

Le ‘informazioni sintetizzate’ generate dai *large language models* e le esigenze di tutela del diritto all’informazione: valori costituzionali e nuove regole.

di

Gianluca Fasano*

Sommario: 1. Introduzione. – 2. I *large language models* producono e diffondono informazioni. – 3. Libertà di pensiero e libero convincimento interiore. – 4. La regolazione europea: dal Regolamento sull’intelligenza artificiale (AI Act) alla Direttiva sulla responsabilità dell’AI (AILD). – 5. Conclusioni.

Introduzione

Lo sviluppo delle tecnologie digitali e dei nuovi media ha riaperto il dibattito, ormai da diversi anni, sulla tutela dei diritti e delle libertà costituzionali nella dimensione digitale e, in particolare modo, di quella della libertà di espressione.

Si è discusso molto della contrapposizione tra l’esigenza di arginare normativamente il fenomeno dei discorsi d’odio in rete, della disinformazione, e quella di non pregiudicare la tutela della libertà di espressione. Il tema non è di semplice impostazione anche per la maggiore complessità di ruoli e responsabilità che si affacciano nella dimensione online e si estende attorno alla rilevanza giuridica da attribuire alle condotte sui social, con specifico riguardo ai margini di azione e ai profili di responsabilità non tanto degli autori della condotta quanto piuttosto degli *host provider*.

In queste pagine non si vuol parlare della *querelle* circa ruolo e responsabilità dei provider ma si vuol allargare il campo di indagine alla cosiddetta “intelligenza

artificiale generativa”, da analizzare muovendo dalle riflessioni condivise dalla dottrina costituzionale sulla libertà di espressione¹.

Per “intelligenza artificiale generativa” si intende un campo dell'intelligenza artificiale (AI) che si concentra sulla creazione di sistemi in grado di generare dati, contenuti o output in modo automatizzato, spesso utilizzando tecniche basate su reti neurali profonde (deep learning), apprendimento automatico (machine learning) e modelli probabilistici in grado di raccogliere una conoscenza molto ampia, ricavandola da enormi quantità di dati, principalmente dal Web, e di produrre testi, immagini, suoni o video. L'addestramento dell'AI generativa può richiedere molte risorse computazionali, finanziarie e molto tempo a seconda della complessità del modello e del volume di dati di addestramento. La qualità dei dati di addestramento, l'architettura del modello e la validazione del modello sono tutti fattori critici per ottenere risultati coerenti.

Nel campo dell'intelligenza artificiale generativa particolare interesse hanno suscitato i *large language models*, progettati per generare testo in modo naturale, consentendo di analizzare, interpretare e rispondere a domande, completare frasi, tradurre testi in diverse lingue, generare codici, scrivere articoli e molto altro ancora². Il loro funzionamento è complesso e può essere così semplificato: fornita una frase come input, il sistema prosegue nella scrittura estraendo dal suo database le parole o le frasi che hanno la maggior probabilità di essere coerenti con l'input (e quelle che le hanno precedute), arrivando anche a scrivere interi articoli o addirittura saggi³. Possono essere utilizzati per una varietà di compiti (come

* Dirigente Tecnologico dell'Istituto di Scienze e Tecnologie della Cognizione del CNR.

¹ Non si disconosce l'importanza dei problemi derivanti dall'utilizzo di enormi quantità di dati nell'addestramento dell'AI generativa, principalmente quello della tutela della proprietà intellettuale di tali fonti e dei dati personali, ma la presente indagine vuole condurre un'analisi sulla natura delle informazioni sintetizzate e sui connessi profili di responsabilità.

