere dichiarato responsabile di contenuti prodotti e diffusi da altri[135].

Per meglio comprendere la diversa natura degli approcci normativi procediamo con ordine.

413-452; L. SCAFFARDI, *Oltre i confini della libertà di espressione: l'istigazione all'odio razziale*, Padova, Cedam, 2009; J. WALDRON, The Harm in hate speech, Cambridge, Harvard University Press, 2012; M. ROSENFELD, *Hate speech in Constitutional Jurisprudence.*
In the Content and Context of Hate Speech, Cambridge University Press, 2012; V. ZENO ZENCOVICH, *Freedom of expression: A Critical Comparative and Analysis, London*, Routledge-Cavendish, 2008.

[134] ORTOLANI, *The Resolution of Content Moderation Disputes under the Digital Services Act*, in Giust. consensuale, 2022, 539

[135] KOSSEFF, *A User's Guide to Section 230, and a Legislator's Guide to Amending It (or Not)*, in 27 Berkeley Tech. LJ, 2022, 757; CANDEUB, *Reading Section 230 as Written*, in 1 J. Free Speech L., 2021, 139.

§§§

2.1. Gli obblighi normativi delle piattaforme digitali in Europa: il Digital service act (DSA)

Al pari del regolamento AI ACT, in ambito europeo il Digital service Act - DSA regolamento UE 2022/ 2065 costituisce un intervento regolatorio di carattere generale relativamente alla disciplina delle piattaforme online, stabilendo regole chiare per garantire la sicurezza e la trasparenza delle attività online[136].

In primo luogo, la normativa in questione impone a tutti gli intermediari (come i social network, siti web con contenuti generati dagli utenti, motori di ricerca, ecc.) di dotarsi di idonee procedure di moderazione trasparenti e di sistemi di notifica e rimozione (notice-and-takedown) per i contenuti illegali.

In secondo luogo, in caso di pubblicazione di contenuti illeciti, dai discorsi di odio in materia di terrorismo fino alla diffamazione alla violazione della privacy e alla pornografia non consensuale, la piattaforma una volta ricevuta una apposita segnalazione ha l'obbligo di rimuoverli tempestivamente, secondo le procedure previste dal DSA. Sul punto, un'ulteriore tutela fornita all'utente del servizio è data dall'obbligo di trusted flaggers e di meccanismi di reclamo, che possono facilitare la segnalazione e rimozione di deepfake nocivi. In sintesi, se l'AI Act disciplina la creazione e

[136] B. CALDERINI *Digital Service Act: cos'è e cosa prevede la legge europea sui servizi digitali,* in Agenza Digitale, 16 aprile 2025; M.FOTI Regolamentazione digitale: il Digital Service Act e le piattaforme online in Altalex, 11 luglio 2024; Guida pratica al Digital Services Act, di TRAINA CHIARINI RICCARDO, POLLICINO ORESTE, PAOLUCCI FEDERICA, eBook, Ed. UTET GIURIDICA, 2024. *Responsabilità degli Internet Service Provider - Obblighi specifici delle piattaforme online - Diritto degli utenti e dei titolari di diritti.*

diffusione di deepfake imponendo obblighi a produttori e utilizzatori, il DSA ne disciplina la circolazione sulle piattaforme, garantendo che i servizi digitali reagiscano in modo trasparente e rapido ai contenuti manipolati quando violano la legge o i diritti altrui.

Inoltre, il regolamento impone anche che i provider pubblichino chiaramente le proprie policy di moderazione (ad es. nei termini di servizio), includendo come vengono trattati i materiali manipolati o ingannevoli. In particolare, le piattaforme molto grandi come ad esempio Facebook, TikTok, YouTube, e tutte quelle considerate VLOPs hanno obblighi aggiuntivi di analisi e mitigazione dei rischi sistemici, tra cui quelli legati alla necessità di evitare che possano verificarsi fenomeni disinformazione online[137].

Tali obblighi aggiuntivi maggiormente si accentuano nei periodi elettorali durante i quali il rischio dei deepfake si trasforma in potenziale minaccia alla libertà di voto degli individui.

Al riguardo, il DSA richiede ai grandi operatori di valutare ed attenuare il rischio di manipolazione elettorale tramite IA, ad esempio rafforzando i sistemi di rilevazione di media falsificati e collaborando a codici di condotta sul contrasto alla disinformazione. In pratica, anche se il DSA non impone un obbligo legale esplicito di etichettare ogni deepfake sulle piattaforme, spinge i big tech ad adottare misure (anche volontarie) come etichettature o watermark per identificare i

[137] HOLMES, *Facebook just banned deepfakes, but the policy has loopholes – and a widely circulated deepfake of Mark Zuckerberg is allowed to stay up,* (Jan. 7, 2020) Insider: https://www.businessinsider.com/facebook-just-banneddeepfakes-but the-policy-has-loopholes-2020-1.; Ci sono numerosi studi di modelli, algoritmi e sistemi volti a riconoscere i deep fakes. Nel 2020, Facebook ha elaborato una nuova policy riguardante i media manipolati e i deep fakes, che prevede la rimozione dal social dei deep fakes creati in modo di far credere ad una persona media che il protagonista del video e/o dell'immagine abbia detto o fatto cose che non ha mai detto o fatto. Dalla nuova policy sono esclusi immagini/video creati come parodia e satira. Simili policy sui media manipolati sono state introdotte da Twitter e da YouTube.

contenuti generati da IA. La Commissione UE ha infatti sollecitato le maggiori piattaforme a rendere riconoscibile l'origine IA dei contenuti in vista delle elezioni europee 2024. Complessivamente, quindi, il DSA si è scelto un approccio regolatorio centralizzato e vincolante, in virtù del quale la piattaforma non viene ritenuta responsabile in automatico, ma se viene notificata del contenuto illecito deve agire rapidamente per rimuoverlo o bloccarne l'accesso[138].

§§§

2.2. Gli obblighi normativi delle piattaforme digitali negli USA: il titolo 47 U.S. Code, § 230 del 1996 del Communications Decency Act

Come detto in precedenza, negli Stati Uniti la normativa primaria di riferimento è costituita dai principi stabiliti nella Section 230 del Communications Decency Act divenuta celebre come «*The Twenty-Six Words That Created the Internet*»[139]. Tale atto normativo è stato approvato l'8 febbraio del 1996 nel momento in cui internet stava esplodendo ed è stato preceduto da alcune decisioni giudiziarie nelle quali veniva affermava la necessità di dettare una disciplina che adeguatamente proteggesse gli utenti da possibili abusi e allo stesso tempo garantisse la libertà di espressione.

[138] BACHELET, *Il rafforzamento del contrasto agli abusi di posizione "non dominante" delle piattaforme digitali* Ballardini B., Isis®. Il marketing dell'apocalisse, Baldini&Castoldi, Milano 2015.
[139] KOSSEF, *The Twenty-Six Words That Created the Internet*, Cornell University Press, 2019.

Infatti, negli anni '90, il *Communication Decency Act* aveva inizialmente rischiato di diventare un bavaglio per il giovanissimo Internet, ancora largamente privo di qualsiasi genere di argine normativo. Il legislatore statunitense, infatti, aveva sentito la necessità di colmare tale lacuna estendendo la disciplina sulle comunicazioni "oscene e indecenti" rivolte a minori di 18 anni e sul divieto di distribuzione di materiali "palesemente offensivi" che fossero disponibili anch'essi per minori di 18 anni.

In origine, la preoccupazione dominante rispetto ai contenuti disponibili su Internet riguardava l'accesso incontrollato alla pornografia da parte dei minori. Nelle prime fasi dell'iter legislativo, infatti, non erano sorte particolari attenzioni nei confronti delle possibili ripercussioni sugli operatori del web[140].

Al contrario, l'origine dell'emendamento che ha introdotto la Sezione 230 va rintracciata in due decisioni assunte da giudici di New York in quegli anni.