² Questi modelli linguistici sono definiti “large” perché al loro interno sono presenti miliardi, e nei sistemi più recenti anche centinaia o migliaia di miliardi, di parametri. Per esempio, il noto GPT-3 (il sistema che alimenta ChatGPT e che è stato superato dall'ancora più grande GPT-4) ha 175 miliardi di parametri, mentre MT-NLG di Nvidia e Microsoft arriva a 530 miliardi. Il più grande in assoluto è però WuDao 2.0 dell'Accademia di Pechino per l'intelligenza artificiale, dotato della strabiliante quantità di 1.750 miliardi di parametri.

³ I *large language models* stanno rivoluzionando anche il settore accademico. ChatGPT, non senza clamore, è stato citato come uno dei 12 autori di un *preprint* sull'utilizzo dello strumento per la

assistenza virtuale, automazione dei processi aziendali, ricerca, scrittura, e altro ancora) e possono essere addestrati ulteriormente su specifici dati o task, con possibilità di incrementare le loro prestazioni in base all'esperienza e all'esposizione a nuove informazioni. Inoltre, sono accessibili da qualsiasi dispositivo connesso al web, consentendo a tutti di poter attingere informazioni da qualsiasi luogo e in qualsiasi momento.

L'utilizzo di questi sistemi non è esente da incognite, in termini di informazioni errate, perché basate su dati di addestramento la cui veridicità non è stata accertata, di bias e discriminazioni, perché se i dati di addestramento contengono pregiudizi questi vengono replicati negli output del sistema⁴, di manipolazione, posto che i risultati potrebbero esser presentati a terzi come autentici⁵, pur non essendolo, e anche in termini di concentrazione di un potere para-statale⁶.

Da qui si comprende come lo sviluppo di queste tecnologie innovative porti ulteriore linfa al dibattito incentrato sulla libertà di espressione nella dimensione digitale e costringe l'interprete a confrontarsi non più e soltanto coi problemi legati alla responsabilità derivante dalla selezione dei contenuti effettuata dai motori di ricerca, ovvero al dovere di sorveglianza sul contenuto dei messaggi veicolati dal *provider*⁷, ovvero al potere derivante dalla moderazione dei contenuti, quanto piuttosto coi problemi derivanti dalla generazione e diffusione ad un pubblico

formazione medica, pubblicato nel dicembre dello scorso anno su <https://www.medrxiv.org/content/10.1101/2022.12.19.22283643v2>.

⁴ M. NADEEM, A. BETHKE, S. REDDY. *Stereoset: Measuring stereotypical bias in pretrained language models*, in ACL-IJCNLP 2021. Secondo gli autori, poiché i modelli linguistici pre-addestrati vengono addestrati su grandi dati del mondo reale, è noto che catturano pregiudizi stereotipati per cui diviene importante quantificare in che misura questi pregiudizi siano presenti in essi.

⁵ Sulle incongruenze e limiti nelle performance di questi sistemi si veda BORJI, A., (2023), *A Categorical Archive of ChatGPT Failures*, arXiv:2302.03494v1, 6Feb2023.

⁶ Si veda O. POLLICINO, *L'efficacia orizzontale dei diritti fondamentali previsti dalla Carta. La giurisprudenza della Corte di giustizia in materia di digital privacy come osservatorio privilegiato*, in *MediaLaws*, 3/2018, 140. Come rilevato da A. SIMONCINI, *Sovranità e potere nell'era digitale*, in T.E. FROSINI, O. POLLICINO, E. APA, M. BASSINI (a cura di), *Diritti e libertà in Internet*, Firenze, 2017, 22, le Big Tech sono in grado di esercitare un potere non soltanto sul piano economico e tecnologico ma che si atteggia come *superiorem non recognoscens*, al pari del potere sovrano e non accetta confini, giuridici o territoriali.