Nella prima decisione del 1991, in *Cubby, Inc., v. CompuServe*, venne affermato che CompuServe non poteva essere ritenuta responsabile per i commenti diffamatori pubblicati in uno dei forum di detta società, in quanto non esaminava i contenuti prima che venissero pubblicati, ma si limitava ad ospitarli sulla propria piattaforma.

Nella seconda decisione del 1995 in *Stratton Oakmont, Inc., v. Prodigy Services Co.* la conclusione fu, invece, diversa: poichè Prodigy svolgeva attività di moderazione sulle sue bacheche online e cancellava determinati messaggi per "offensività e cattivo gusto", venne affermato che potesse essere ritenuta responsabile per i contenuti pubblicati sulla sua piattaforma.

---

[140] YOO, *The first emendment, common carriers, and public accommodations: net neutrality, digital platforms, and privacy*, in 1 J. Free Speech L., 2021, 463.; SITARAMAN, *Deplatforming*, in Yale Law J., 2023, 497 s., spec. 553 s., sia pure evidenziando che «the American tradition has not been one of either an absolute duty to serve or an absolute right to exclude. Instead, the American tradition has been one of reasonable deplatforming».

In tale contesto giurisprudenziale è così accaduto che due rappresentanti repubblicani al Congresso, Ron Wyden e Chris Cox, proposero dunque un emendamento per escludere la responsabilità dei provider per i contenuti pubblicati dagli utenti, anche nel caso in cui venisse svolta attività di moderazione sulla piattaforma. In tale emendamento, il cui accoglimento ha portato alla attuale formulazione del comma 1 della sezione 230, veniva espressamente affermato che: "*No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider*."[141]

Secondo alcuni autori statunitensi su queste 26 parole, che nella interpretazione comune hanno introdotto il principio di immunità delle piattaforme web, sono state poste le basi per lo sviluppo della mastodontica industria americana del web[142].

Accanto al menzionato principio di immunità, la Sezione 230 prevede anche un ulteriore disposizione meglio nota come quella del "good clause" in virtù della quale le piattaforme non perdono la descritta immunità se moderano attivamente i contenuti ("good faith moderation").

In particolare, il testo in questione stabilisce che: "*Civil liability*

"No provider or user of an interactive computer service shall be held liable on account of"

(A) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or

[141] "*Nessun fornitore o utente di un servizio informatico interattivo è trattato come l'editore o il portavoce di informazioni fornite da un altro fornitore di contenuti informativi*"

[142]https://www.cornellpress.cornell.edu/book/978150714412/the-twenty-six-words-that-created-the-internet

(B) any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph" [143]

L'approvazione di tali disposizioni ha consentito di limitare la responsabilità delle piattaforme digitali nei confronti dei contenuti degli utenti, sia in presenza che in assenza di una attività di moderazione del dibattito da parte delle stesse piattaforme.

Ciò nonostante, in seguito all'approvazione del *Communications Decency Act*, numerose proteste da parte di associazioni per i diritti civili hanno messo in dubbio la costituzionalità del resto della normativa e dei divieti in essa contenuti, considerati in contrasto con il Primo Emendamento (che protegge la libertà di pensiero e di espressione).

Nel 1997, il caso *Reno v. American Civil Liberties Union* determinava una pronuncia della Corte Suprema con la quale veniva dichiarata l'incostituzionalità della normativa in questione nella parte in cui limitava la pubblicazione di contenuti "osceni e indecenti", sul presupposto che in tale ambito potesse rientrare anche la pubblicazione di materiali inerenti la salute, come ad esempio le tecniche per la prevenzione della diffusione dell'AIDS.

Tale interpretazione ha permesso alle piattaforme di Internet di crescere e prosperare più facilmente senza essere soffocate da costose verifiche dei contenuti e senza dover limitare la libertà di espressione online.

Tuttavia, negli ultimi anni, si sono succedute numerose ed aspre critiche verso la Sezione 230, poiché consente alle piattaforme di

[143] "Nessun fornitore o utente di un servizio interattivo per computer sarà ritenuto responsabile per:
(A) qualsiasi azione volontariamente intrapresa in buona fede per limitare l'accesso o la disponibilità di materiale che il fornitore o l'utente ritenga osceno, lascivo, sudicio, eccessivamente violento, molesto o altrimenti discutibile, se tale materiale è protetto costituzionalmente; o
(B) qualsiasi azione intrapresa per consentire o rendere disponibili ai fornitori di contenuti informativi o ad altri i mezzi tecnici per limitare l'accesso al materiale descritto nel paragrafo"

tollerare contenuti diffamatori, disinformazione e incitamenti alla violenza. Inoltre, diversi autori sostengono che se da un lato le piattaforme garantiscono il diritto di libera espressione, dall'altro non fanno ancora abbastanza per rimuovere i contenuti offensivi e proteggere gli utenti.

Sul punto, un recente saggio[144] ha affrontato il tema dell'utilizzo di Facebook come canale di traffico illecito di reperti archeologici provenienti dal Medio Oriente sottolineando come la piattaforma, grazie alla sua capillarità e facilità d'uso, venisse sfruttata da trafficanti e gruppi criminali (inclusi gruppi legati all'ISIS) per vendere e acquistare reperti archeologici rubati. Nel saggio in questione gli autori denunciano la mancanza di controlli efficaci da parte della piattaforma sottolineando la necessità di maggiore responsabilità da parte delle big tech, di una cooperazione internazionale più forte e di strumenti investigativi dedicati al contrasto di questi fenomeni online. Nel giugno 2020, dopo un'intensa pressione da parte di gruppi come il ATHAR Project, l'Antiquities Coalition e i media (es. BBC, UNESCO), Facebook (oggi Meta) ha annunciato una svolta nelle sue politiche vietando esplicitamente la vendita, l'acquisto e lo scambio di tutti i "historical artifacts" (reperti storici) su Facebook e Instagram. La policy include manufatti di significativo valore storico, culturale o scientifico, compresi manoscritti antichi, monete, sigilli e lapidi.

Per ovviare a tali problematiche, negli ultimi anni il Congresso ha discusso vari disegni di legge per limitare l'immunità della Section 230 e contestualmente introdurre maggiori obblighi di trasparenza e di rendicontazione per le piattaforme, al fine di ulteriormente responsabilizzare le piattaforme in questione..

Finora, però, nessuna riforma organica è stata approvata.

---

[144] A. AL AZM e K. A. PAUL, *How Facebook Made It Easier Than Ever to Traffic Middle Eastern Antiquities*, in World Politics Review, August 2018, reperibile al sito worldpoliticsreview.com/insights/25532/howfacebook-made-it-easier-than-ever to traffic-middleeastern-antiquities

In conclusione, negli USA non esiste una normativa generale come il DSA europeo; il sistema è basato sulla Section 230 e su un mosaico di leggi settoriali, integrate da iniziative statali e proposte di riforma non ancora approvate.
Negli USA le piattaforme sono protette dall'immunità e hanno grande discrezionalità nella moderazione. Invece, nell'UE, con il Digital Services Act, si è scelto un approccio regolatorio centralizzato e vincolante, che impone obblighi specifici di trasparenza, rendicontazione e responsabilità, specialmente per le grandi piattaforme.
Tale differenza normativa, dal punto di vista pratico comporta che ad esempio Facebook negli USA ha ampia libertà: non è legalmente responsabile e decide se rimuovere secondo i suoi termini; invece, Facebook nel sistema UE ha obblighi chiari e vincolanti dovendo agire su segnalazione, garantire trasparenza e cooperare con autorità.

§§§§§

3. Tra responsabilità e deresponsabilizzazione: l'evoluzione della giurisprudenza statunitense sulle piattaforme digitali ed i riflessi in ambito europeo.

I diversi approcci normativi e il conseguente dibattito politico sulle responsabilità delle piattaforme digitali si riflettono anche in ambito giudiziario ed in particolare in quello statunitense. In tale ultimo ambito, le pronunce dei giudici hanno posto al centro della questione la tenuta e l'attualità della Sezione 230 del *Communications Decency Act* del 1996. Conseguentemente, per la lora ampia portata di carattere generale, le pronunce che di seguito esamineremo si sono trasformate in riferimenti per l'intera materia, offrendo, in tal modo, imprescindibili spunti per

meglio delineare il ruolo e la responsabilità delle piattaforme digitali.