⁷ Il *Digital Service Act* [REGOLAMENTO (UE) 2022/2065], all'articolo 8, prevede espressamente l'assenza di obblighi generali di sorveglianza a carico dei prestatori di servizi intermediari sulle informazioni che tali prestatori trasmettono o memorizzano.

indeterminato di informazioni sintetizzate da sistemi definiti 'autonomi'. Non ci si riferisce, quindi, alla fase che si apre dopo l'acquisizione delle informazioni, in cui è libertà di ciascuno farne l'uso che più aggrada (uso secondario). Uso che qui non rileva, sia per rappresentare una mera eventualità sia per essere regolato dall'ordinario principio del *neminem laedere*. Ci si riferisce, invece, alla fase generativa dell'informazione, resa possibile dalla e nell'interazione uomo-macchina, ove quest'ultima prende parte all'esercizio della libertà di pensiero del primo.

L'analisi muove dalla convinzione che l'ordinamento costituzionale possiede già delle solide guarentigie idonee a presidiare tale libertà e che l'interpretazione e le tecniche ermeneutiche consentiranno di attualizzare principi e regole vigenti rispetto all'evoluzione dei mezzi, sì da adattarsi a una realtà in continuo cambiamento⁸. Ciò nondimeno, i *large language models* si affacciano prepotentemente sulla scena giuridica per una caratteristica dirompente, la capacità di generare e veicolare informazioni sintetizzate da sistemi di auto-apprendimento (*self-learning*), una capacità di origine tecnica che, per via della larghissima diffusione e della pervasività che consegue, obbliga a una più approfondita riflessione sull'esercizio della «libertà di tentare di persuadere gli altri»⁹. Sarebbe auspicabile, in altre parole, che il legislatore si soffermasse, anche per ragioni di certezza del diritto, sui nuovi profili giuridici inerenti alla responsabilità derivante dalla creazione di "informazioni sintetizzate".

I *large language models* producono e diffondono informazioni

All'interno della comunità scientifica si discute se tali sistemi si limitino a combinare e riprodurre, in modo molto sofisticato, informazioni apprese dai dati di addestramento oppure se fanno qualcosa in più. L'argomento di discussione è se questi modelli siano creativi, originali, se apprendano veramente o semplicemente

⁸ Vedi C. SALAZAR, *Umano troppo umano...o no?*, in *BioLaw Journal*, 1, 2014, 257, per il quale «le norme giuridiche già esistenti siano inadeguate e insufficienti per regolare in tutte le sue implicazioni l'interazione tra gli uomini e queste macchine, all'evidenza "diverse" da tutte le altre».

⁹ A. C. JEMOLO, *I problemi pratici della libertà*, Milano, 1972, 47.

memorizzino i dati di addestramento e li riproducano in modo puramente statistico, combinando e riproducendo informazioni apprese dai dati di addestramento.

Quanto all'originalità, il meccanismo di coerenza tra output e dati di addestramento sembra impedire che la si possa riconoscere, essendo gli output una riflessione degli input ricevuti durante l'addestramento. Tuttavia, ciò non esclude che nel prossimo futuro lo sviluppo scientifico e tecnologico possa portare ad un avanzamento del modello che gli consentirà di andare oltre la mera riproduzione di informazioni apprese durante l'addestramento e di creare qualcosa di veramente nuovo e originale.

Quanto alla capacità di comprendere, gli studiosi di scienze cognitive concordano con i *computer scientists* nel mostrarsi ancora profondamente scettici rispetto alle sue concrete capacità di comprensione dei messaggi veicolati¹⁰. Si tratta di motori statistici (*transformers*) che elaborano quantità impressionanti di testi e producono un testo con il più alto valore di probabilità riferito alla conformità agli schemi conversazionali appresi. In pratica, l'AI generativa consente di 'apprendere' le regolarità di forma con cui una lingua si esprime (sintassi) ma non la semantica. Tuttavia, non possiamo escludere che la ricerca approdi a nuovi e differenti risultati che ci aiuteranno a comprendere meglio la natura profonda di questi algoritmi. Già si rinvergono in letteratura opinioni secondo cui la previsione realizzata con modelli di linguaggio neurale pre-addestrati è supportata, almeno in parte, da rappresentazioni dinamiche del significato¹¹ e non esclusivamente dalla modellazione accurata delle statistiche di co-occorrenza di parole.