Sul punto, la trasformazione delle piattaforme in "colossi digitali", come tali detentrici di un potere seppur privato ma di dimensioni rilevanti, ha ulteriormente alimentato il dibattito intorno alla effettiva tenuta dell'interpretazione della *Section* 230 basata, come visto, sul principio di immunità in quanto collegata all'assenza di un obbligo generale di sorveglianza in via preventiva delle informazioni trasmesse o memorizzate da parte delle stesse piattaforme digitali.

In particolare, la tesi del principio di immunità è stata messa in discussione da coloro i quali hanno ritenuto che la stessa si fondasse[145]su un passaggio isolato della disposizione della Sezione 230[146] che, tuttavia, non teneva conto del testo nel suo complesso, del contesto giuridico nel quale è stata approvato e delle intenzioni del legislatore il cui obiettivo, in realtà, era quello di colmare le lacune del *common law* in materia di responsabilità delle società che distribuiscono contenuti di terzi, estendendo a Internet le regole già sviluppate per i distributori di contenuti offline, come le librerie e le edicole.

Secondo tale impostazione, il *Communication Decency Act*, del quale la *Section* 230 è parte, non voleva introdurre una ampia e generale immunità per le piattaforme, ma cercava di incoraggiare i fornitori di servizi di telecomunicazione e di informazione a implementare nuove tecnologie e politiche per bloccare o filtrare il materiale offensivo e al tempo stesso tutelare le piattaforme per lo svolgimento di questa attività nei confronti di azioni giudiziarie.

---

[145] F.BENATTI G.PORTONERA, *La responsabilità di diritto civile degli Internet service providers. Spunti dalla comparazione con la giurisprudenza statunitense* 2024, La nuova giurisprudenza civile commentata, pag.476-485 dove viene dato ampio conto del dibattito in merito alla effettiva sussistenza del principio di immunità.

[146] *no provider or user of an interactive computer service shall be treated as the publisher or speaker of any informa- tion provided by another information content provider*.

In tale contesto di dibattito politico e dottrinale che vede contrapporsi due distinte impostazioni, le scelte della Corte Suprema hanno finito per ulteriormente influenzare il dibattito in materia soprattutto con riferimento a quei casi giudiziari che si proponevano di affermare la responsabilità delle piattaforme digitali per la diffusione di materiali inerenti al terrorismo.

§§§

3.1. Internet liability e prime crepe giudiziarie sul concorso degli algoritmi negli illeciti on line.

Le decisioni giudiziarie sulla asserita responsabilità delle piattaforme digitali per la diffusione di materiali inerenti al terrorismo si inseriscono nel più ampio contesto della "*internet liability*", che, come visto, è tutt'ora al centro di un dibattito acceso che vede come protagonista l'effettiva tenuta della Sezione 230 del *Communications Decency Act* del 1996 che nella interpretazione prevalente ha fornito negli ultimi tre decenni una sorta di immunità alle piattaforme online per il contenuto generato dagli utenti in virtù del quale se un utente pubblica un contenuto offensivo, diffamatorio o comunque illecito su una piattaforma, infatti, questa non ne può essere ritenuta responsabile.
Un primo vaglio della tenuta di tale principio nel caso in cui il ruolo della piattaforma non sia passivo ma attivo attraverso l'utilizzo di appositi algoritmi, si ha nella controversia Force v. Facebook[147].
I fatti oggetto di causa si collocavano tutti tra il 2014 e il 2016 e vedevano coinvolti cinque cittadini americani rimasti uccisi nel

[147] A.BACCIN *Responsabilità penale dell'internet service provider e concorso degli algoritmi negli illeciti online: il caso force v. facebook* in Sistemi Penali n.5/20

corso di svariati attacchi terroristici perpetrati da Hamas, la nota organizzazione islamista palestinese, in territorio israeliano[148].

A seguito dei fatti accaduti, i familiari intentavano una azione legale nei confronti della piattaforma digitale Facebook accusandola di aver fornito supporto ai terroristi tramite la possibilità concessagli di utilizzare la piattaforma per la diffusione del loro messaggio di propaganda e per l'organizzazione degli attacchi.

Alla base della presunta responsabilità della piattaforma vi sarebbe stato il funzionamento dell'algoritmo e la tendenza a creare "*echo-chambers*" favorendo il rinvenimento di contenuti compatibili con le idee o i gusti degli utenti.

Nel 2019, investita della questione, la US Second Circuit Appeals Court statuiva – per la prima volta – che la Sezione 230 proteggeva le piattaforme come Facebook non solo dal punto di vista penalistico ma anche da eventuali azioni civili intentate da vittime di terrorismo.

In seguito al ricorso presentato dai parenti delle vittime, la Corte Suprema confermava la decisione respingendo la richiesta di esaminare la questione nonostante la dissenting opinion del giudice Katzman.

In particolare, il dissenso di Katzman veniva motivato sul presupposto che: "*evidenze sempre maggiori suggeriscono che i provider abbiano progettato i loro algoritmi per guidare gli utenti verso contenuti e persone con cui gli utenti stessi siano d'accordo e che lo abbiano fatto troppo bene, spingendo le anime sensibili sempre più in basso verso percorsi oscuri"*. Tali

[148] Le vittime sono: Yakov Naftali Fraenkel, un teenager rapito dagli uomini di Hamas al ritorno da scuola e successivamente ucciso a colpi di pistola a Gush Etzion; Chaya Zissel Braun, neonata di 3 mesi travolta d un'automobile scagliatasi contro la folla in una stazione a Gerusalemme; Richard Lakin, dapprima colpito con un'arma da fuoco e, successivamente, finito a coltellate durante un'incursione di Hamas in un autobus a Gerusalemme; Taylor Force, studente pugnalato mentre passeggiava sul lungomare di Tel Aviv e Menachem Mendel Rivkin, pugnalato al collo da un esponente di Hamas mentre si recava in un ristorante nei pressi di Gerusalemme, unico sopravvissuto nonostante le gravissime ferite riportate

affermazioni aprivano un primo varco per il riconoscimento della responsabilità delle piattaforme nel caso in cui le stesse avessero avuto un ruolo attivo nelle scelte effettuate dagli utenti.
La questione della responsabilità dell'internet provider veniva riproposta e nuovamente respinta dalla Corte suprema nel caso Malwarebytes, Inc. v. Enigma Software Group USA, LLC. in cui la dissenting opinion del Giudice Thomas riprendeva le argomentazioni svolte dal giudice Katzman in Force v. Facebook Inc.
Al di là dell'esito finale delle vicende appena descritte, con le decisioni, in questione i giudici statunitensi hanno aperto anche in ambito europeo un primo momento di riflessione sull'adeguatezza del quadro normativo vigente rispetto all'evoluzione del contesto tecnologico e alla conseguente necessità di ridefinire la regola di responsabilità di diritto civile (comprensiva sia di quella *ex delicto* che di quella ex *contractu*) cui assoggettare gli Internet service providers (ISP).
Tale momento di riflessione si è ancor di più sviluppato con le nuove richieste di intervento formulate alla Corte suprema statunitense in occasione di altre due vicende processuali: *Taamneh* e *Gonzalez.*

§§§

3.2. La (non) neutralità delle piattaforme digitali: spunti dai casi Google v Gonzalez e Twitter v. Taamneh

Il primo caso riguarda Nohemi Gonzalez, cittadina statunitense, veniva uccisa nell'attacco terroristico del Bataclan di Parigi, nel 2015.
Il giorno seguente, l'ISIS rivendicava la paternità dell'attacco, rilasciando una dichiarazione scritta e un video di YouTube.

Il padre di Gonzalez, quindi, intenteva un'azione legale nei confronti di Google, Twitter e Facebook, sostenendo principalmente che:

- Google avesse aiutato e favorito il terrorismo internazionale consentendo all'ISIS di utilizzare la sua piattaforma, in particolare YouTube, "*per reclutare membri, pianificare attacchi terroristici, lanciare minacce terroristiche, instillare paura e intimidire le popolazioni civili*.";
- che proprio l'utilizzo di algoritmi informatici che suggeriscono contenuti per gli utenti in base alla loro cronologia di visualizzazione aiutasse l'ISIS a diffondere il suo messaggio;
- che il sistema di monetizzazione di Google su YouTube avrebbe anche fatto sì che l'algoritmo valutasse e approvasse contenuti provenienti dall'ISIS, comportando una condivisione di guadagni con soggetti riconducibili all'organizzazione terroristica.

Nei primi due gradi di giudizio la *motion to dismiss* cioè la richiesta di archiviazione del caso da parte di Google veniva accolta, come avvenuto nelle precedenti vicende giudiziarie.

Il secondo caso riguarda Nawras Alassaf rimasto vittima nell'attacco terroristico al Reina di Istanbul nel 2017. Taamneh, uno dei familiari della vittima, intentava azione legale nei confronti di Google, Twitter e Facebook accusandoli di favoreggiamento per non aver adottato misure significative per prevenire l'uso dei loro servizi per scopi di terrorismo.

In questo caso, dopo l'iniziale rigetto in primo grado, la Court of Appeals for the Ninth Circuit ha ribaltato la decisione, ritenendo che vi fosse un collegamento diretto tra la diffusione del messaggio dell'ISIS da parte delle piattaforme social e i danni causati alle vittime degli attacchi.

Entrambi i casi citati sono stati trattati congiuntamente dalla Corte Suprema sebbene nel caso Twitter v. Taamneh abbia incentrato la

motivazione esclusivamente sul tema dell'applicabilità della responsabilità per favoreggiamento di cui all'Anti-Terrorism Act evitando di affrontare in maniera diretta nella decisione Google v. Gonzalez il delicato tema relativo alla perdurante attualità del principio di immunità e della Sezione 230.

In particolare, nel percorso argomentativo della sentenza Twitter v. Taamneh del 18 maggio del 2023[149], i giudici della Corte suprema si confrontano con la rilevanza della consapevolezza, in capo al provider, della presenza di clienti/utenti che utilizzano il servizio fornito per scopi illeciti (ad es. la presenza dell'ISIS su YouTube). A tal proposito, la Corte Suprema si è affidata al concetto di neutralità dell'azione del provider (già utilizzato dalla Corte d'Appello del 9° Circuito rispetto all'operato dell'algoritmo) ed ha evidenziato come non si possa trasformare una "distante inerzia" in consapevole e sostanziale assistenza all'attività terroristica; ha dunque ritenuto insufficiente, di per sé, l'osservazione per cui le piattaforme in esame fanno qualcosa in più che trasmettere informazioni per miliardi di persone (tramite l'analisi delle preferenze degli utenti).

Inoltre, con riferimento specifico a Google e al sistema di monetizzazione di YouTube, i ricorrenti non avrebbero portato evidenze concrete di un contributo notevole fornito all'ISIS o ai suoi appartenenti, né in termini di importo delle somme corrisposte, né rispetto al numero di account e contenuti approvati dalla piattaforma.

Sulla base di tali considerazioni, i giudici hanno così escluso ogni forma di diretta assistenza da parte di Google a favore all'ISIS, né in occasione dell'attacco di Istanbul del 2017, né nelle altre attività di natura terroristica dell'organizzazione.

Nella motivazione della sentenza Gonzalez[150], i giudici statunitensi in maniera pilatesca hanno evitato di addentrarsi in merito al rinnovato dibattito sviluppatosi sulla perdurante

[149] Sentenza Twitter v. Taamneh in www.supremecourt.gov
[150] Sentenza Google v Gonzalez in www.supremecourt.gov

attualità della deresponsabilizzazione delle piattaforme digitali e della attuale tenuta delle Sezione 230. Al riguardo i giudici hanno espressamente affermato di rifiutarsi di affrontare il tema dell'applicazione della Sezione 230 ad una vicenda che "sembra poter rivendicare poche pretese di risarcimento in quanto il ricorso dei querelanti deve essere respinto sulla base della nostra decisione su Twitter".

Secondo quanto riportato dai media statunitensi, in udienza i giudici della Corte Suprema avevano già manifestato notevole perplessità di fronte all'opportunità di decidere sul futuro di Internet, laddove dovrebbe essere il legislatore ad intervenire per determinare una simile svolta: "*Isn't it better to keep it the way it is, for us, and to put the burden on Congress to change that and they can consider the implications and make these predictive judgments?*"

La decisione quasi pilatesca della Corte Suprema nella vicenda Gonzalez ha rinverdito le posizioni di chi, come gli attivisti per i diritti digitali, sostiene che la Sezione 230 deve rimanere parte dell'ordinamento statunitense in quanto espressione della libertà online e della riconosciuta possibilità alle piattaforme online di rimuovere contenuti offensivi senza necessariamente censurare la libertà di espressione[151].

§§§

3.3. Prospettive di oggettivizzazione della responsabilità` della piattaforma.

Come già condivisibilmente evidenziato in dottrina, i descritti orientamenti delle Corti statunitensi che sostengono la neutralità

[151] R.DE VITA *US Supreme Court: Gonzalez v. Google e l'internet liability* in https://www.devita.law 3 giugno 2023

delle piattaforme online e di fatto l'irresponsabilità delle stesse risultano difficilmente condivisibili.

Invero l'oramai noto funzionamento degli algoritmi di analisi delle preferenze degli utenti sono una evidente conferma della assoluta centralità e del ruolo attivo dei provider nella tematica in questione.

Le piattaforme hanno il potere di raggiungere milioni di persone in tutto il mondo, ma da questo potere deriva anche la responsabilità di assicurarsi che il contenuto pubblicato su di esse non danneggi gli utenti o quantomeno di adoperarsi concretamente in tal senso. Anche nel diritto europeo si riscontra ormai una attenzione alle dimensioni quantitative delle piattaforme, soprattutto di quelle particolarmente grandi c.d. VLOPS (*very large online platforms and search engines*, secondo la terminologia impiegata dal *Digital Service Package*) al fine di giustificare l'applicazione di una disciplina più rigorosa rispetto a quella comunemente data.

Seppur in materia differente da quella del terrorismo, anche in Italia recenti vicende assurte agli onori della cronaca hanno riproposto l'attualità del tema in questione[152]

e la conseguente necessità di una maggiore regolamentazione e supervisione delle piattaforme online. Se da un lato l'esperienza americana e la Sezione 230 ha fornito per lungo tempo un'immunità funzionale allo sviluppo di Internet, dall'altro a livello globale permangono le preoccupazioni per la sicurezza degli utenti e la diffusione di contenuti che comportano pericoli tanto online quanto offline.

Nella consapevolezza della difficoltà di individuare un corretto punto di equilibrio, è però necessario, prima e ben più di una pronuncia giurisdizionale, un intervento normativo che affronti il problema con lo sguardo critico dell'attualità e che possa

[152] Foto rubate, chiude il sito sessista. Caccia agli utenti che commentavano, in Corriere della sera del 29 agosto 2025

soddisfare le (spesso) contrapposte esigenze di sicurezza e libertà di espressione.

Se dal punto di vista penalistico risulta assai arduo ipotizzare una responsabilità delle piattaforme digitali alla luce dei principi esposti nel caso Twitter v. Taamneh essendo necessario che la stessa piattaforma abbia volontariamente e consapevolmente tenuto un comportamento a supporto dell'illecito per poter essere configurata una sua responsabilità a conclusioni diverse si può giungere in ambito civilistico.

In tale ambito non sembra irragionevole ipotizzare una ridefinizione della regola della responsabilità civile delle piattaforme digitali facendo ricorso all'istituto della responsabilità oggettiva.