Allo stato dell'attuale sviluppo tecnologico, possiamo osservare che, differentemente dai processi cognitivi umani, in cui la comprensione discende dalla

¹⁰ G. MARCUS, E. DAVIS, *GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about*, in *MIT Technology Review*, 22 agosto 2020, www.technologyreview.com/2020/08/22/1007539/gpt3-openai-languagegenerator-artificial-intelligence-ai-opinion/.

¹¹ LI B., NYE M., ANDREAS J., (2021), *Implicit Representations of Meaning in Neural Language Models*, Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, pages 1813–1827, <https://doi.org/10.48550/arXiv.2106.00737>.

capacità di rappresentarsi il mondo e di concettualizzare ed elaborare queste rappresentazioni, questi sistemi di AI non si rapportano al modello del mondo, per cui si dubita che possano sviluppare una vera capacità di comprensione¹².

Sul piano giuridico, l'analisi non può che essere coerente alla natura 'tecnica' dell'output generato. Tali sistemi utilizzano sofisticati modelli computazionali che eseguono un 'task di completamento': l'istruzione di fondo data all'AI generativa è quella di completare il discorso introdotto nella stringa di domanda. Non vi è spazio, quindi, né per la comprensione del messaggio né per un moto espressivo del suo contenuto. In definitiva, pare condivisibile l'assunto che l'output della macchina non è frutto di un pensiero consapevole: il task di completamento si sostanzia in un'associazione "tecnica" di parole, non assimilabile, a una manifestazione di pensiero. Conseguentemente, non può riconoscersi alle macchine la libertà di esprimere pensieri, idee od opinioni.

Un riferimento giurisprudenziale, rispetto a tale argomentazione, può rinvenirsi nella giurisprudenza francese pronunciata con riferimento alla "google suggest", una funzionalità integrata nel motore di ricerca di Google Inc. che propone agli utenti del web di effettuare una ricerca a partire dalle prime lettere della parola digitata, grazie a un menù a tendina che 'suggerisce' un elenco di dieci possibili *query*. Un semplice click sulla query proposta e si evita di dover digitare il testo completo della ricerca. Ebbene, tale funzione di completamento automatico, analizzata nel caso concreto sotto il profilo della lesione dell'onore della persona citata nei risultati delle query accanto a notizie disonoranti, non rientra nell'alveo concettuale della libertà di espressione posto che "ne constitue pas l'expression d'une pensée la simple juxtaposition de mots résultant uniquement d'un processus de classement systématique et automatisé, destiné à faciliter des recherches sur internet"¹³. In tale ultimo passaggio, la *Cour* esclude la qualificazione dell'output

¹² Vedi J. M. BISHOP, 2021, *Artificial intelligence is stupid and causal reasoning will not fix it*, in *Frontiers in Psychology*, 11, 2021, per il quale i large language models non pensano, non ragionano e non capiscono.

¹³ Cour de Cassation, Chambre civile 1, 19 June 2013, 12-17.591. Per un approfondimento sul tema: S. KARAPAPA, M. BORGHI, *Search engine liability for autocomplete suggestions: personality*,

della macchina come frutto del pensiero estendendo la sua riflessione sul piano dell'ontologia dei sistemi tecnici utilizzati e a prescindere dalla considerazione, pur paventata in altre decisioni giurisprudenziali, che essendo le libertà posizioni giuridiche che fanno capo a persone esse non possono esser riconosciute alle macchine¹⁴.

Ma, al di là del quesito prettamente epistemologico circa la creatività, originalità o capacità di apprendimento dei *large language models*, non v'è dubbio che essi, pur non esprimendo un pensiero, producano informazioni testuali comprensibili per le persone, e a queste indirizzate.