Sul punto in dottrina[153] è stato evidenziato come

la giurisprudenza italiana aveva provato ad utilizzare l'art. 2050 del Codice civile (responsabilità per l'esercizio di attività pericolose) per fondare una responsabilità autonoma delle piattaforme sul presupposto che tale attività presentasse "profili di pericolosità" per l'uso illecito che gli utenti potevano fare dei servizi in questione.

Questo orientamento è stato però smentito dalla direttiva e-commerce (dir. 2000/31/CE, recepita dal d.lgs. 70/2003), che ha introdotto un regime di "irresponsabilità" avendo escluso l'obbligo per le piattaforme di cercare attivamente contenuti illeciti. Rispetto all'art. 2050, ciò equivale all'eliminazione dell'obbligo di adottare "tutte le misure idonee a evitare il danno"

Tuttavia, il tema della responsabilità oggettiva delle piattaforme è tornato d'attualità a seguito dell'entrata in vigore del diritto d'autore online disciplinato dall'art. 102-septies l. 633/1941[154]

[153] F.BENATTI G.PORTONERA, *La responsabilità di diritto civile degli Internet service providers. Spunti dalla comparazione con la giurisprudenza statunitense* 2024, La nuova giurisprudenza civile commentata, pag.476-485 dove viene dato ampio conto del dibattito in merito all'evoluzione in ambito civilistico della responsabilità oggettiva.

[154] L'articolo 102-septies della legge sul diritto d'autore (Legge 633/1941)stabilisce la responsabilità dei prestatori di servizi di condivisione di contenuti online (come le piattaforme online che

(che recepisce l'art. 17 dir. 2019/790/UE) il quale parlando di "massimi sforzi" e di "elevati standard di diligenza professionale", ripropone espressioni che ricordano il meccanismo dell'art. 2050: responsabilizzazione del soggetto, salvo esonero se dimostra di aver fatto tutto il possibile.

Tale evoluzione normativa in tema di diritto d'autore potrebbe essere una efficace base per la predisposizione di un intervento di carattere normativo che estenda i principi delineati nell'art.2050 c.c. anche alla logica della responsabilità oggettiva al settore delle piattaforme digitali in un'ottica di sopportazione dei costi dell'attività di impresa.

§§§§§

4.Conclusioni e proposte

I casi trattati dalla giurisprudenza statunitense dimostrano sempre di più la necessità di individuare anche in ambito normativo un punto di equilibrio tra la libertà di espressione su cui si è fondata la nascita di internet e la protezione degli utenti online.

Indubbiamente, si tratta di obiettivi non semplici da realizzare sia in ambito penalistico che civilistico.

Dal punto di vista penalistico la previsione di obblighi di controllo in capo ai gestori di piattaforme (siano persone fisiche o enti), rischia di confliggere con i principi generali del diritto penale attesa la necessità di valutare attentamente la concreta esigibilità delle misure preventive, non potendosi automaticamente

consentono il caricamento di contenuti da parte degli utenti) per atti non autorizzati di comunicazione al pubblico e messa a disposizione di opere protette. Salvo che non abbiano ottenuto la necessaria autorizzazione ai sensi dell'articolo 102-sexies, le piattaforme sono tenute a dimostrare di aver adottato tutte le misure necessarie per rispettare la legge, considerando vari fattori come la tipologia del servizio, il pubblico e la qualità dei contenuti caricati

configurare una responsabilità da posizione a carico della piattaforma stessa.

Ancora, l'attribuzione a determinati soggetti (piattaforme social o autorità governative) di poteri e obblighi di repressione e persino di prevenzione rispetto al verificarsi di determinati illeciti – anche tramite interruzione delle condotte che condurrebbero al verificarsi dell'evento atteso – sollevano indubbiamente un problema di bilanciamento degli interessi e di tutela di diritti costituzionalmente garantiti, in particolare con la libertà di manifestazione del pensiero, anche negli spazi virtuali.

Il tema diventa delicato soprattutto in ambito di influenza sul mercato politico, e in generale quando si discorre di utilizzo di piattaforme di social media, che vedono ormai il web come strumento di espressione potenziata del proprio pensiero e forma di estrinsecazione pubblica della propria personalità.

La possibilità concessa dall'ordinamento (a un singolo soggetto, agli organi di una piattaforma ovvero ad un'autorità amministrativa) di intervenire bloccando la possibilità di condividere o pubblicare un determinato contenuto, "oscurandolo" e rimuovendolo all'accesso della platea di destinatari a cui era diretto, può costituire – specialmente si si tratta di materiale ideologicamente orientato – una forma di censura del pensiero, e divenire essa stessa forma di oppressione di minoranze e di repressione del dissenso politico.

Proprio nel caso delle recenti elezioni U.S.A., gli interventi di rimozione, da parte delle piattaforme social più note, dei profili e dei contenuti ritenuti costituire delle fake news hanno provocato la reazione di numerosi cittadini che avevano pubblicato e commentato tali contenuti e che hanno lamentato la violazione della propria libertà di manifestazione del pensiero.

L'amministratore delegato Zuckeberg, in un editoriale pubblicato sul Washington Post, ha chiesto ai governi regole stringenti in ordine a quali siano i contenuti accettabili e quali no sostenendo

che il peso della discrezionalità non può ricadere su aziende private.

La tendenza che si evince dall'osservazione delle proposte all'esame delle istituzioni nazionali è una responsabilizzazione degli operatori che sono chiamati a vagliare nel merito i contenuti e a procedere alla eventuale rimozione in base a una sostanziale discrezionalità, essendo il parametro di riferimento per il riconoscimento di materiale riconducibile all'incitamento all'odio o alla violenza piuttosto vago e indeterminato.

In questo modo il rischio con il quale nel prossimo immediato futuro occorrerà confrontarsi è di configurare una sorta di sistema di giustizia privata parallela, con le aziende del web tenute a intervenire in prima battuta a effettuare operazioni che, possono diventate una forma di censura all'espressione individuale.

De iure condendo sempre più pressante diventa in ogni caso la necessità di affrontare normativamente l'obbligo dell'identità digitale a carico di chi utilizza i social e l'estensione della normativa antiterrorismo e delle relative procedure d'urgenza per oscurare i siti sessisti.

FONTI: BIBLIOGRAFICHE, SITOGRAFICHE, NORMATIVE E DI GIURISPRUDENZA

BIBLIOGRAFIA

AA. VV., Contrasto multilivello al terrorismo internazionale e rispetto dei diritti umani, G. Giappichelli Editore, Torino 2012.

AL AZM e K. A. PAUL, *How Facebook Made It Easier Than Ever to Traffic Middle Eastern Antiquities*, in World Politics Review, August 2018, reperibile al sito worldpoli

ARESU *Geopolitica dell'intelligenza artificiale* Feltrinelli 2024

BACCIN *Responsabilità penale dell'internet service provider e concorso degli algoritmi negli illeciti online: il caso force v. facebook* in Sistemi Penali n.5/20

BACHELET, *Il rafforzamento del contrasto agli abusi di posizione "non dominante" delle piattaforme digitali* Ballardini B., Isis®. Il marketing dell'apocalisse, Baldini&Castoldi, Milano 2015.

BASSU *Istigazione all'odio, terrorismo e sicurezza nell'era digitale: c'è un limite alla libertà di espressione?* in Diritto Pubblico Europeo, 2019

BASSU *Istigazione all'odio, terrorismo e sicurezza nell'era digitale: c'è un limite alla libertà di espressione?* in Diritto Pubblico Europeo,2

BASSU, *Terrorismo e costituzionalismo. Percorsi comparati*, Torino, Giappichelli, 2010;

BENATTI, PORTONERA, *La responsabilità di diritto civile degli Internet Service Providers. Spunti dalla comparazione con la giurisprudenza statunitense,* in SAGGI, 2024.