Nella giurisprudenza nazionale possiamo rintracciare concreti riferimenti a tale argomentazione. Sempre relativamente alla funzionalità 'google suggest', il Tribunale di Milano (Sezione I Civile - Ordinanza 23 maggio 2013) è giunto ad affermare che "sebbene le parole così aggregate non costituiscano frasi logicamente e grammaticalmente complete, non vi è dubbio che talvolta l'associazione di talune parole - per la naturale concordanza tra loro - sia immediatamente evocativa di un concetto finito e di un'informazione riferibile al soggetto il cui nominativo è stato digitato". In altri termini, pur non venendo in rilievo l'aspetto della manifestazione di un pensiero, di certo viene veicolato un determinato contenuto informativo¹⁵. E se questa conclusione si può raggiungere con riferimento a poche parole messe in fila da un suggerimento automatizzato, non diversamente possiamo argomentare con riferimento a frasi grammaticalmente e semanticamente complete prodotte dai *large language models*. Questi sono strumenti molto più complessi del 'google suggest' e i loro risultati sono evocativi di un pensiero umano 'vero e proprio'.

privacy and the power of the algorithm, in *International Journal of Law and Information Technology*, 3, 2015, 261-289.

¹⁴ Tribunal de Grande Instance de Paris, 17ème chambre, 8 settembre 2010, M. X./Google Inc., Eric S. et Google France. Più pertinente appare il percorso argomentativo del Tribunal de Grande Instance de Paris, 23 ottobre 2013, secondo cui «la liberté d'expression, laquelle ne s'entend, évidemment, que comme celle d'un être humain et à laquelle ne peut être assimilé le produit de calculs effectués par une machine».

¹⁵ In tal senso si veda il Bundesgerichtshof, VI ZR 269/12, 14 maggio 2013, che aveva ritenuto relativamente ad un sistema automatizzato di completamento (suggerimento di query) che l'uso del termine "Scientology" in connessione con il nome di una persona realmente esistente è adatta per evocare un'idea che è significativa in sé.

Dunque, i sistemi di *large language models* sono in grado di generare una narrativa testuale per mezzo di un'attività che non è meramente riproduttiva di fonti già esistenti, sono talmente pervasivi, consentendo a tutti di poter attingere informazioni da qualsiasi luogo e in qualsiasi momento, generalisti perché non hanno limiti di materie su cui pronunciarsi, persuasivi per via del loro stile narrativo calibrato sul profilo della persona e corrispondente alle sue aspettative di conoscenza, dall'aspetto confidenziale considerato il contesto in cui viene veicolata l'informazione sintetica, da risultare molto convincenti e seduttivi, in grado di inculcare un forte affidamento sui contenuti trasmessi.

In tale contesto si formeranno sempre più spesso idee, opinioni, convinzioni, quell'*humus* in grado di trasformare i moti della coscienza in colui che riceve informazioni indistinguibili da quelle prodotte dalle persone. Ed è questo l'aspetto che più rileva sul piano giuridico (ed etico), dacché le persone che recepiscono quel linguaggio sono portate a intendere quelle informazioni come pensieri, opinioni e idee, sono portate a realizzare una similitudine tra pensiero umano e informazioni sintetizzate, sono portate a vivere una suggestione che può incidere, con differenti gradazioni, sulla libertà di formare i propri convincimenti interiori e di conseguentemente agire: d'altronde, ciò che decidiamo dipende innanzitutto da ciò che sappiamo e crediamo. Allora, ancorché non possa assumersi che l'output dei *large language models* corrisponda alla manifestazione di un pensiero, tanto meno all'esercizio di libertà, deve convenirsi che le associazioni di parole prodotte implicano un significato per i destinatari, un significato fatto palese dal senso comune delle parole, secondo il linguaggio naturale, che va a influire sulla libertà di formazione del pensiero; implicano un significato che va a incidere sulla libertà di raccogliere le informazioni, di informarsi, su quella che viene indicata come il profilo passivo della libertà di informazione; implicano un significato che appartiene ad una relazione, ove la libertà è garantita «perché l'uomo possa unirsi all'altro uomo nel pensiero e col pensiero ed eventualmente insieme operare»¹⁶;