BENATTI, PORTONERA, *La responsabilità di diritto civile degli Internet service providers. Spunti dalla comparazione con la giurisprudenza statunitense* 2024, La nuova giurisprudenza civile commentata, pag.476-485

BERLIN, "*Terrorisme et demande du statut de réfugié : quand le droit rejoint le bon sens, La Semaine Juridique - édition générale*" 2017 nº 7-8 p.312-313 (FR)

BETTI S., Le armi del diritto contro il terrorismo. Un esperto Onu fra diplomazia, codici e assistenza legale, Franco Angeli, Milano 2008.

BLAKKARLY, A gay sex tape is threatening to end the political careers of two men in Malaysia, SBS News, 17 giugno 2019.

BONETTI, *Terrorismo, emergenza e costituzioni democratiche*, Bologna, Il Mulino, 2006; T. E.

BRELAND, The Bizarre and Terrifying Case of the "Deepfake" Video that Helped Bring an African Nation to the Brink, Mother Jones, 15 marzo 2019;

CALDERINI Digital service act: cos'è e cosa prevede la legge europea sui servizi digitali in Agenza Digitale, 16 aprile 2025

CAPPELLINI, *"Machina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale*", in Criminalia: Annuario di scienze penalistiche, Edizioni ETS, 2018, Pisa:

CAPPIELLO, *The EU and the AI ACT. Was it worthwhile to be the first?*, in CERIDAP, 4, 2024, 235 e ss., ;

CASIERE, *A brave new (digital) world. Tra terrorismo e autoritarismo digitale,* in COSCIENZA E LIBERTÀ, 2024, 67.

CASIERE, *Intelligenza artificiale, terrorismo e responsabilità delle piattaforme digitali. Tra Stati Uniti e Unione Europea,* in COSCIENZA E LIBERTÀ, 2024, 68.

CATANZARITI *L'AI Act alla prova delle sfide globali: potenzialità e limiti di un modello regolatorio*, in Giustizia Insieme maggio 2025;

CHESNEY – D.K. CITRON, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security, 107 California Law Review 1753 (October 15, 2019).

CISTERNA, *L'evanescenza delle condotte condiziona le norme sul terrorismo* in *Guida al diritto*, 2025, n. 16,

COPPOLA *Intelligenza artificiale Metaverso il sistema penale: prevenzione, repressione, opportunità, rischi* Wolters Kluwer Cedam 2025

COPPOLA. *Commisurazione della pena e intelligenza artificiale: una ipotesi di lavoro con l'algoritmo Ex-Aequo,* in *Arch. pen.,* 2/2023

CORRIDORI, *Machina delinquere non potest?* in Giustizia Insieme del 19 maggio 2022

Council of Europe, Cyberterrorism: the use of Internet for terrorist purposes, Strasburg 2007

DAL POZZO, *Protezione dei dati personali e diritti fondamentali della persona: le nuove norme sui «codici di prenotazione» (PNR),* in RIVISTA DI DIRITTO INTERNAZIONALE PRIVATO E PROCESSUALE, pag. 1020, 2016, 4.

DE CARIA, Il dibattito sull'online speech: alla ricerca di un fondamento ideologico solido e coerente, tra liberta` di espressione, ''economic speech'' e liberta` di iniziativa economica, in Dir. econ., 2023, 315.

DE GREGORIO, Digital Constitutionalism in Europe. Reframing Rights and Powers in the Algorithmic Society, Oxford University Press, 2022.

DE MINICO, *Costituzione. Emergenza e terrorismo*, Jovene, Napoli 2016;

DE VERGOTTINI, *Guerra e Costituzione*, Bologna, Il Mulino, 2004

DE VITA *US Supreme Court: Gonzalez v. Google e l'internet liability* in https://www.devita.law 3 giugno 2023

DI NUOVO, *L'intelligenza artificiale tra percezioni soggettive e regolazioni normative*, in Giustizia insieme, maggio 2025, pag.2

DI TANO, Hate speech online: scenari, prospettive e criticità giuridiche del fenomeno, in Ciberspazio e diritto, 51 (2/3), 2014, pp.

DI VIZIO, *Prevenzione e investigazione: l'uso di IA, big data e soluzioni tecnologiche in ambito finanziario e nel contrasto al riciclaggio (AML) e al finanziamento del terrorismo (CFT),* in DISCRIMEN, 2024.

DI VIZIO, *Prevenzione e investigazioni: l'uso di IA, Big Data e soluzioni tecnologiche in ambito finanziario e nel contrasto al riciglaggio e al finanziamento del terrorismo*. Relazione al corso "*L*a digital transformation*: evoluzione del contesto e profili di*

impatto di diritto penale sostanziale e processuale", organizzato dalla Scuola Superiore della Magistratura in collaborazione con la Scuola di Polizia economico-finanziaria della Guardia di finanza, Cod. FPFP23015, 16 novembre 2023, Lido di Ostia (Roma).

FERRI, *L'Unione europea e la nuova disciplina sull'intelligenza artificiale: questioni e prospettive*, Napoli, 2024

FOTI Regolamentazione digitale: il Digital Service Act e le piattaforme online in Altalex, 11 luglio 2024

FROSINI, *Il diritto costituzionale alla sicurezza, in forumcostituzionale.it*, 2006;

GALLI, *Prevenzione del terrorismo nell'Unione Europea: un nuovo ruolo e responsabilità per le piattaforme informatiche?* in RASSEGNA DI DIRITTO PUBBLICO EUROPEO, 2019, XVIII.

GELLERT–JANSSEN, The Impact of Artificial Intelligence on European Tort Law: Taking Stock of an Ongoing Process, in The Future of European Private Law, a cura di Janssen–Lehmann–Schulze, Nomos/Hart, 2023, 175 e 193.

GENNUSO, *"Tutto in una definizione? La nuova direttiva antiterrorismo dell'Unione europea e i confini del terrorismo"* in *Quaderni costituzionali* / n.3 2017 pag.651

GIARDINI, Le regole dell'informazione: dal cartaceo al bit, G. Giappichelli Editore,Torino 2014.

GRAZIANI, *Terrorismo internazionale, radicalizzazione e tecnologia,* in FIDERALISMI.IT, 2023.

GROPPI (a cura di), *Democrazia e terrorismo*, Napoli, Editoriale Scientifica, 2006;

GROPPI (a cura di), *Democrazia e terrorismo*, Napoli, Editoriale Scientifica, 2006;

HALLEVY, "*The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*". Akron Intellectual Property Journal

HARWELL, 'Sexist' videos edited to make Nancy Pelosi look drunk go viral, with Trump's help, The Independent, 24 maggio 2019.

HESS, KALB., The media and the war on terrorism, Brookings Institution Press,

ical Comparative and Analysis, London, Routledge-Cavendish, 2008.

In the Content and Context of Hate Speech, Cambridge University Press, 2012;

INGLESE, *Il regolamento sull'intelligenza artificiale come atto per il completamento e il buon funzionamento del mercato interno?* in Rivista Quaderni AISDUE - *L'Unione europea e la nuova disciplina sull'intelligenza artificiale: questioni e prospettive*, a cura di F. FERRI, Napoli, 2024, 71 e ss.;

ITALIANO, *Intelligenza artificiale: passato, presente, futuro*, cit., p. 222

IURILLI, *Il diritto naturale come limite e contenuto dell'intelligenza artificiale. Prime riflessioni sul nuovo Regolamento Europeo "AI Act"*, in Judicium, 24 giugno 2024.

KOSSEF, *The Twenty-Six Words That Created the Internet*, Cornell University Press, 2019.

KOSSEFF, A User's Guide to Section 230, and a Legislator's Guide to Amending It (or Not), in 27 Berkeley Tech. LJ, 2022, 757; CANDEUB, Reading Section 230 as Written, in 1 J. Free Speech L., 2021, 139.

KOSTORIS, ORLANDI, *Contrasto al terrorismo interno e internazionale,* G. Giappichelli Editore, Torino 2006. Maggioni M., Terrore mediatico, Editori LaTerza, Bari 2015.

LAZCOZ MORATINOS, *Human oversight (article 14), in The EU regulation on artificial intelligence: a commentary* a cura di A. Huergo Lora, Milano, 2025, 243 e ss.

LEGG *Is There an Elegant Universal Theory of Prediction?* Proceedings of the 17th International Conference on Algorithmic Learning Theory (ALT 2006) Serie: *Lecture Notes in Computer Science* (LNCS), vol. 4264 Editore: Springer, Berlin, Heidelberg Pagine: 274–287

MARIOTTI, "*Status di rifugiato e partecipazioni alle attivita' di un gruppo terroristico", Giurisprudenza italiana* 2017 p.579

MARTORANA *Lotta al deepfake: stato dell'arte nell'UE* in www.altalex.com

MC CARTHY, MlNSKY, ROCHESTER, SHANNON, *A proposal or the Dartmouth Summer Research Project on Artificial lntelligence,* 31 agosto 1955, disponibile al link http://jmc.stanford.edu/articles/dart

MERVOSH, Distorted Videos of Nancy Pelosi Spread on Facebook and Twitter, Helped by Trump, The New York Times, 24 maggio 2019.

MITSILEGAL e SALVI, *Digital Exceptionalism, Freedom of Expression and the Rule of Law: The Case of Targeting Terrorism Content Online,* in RIVISTA EUROJUST, 2024, 2.

MONJE JR., Twitter letter to Chairman Schiff, Twitter, 31 luglio 2019;

Nacos, Mass-Mediated Terrorism, The Central Role of the Media in Terrorism and Counterterrorism, Rowman and Littlefield Publishers Inc., USA 2007.

NORDIO, *D.L. sicurezza quei giudici irriverenti verso il Colle. Danno per tutte le toghe* in www.ilmessaggero.it del 02.07.25

ONORATI, *I.A., politica e reati contro la personalità dello stato,* in ATTI DEL WORKSHOP FONDAZIONE VITTORIO OCCORSIO (Intelligenza artificiale e giurisdizione penale), 2021.

ONORATI, *I.A., politica e reati contro la personalità dello Stato*, in *Sistema penale*, 13 giugno 2022, all'interno della raccolta degli atti del workshop della Fondazione Occorsio su *Intelligenza artificiale e giurisdizione penale.* pag 105

ONORATI, *I.A., politica e reati contro la personalità dello stato,* in ATTI DEL WORKSHOP FONDAZIONE VITTORIO OCCORSIO (Intelligenza artificiale e giurisdizione penale), 2021

PIRAINO, Spunti per una rilettura della disciplina giuridica degli internet service providers, in AIDA, 2017, 498 s; ORTOLANI, The Resolution of Content Moderation Disputes under the Digital Services Act, in Giust. consensuale,

POLLICINO e DE GREGORIO, Hate speech, una prospettiva di diritto costituzionale comparato, in Giornale di Diritto amministrativo, 4, 2019, pp. 421-436;

POLLICINO, PITRUZZELLA e QUINTARELLI, Parole e potere: libertà di espressione, hate speech e fake news, Milano, Egea,

POLLICINO, Potere digitale, e RESTA, Poteri privati e regolazione, entrambe in Enc. del dir., I tematici, V, Potere e Costituzione, Giuffre` Francis Lefebvre, 2023.

POMPILI, *Il dissenso nelle nuove fattispecie di reato e nelle aggravanti introdotte con il DDL sicurezza*, in *Penale diritto e procedura,* 15 ottobre 2024.

ROMANELLI, Criminalità organizzata e terrorismo: la circolazione dei modelli criminali e degli strumenti di contrasto, in Sistema penale, 20 dicembre 2019.

ROMANELLI, *Criminalità organizzata e terrorismo: la circolazione dei modelli criminali e degli strumenti di contrasto, op.cit.*.

ROMANELLI, *Intelligenza artificiale, influenza sul mercato politico e reati contro la personalità dello Stato. La criminalità terroristica,* in SISTEMA PENALE.

ROMANELLI, *Intelligenza artificiale, influenza sul mercato politico e reati contro la personalità dello Stato. La criminalità terroristica,* in Sistema Penale 29 giugno 2023.

ROSENFELD, Hate speech in Constitutional Jurisprudence.

RUOCCO *L'avanzata delle leggi contro i deepfake* in www.agendadigitale.eu;

SACCHETTI, Il contrasto alla propaganda *terroristica nell'ambito dell'Unione Europea: tutela attuale e prospettive future,* in RIVISTA EUROJUST, 2019, 4.

SANTINI," *Una prima lettura della direttiva 2017/541 sulla lotta contro il terrorismo che sostituisce la decisione quadro Gai 2002",* in *Diritto penale contemporaneo* 2017, fasc.4 luglio 2017

SANTONI, *Il futuro della regolamentazione IA in CINA ecco i possibili scenari.* in Agenda Digitale del 1 aprile 2025

SANTOSUOSSO, B. MARONE, *Regole per l'IA: cosa può imparare l'Italia dalle strategie Usa e UK*, in Agenda Digitale, 21 novembre 2023.

SARTOR, *L'intelligenza artificiale e il diritto,* Giappichelli, 2022, 23 e ss

SCAFFARDI, Oltre i confini della libertà di espressione: l'istigazione all'odio razziale, Padova, Cedam, 2009;

SCHARFFS, A Commitment to Religious Freedom as the Bond that Makes Us Free, in The Review of Faith & International Affairs, n. 4/22, p. 24.

SITARAMAN, Deplatforming, in Yale Law J., 2023, 497 s., spec. 553 s., sia pure evidenziando che «the American tradition has not been one of either an absolute duty to serve or an absolute right to exclude. Instead, the American tradition has been one of reasonable deplatforming

SPIGNO, Discorsi d'odio: modelli costituzionali a confronto, Milano, Giuffré, 2018;

TESTAÌ, *Le indagini patrimoniali nel contrasto alla criminalità e al terrorismo,* in RIVISTA TRIMESTRALE DELLA SCUOLA DI PERFEZIONAMENTO PER LE FORZE DI POLIZIA, 2022, 2.

TORCHIA, *Pubblica amministrazione e transizione digitale*, in Giorn. dir. amm., 6, 2024, 729).

TURING, *Computing machinery and intelligence,* in *Mind,* UX. 1950,433.

UN News *Terror threat posed by ISIL 'remains volatile and complex,' Security Council hears* in news.un.org

VEDASCHI, *Intelligenza artificiale e misure antiterrorismo alla prova del diritto costituzionale,* in *Consulta Online* 17 febbraio 2020, 5

VEDASCHI, *The dark side of counterterrorism: arcana imperii and salus rei publicae*, in The American Journal of Comparative Law, 66, 4, 2018, pp. 877-926;

VICINANZA *La responsabilità delle piattaforme digitali nei confronti dei contenuti illegali: dal caso Telegram al Digital Services Act,* in Quaderni AISDUE

VIGANO', *"Terrorismo di matrice islamica-fondamentalistica e art.270-bis c.p. nella recente esperienza giurisprudenzale,* in *Cass.pen.,*2007, pag. 3953 ss

WALDRON, The Harm in hate speech, Cambridge, Harvard University Press, 2012;

YADAV, AI-Driven Digital Forensics. International Journal of Scienti5c Research & EngineeringTrends, Vol. 10 (2024), Issue 4, pp. 1673-1681.

YOO, The first emendment, common carriers, and public accommodations: net neutrality, digital platforms, and privacy, in 1 J. Free Speech L., 2021, 463.

ZENCOVICH, Freedom of expression: A Crit-

ZICCARDI, *Internet, controllo e libertà: Trasparenza, sorveglianza e segreto nell'era tecnologica,* Raffaello Cortina Editore, Milano 2015.