¹⁶C. ESPOSITO, *La libertà di manifestazione del pensiero nell'ordinamento italiano*, Milano, Giuffrè, 1958, p. 9.

implicano un significato che appartiene alla «libertà di tentare di persuadere gli altri»¹⁷; implicano un significato che, se nell'immediato tende a soddisfare una richiesta di informazioni, nel lungo periodo potrebbe ridurre le nostre abilità cognitive, la capacità umana di interpretare il mondo, di analizzare criticamente le informazioni e sviluppare competenze intellettuali. E, se una dipendenza eccessiva dalla tecnologia può ridurre la capacità delle persone di svolgere compiti senza assistenza tecnologica, ciò si tradurrebbe, nel contesto dei *large language models*, in un *vulnus* per un «pieno sviluppo della persona umana»¹⁸.

Non si ritiene, sia detto ad evitare equivoci, che i *large language models* abbiano il diritto di usare della libertà costituzionalmente garantita ma, senz'altro, partecipano al suo esercizio, dal momento in cui intervengono nello svolgimento di quella libertà. Tanto è sufficiente non solo per esigere standard in termini di qualità, veridicità e correttezza, ma soprattutto per porsi l'interrogativo sul modello normativo di tutele che deve presidiare la libertà di formazione pensiero, struttura intima della persona umana.

Libertà di pensiero e libero convincimento interiore

La prospettiva che si considera muove dalla constatazione che, a prescindere dalle capacità creative e inventive, in discussione v'è una tale mole di informazioni messa in circolazione con modalità talmente incessanti, pervasive e sofisticatamente persuasive da sollevare forti preoccupazioni sulla effettività della libertà di pensiero, nella sua dimensione fondazionale di formare, adottare e difendere le proprie convinzioni senza interferenze esterne, nell'esercizio del più libero convincimento interiore.

Non possiamo ignorare i rischi legati alla diffusione di questi modelli, come la concentrazione del potere nelle mani di poche grandi aziende private. Inoltre, c'è il pericolo che le persone vengano confinate nelle proprie convinzioni attuali a causa

¹⁷A. C. JEMOLO, *I problemi pratici*, cit.

¹⁸ Senza considerare l'ulteriore effetto della "oggettività algoritmica", che si presenta uguale o superiore al tipo di legittimazione che deriva dalle forme di conoscenza tradizionale (basate sulla scienza, le istituzioni pubbliche, i media), cfr. D. BOULLIER, *Sociologie du numérique*, Paris, Armand Colin, 2016, p. 149.

di algoritmi che mostrano loro solo informazioni che confermano ciò che già pensano. Questo è dovuto alle profilazioni digitali che limitano la diversità di opinioni. Inoltre, la raccolta massiva di opinioni delle persone è idonea a realizzare una sorveglianza digitale. In questo contesto, emerge con chiarezza la necessità di una riflessione critica sull'utilizzo responsabile delle tecnologie digitali, la quale non solo si configura come imprescindibile, ma si rivela anche insufficiente se non accompagnata da una regolamentazione che guidi uno sviluppo responsabile delle nuove tecnologie.