SITOGRAFIA

ANSA REDAZIONE, “Assistenti poco smart, Alexa propone sfida rischiosa a bimba”, Milano, 30/12/2021

ANSA, “Tesla: Elon Musk annunc

Carta nazioni unite e risoluzioni, in www.un.org

Convenzione di Ginevra, in www.admin.ch

Cyber- and AI- crime: rassegna delle novità dicembre 24 www.sistemapenale.it/it/scheda/cyber-and-ai-crime-rassegna-delle-novita

Fonti comunitarie, in www.europa.eu

https://www.garanteprivacy.it/home/docweb/-/docweb display/docweb/9512278

https://www.sistemapenale.it/it/scheda/cyber-and-ai-crime-rassegnadelle novità dicembre 2024-aprile 2025

https://www.sistemapenale.it/it/scheda/cyber-and-ai-crime-rassegna-delle-novita-dicembre-2025-aprile-2025 Cyber- and AI- crime

https://www.spectrumlocalnews.com

Rapporto Veridas, in https://veridas.com

LETTERATURA GRIGIA

AIGA (ASSOCIAZIONE ITALIANA GIOVANI AVVOCATI), 22 e 23 ottobre 2024;

Cattedra di Diritto penale 2, Titolari di insegnamento – Prof.ri E. GALLUCCI E M.N. MASULLOdel 18 marzo 2020, Luiss Guido Carli, Dipartimento di Giurisprudenza, Cattedra di Diritto penale 2,

HOLMES, Facebook just banned deepfakes, but the policy has loopholes – and a widely circulated deepfake of Mark Zuckerberg is allowed to stay up, (Jan. 7, 2020)

SABELLA P.M., "Il delitto di diffamazione. Struttura del fatto tipico nella dimensione offline e online", Lezione

SABELLA P.M., "La manifestazione odiosa del pensiero in Internet. Responsabilità individuali e dell'Internet Service Provider", Lezione del 23 marzo 2020, Luiss Guido Carli, Dipartimento di Giurisprudenza,

Se l'intelligenza artificiale finisce in tribunale ChatGpt diffama un sindaco australiano che ora vuole denunciarla, la Repubblica, 9 aprile 2023

NORMATIVA

LEGISLAZIONE STATI UNITI D'AMERICA

Legislazione Stati Uniti d'america *First Amendment to the United States Constitution.*

Legislazione Stati Uniti d'america *Section 230 of the Communications Decency Act, 47 U.S.C. § 230.*

Legislazione statunitense rilevante in materia di cybersecurity e antiterrorismo (*USA PATRIOT Act*, *Cybersecurity Act of 2015*).

LEGISLAZIONE UNIONE EUROPEA

AI ACT, Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio del 13 giugno 2024 entrato in vigore nell'agosto del 2025 che modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale), pubblicato sulla Gazzetta Ufficiale dell'Unione Europea Serie L del 12 luglio 2024.

Digital Service Act (meglio noto come DSA) Regolamento U.E. 2022/2065.

Digital Service Act (meglio noto come DSA) Regolamento U.E. 2022/2065.

Direttiva (UE) 2017/541 del Parlamento europeo e del Consiglio, del 15 marzo 2017, sulla lotta contro il terrorismo e che sostituisce la decisione quadro 2002/475/GAI del Consiglio e che modifica la decisione 2005/671/GAI del Consiglio.

Documenti e strategie dell'Unione Europea in materia di cybersecurity e lotta al terrorismo (*EU Cybersecurity Strategy*, *Counter-Terrorism Agenda*)

L'AML Package (Pacchetto antiriciclaggio UE) comprende i seguenti atti normativi: VI Direttiva Antiriciclaggio – Direttiva (UE) 2024/1640 (6AMLD), adottata il 31 maggio 2024. Modifica la Direttiva (UE) 2019/1937 (Whistleblower Protection) e abroga la precedente Direttiva (UE) 2015/849 (4AMLD); Regolamento Antiriciclaggio – Regolamento (UE) 2024/1624, anch'esso del 31 maggio 2024

Parlamento europeo *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 e alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (Regolamento generale sulla protezione dei dati - GDPR).*

Parlamento europeo *Regolamento (UE) 2022/2555 del Parlamento europeo e del Consiglio, del 14 dicembre 2022,*

relativo a misure per un elevato livello comune di cybersicurezza in tutta l'Unione, che modifica il regolamento (UE) n. 910/2014 e la direttiva (UE) 2018/1972 e abroga la direttiva (UE) 2016/1148 (direttiva NIS 2).

Proposta di Regolamento del Parlamento europeo e del Consiglio che stabilisce norme armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e che modifica alcuni atti legislativi dell'Unione.

Rapporto congiunto Eurojust ed Europol su "Common Challenges in Cybercrime" pubblicato il 31 gennaio 2025

Rapporto della Cyber Security Report pubblicato da TIM e dalla Cyber Security Foundation il 12 giugno 2025

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 e alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (Regolamento generale sulla protezione dei dati - GDPR).

Regolamento (UE) 2022/2554 del Parlamento europeo e del Consiglio, del 14 dicembre 2022, relativo alla resilienza operativa digitale per il settore finanziario e che modifica i regolamenti (CE) n. 1060/2009, (UE) n. 648/2012, (UE) n.575/2013, (UE) n. 909/2014 e (UE) 2016/1011 (Digital Operational Resilience Act - DORA).

Risoluzione delle nazioni unite: 1373 del 2001;

LEGISLAZIONE ITALIA

Art. 270-bis c.p.: Associazione con finalità di terrorismo anche internazionale.

Art. 270-quinquies c.p.: Arruolamento con finalità di terrorismo.

Art. 270-sexies c.p.: Definizione di finalità di terrorismo secondo il diritto interno e internazionale.

Art. 280-281 c.p.: Attentato per finalità terroristiche e altri gravi reati.

DDL 1146-B

Decreto Legislativo n. 65/2018 di recepimento della direttiva NIS

Decreto-legge n. 374/2001 (convertito nella Legge n. 438/2001)

Decreto-legge n. 7/2015 (convertito nella Legge n. 43/2015)

Dl sicurezza: relazione dell'ufficio del massimario della Corte di cassazione in Sistema penale.it giugno 2025.

Legge n. 155/2005 (Legge Pisanu)

GIURISPRUDENZA

BRANDSTETTER V. AUSTRIA, CEDU.

DAUBERT V. MERRELL DOW PHARMACEUTICALS, INC., 113 S.Ct. 2786(1993).

MALENCHIK V. STATE, 928 N.E.2d 564, 574 (Ind. 2010).

Sentenza della Cassazione Penale, Sez. III,

Sentenza Google v Gonzalez in www.supremecourt.gov

Sentenza Twitter v. Taamneh in www.supremecourt.gov

Legge n. 15 del 6 febbraio 1980 (Legge Cossiga)

Cass. pen., Sez. feriale, Sent., (data ud. 12/09/2013) 16/12/2013, n. 50620

Direttiva UE 2017/541 del 15 marzo 2017 «sulla lotta contro il terrorismo e che sostituisce la decisione quadro 2002/475/GAI del Consiglio»

Cass., sez. V, 13 luglio 2017 (dep. 3 novembre 2017), n. 50189;

Cass., Sez. VI, sent. 19 dicembre 2017 (dep. 29 marzo 2018), n. 14503

European Parliament resolution of 3 May 2022 on artificial intelligence in a digital age (2020/2266(INI)).

Risoluzione del Parlamento europeo del 3 maggio 2022 sulle implicazioni della guerra contro l'Ucraina per i diritti umani e gli obblighi degli Stati membri dell'UE nei confronti dei rifugiati provenienti dall'Ucraina (2022/2634(RSP)).

Decisione (UE) 2023/436 è in vigore dal 14 febbraio 2023

CGUE (C-634/21, Schufa Holding (Scoring), 7 dicembre 2023).

l. 28 giugno 2024, n. 90, in materia di rafforzamento della *cybersicurezza* nazionale e di reati informatici

l. 28 giugno 2024, n. 90, in materia di rafforzamento della *cybersicurezza* nazionale e di reati informatici

Proposta di legge n.675 depositata il 3 dicembre 2024 in www.assemblee-nationale.fr

Sentenza della Cassazione Penale, Sez. II, n. 21987 del 20 maggio 2019.

Cyber- and AI- crime: rassegna delle novità dicembre 24 www.sistemapenale.it/it/scheda/cyber-and-ai-crime-rassegna-delle-novita