Si conviene che l'ordinamento costituzionale possiede delle solide garanzie idonee a presidiare tale libertà, ricavabili dall'intera Costituzione, ricca di disposizioni che in varia misura si riferiscono al libero convincimento interiore e che rappresentano la bussola su cui operare anche nell'era digitale. Basti pensare al principio personalistico che informa l'intero impianto costituzionale (art. 2 C)¹⁹; al principio di eguaglianza, posto a tutela dello sviluppo della persona umana (art. 3, co. 2 C); al principio di segretezza nel corrispondere con uno o più destinatari determinati (art. 15 C); alla libertà di formazione ed espressione del pensiero, sotto il profilo della libertà all'informazione (art. 21 C); alla libertà di associazione, in cui il vincolo associativo può esser guidato da un convincimento ideale condiviso (art. 18 C); alla libertà di fede religiosa, che coinvolge la sfera delle convinzioni personali e delle espressioni di fede (art. 19 C); alla libertà di insegnamento e ricerca, che assicura lo sviluppo e la diffusione della conoscenza e il progresso culturale (art. 33 C); per citarne solo alcuni²⁰.

Alla luce di questa ricca cornice normativa la libertà di formazione del convincimento interiore potrebbe apparire parcellizzata nelle diverse forme di

¹⁹ Nella sentenza n. 467 del 1991 la Corte costituzionale afferma che: «a livello dei valori costituzionali, la protezione della coscienza individuale si ricava dalla tutela delle libertà fondamentali e dei diritti inviolabili riconosciuti e garantiti all'uomo come singolo, ai sensi dell'art. 2 della Costituzione, dal momento che non può darsi una piena ed effettiva garanzia di questi ultimi senza che sia stabilita una correlativa protezione costituzionale di quella relazione intima e privilegiata dell'uomo con se stesso».

²⁰ La Corte costituzionale, sent. 43 del 1997, ha concluso che gli articoli 2, 3, 19 e 21 «contengono un insieme di elementi normativi convergenti nella configurazione unitaria di un principio di protezione dei cosiddetti diritti della coscienza».

protezione previste dalle diverse disposizioni costituzionali. E conseguentemente indebolita - e frammentata - nella tutela accordata. Eppure, essa vive un momento di unificazione, di garanzia della coerenza tra foro interno ed esterno, di garanzia dell'autenticità del pensiero, proprio grazie al riconoscimento della libertà di espressione come diritto fondamentale dell'individuo, e per l'individuo. In tale prospettiva, la manifestazione del pensiero da forma e rilievo esterno a una determinazione interiore, si fonde con essa e potrà dirsi autenticamente libera se la stessa formazione del convincimento interiore è stata libera. L'art. 21 Cost. (e, per il foro interno in materia religiosa, l'art. 19 Cost.) rappresenta(no) un fondamento costituzionale solido, robusto e completo per garantire una protezione piena della persona, in un panorama sempre più complesso e sofisticato in cui si affacciano macchine - i *large language models* - in grado di persuadere, convincere e vincere il pensiero autentico.

Per tale ragione, la libertà di espressione del pensiero è quella che ha ricevuto più attenzioni nello studio dell'interazione tra diritti e mondo digitale. L'avvento del web ha rivoluzionato il modo di produrre, diffondere e usare l'informazione, e proprio la condivisione delle idee è stata maggiormente trasformata con internet, anche per via della propensione di sostituire le relazioni umane con le 'connessioni'²¹. È utile, quindi, ripercorrere le riflessioni della dottrina costituzionale sulla libertà di espressione, tradizionalmente intesa come diritto di libertà, cioè come posizione del singolo a esprimere le proprie idee liberamente in assenza di qualsiasi ingerenza del potere pubblico.

Al momento di esternare le proprie idee, è vero, si esercita una libertà proclamata, ma ancor prima si esprime una personalità, si dà voce alla persona. Questo aspetto viene lucidamente evidenziato dalla teoria 'individualista'²², il cui più autorevole

²¹ Z. BAUMAN, *Amore liquido: Sulla fragilità dei legami affettivi*, Roma, 2007.

²² Concezione individualista o liberale, per la quale il diritto di manifestare il proprio pensiero è attribuito all'uomo in quanto tale e a suo vantaggio, non all'uomo nella qualità di membro di una comunità e a vantaggio della stessa e dei suoi valori. P. BARILE, *Diritti dell'uomo e libertà fondamentali*, il Mulino, Bologna, 1984, p. 229, sostiene che la libertà di manifestazione del pensiero presenta un duplice profilo, uno individualistico, l'altro funzionale «in quanto tendente al buon funzionamento del regime democratico, che sulla partecipazione si fonda». Similmente, C. MORTATI, *Istituzioni di diritto pubblico*, II, Padova, 1969, 977s, per il quale nei

sostenitore, Carlo Esposito, afferma che la Costituzione “non intende attribuire un diritto funzionale all'individuo, ma un diritto individuale” alla persona, un diritto attribuito all'uomo in quanto tale “indipendentemente dai vantaggi o dagli svantaggi che possano derivarne alla Stato”²³. Le ragioni ideali del riconoscimento di questa libertà risiedono nel valore della persona umana, in coerenza al principio ‘personalistico’ che informa l'intera Costituzione (art. 2 C.), ove si «esalta la persona, assunta come il fondamento e il fine dell'ordine»²⁴. Non nega, vieppiù, il valore sociale dei diritti di comunicazione e manifestazione del pensiero, posto che il relativo esercizio presuppone la vita in società e il loro libero svolgimento avviene necessariamente nella relazione tra individui. Ma nega che essi siano riconosciuti per finalità sociali, per le «utilità sociali delle unioni di pensiero». Il diritto di potersi esprimere liberamente senza ingerenza delle autorità è garantito all'uomo «per l'appagamento egoistico dei suoi bisogni e desideri individuali»²⁵. Così intesa, la libertà di espressione è un catalizzatore per la formazione delle coscienze. Prima ancora della libera espressione, e per questa finalità, è necessario che sia mantenuto libero il foro interno e questo, nell'era contemporanea non è minacciato solo dalla forza fisica, ma anche dalla forza della persuasione (e della manipolazione) che si esprime in forme sempre più sofisticate, cui anche i *large language models* appartengono. Il foro interno è il luogo nel quale le idee prendono forma e dal quale escono, si diffondono e si confrontano, grazie alla libertà di manifestazione del pensiero; e, in questo processo dinamico, la libertà di

sistemi democratici «la garanzia di buon funzionamento del sistema poggia proprio sulla più ampia libertà di manifestazione del pensiero, essendo essa che alimenta la forza sociale di base, che è la pubblica opinione».

²³ C. ESPOSITO, *La libertà*, cit. La Corte costituzionale sin da risalenti pronunce ha infatti asserito che tale libertà «è tra le libertà fondamentali proclamate e protette dalla nostra Costituzione» (sentenza n. 9 del 1965), sottolineando che il diritto previsto dall'art. 21 Cost. è «il più alto, forse,» dei «diritti primari e fondamentali» sanciti dalla Costituzione (sentenza n. 168 del 1971).

²⁴ P. COSTA, *Diritti fondamentali*, in *Enc. dir.*, Annali, II, 2008, p. 396.

²⁵ C. ESPOSITO, *La libertà*, cit., p. 9. In merito ai rapporti tra libertà di espressione del pensiero e democraticità della Stato l'Autore chiarisce il fenomeno in questi termini: «Si vuole solo affermare che non la democraticità dello Stato ha per conseguenza il riconoscimento di quella libertà, sicché possa determinarne la funzione ed i limiti, ma che le ragioni ideali del riconoscimento di quella libertà (e cioè del valore della persona umana) porta tra le tante conseguenze anche alla affermazione dello Stato democratico».

espressione diventa la linfa vitale di un ordinamento che rispetta la diversità di pensiero e promuove «il pieno sviluppo della persona umana».

La connessione con l'art. 3, comma 2, della Costituzione è essenziale, poiché questa disposizione non sancisce solo l'uguaglianza dei cittadini di fronte alla legge, indipendentemente dalle loro opinioni politiche, religiose o sociali, ma esige che siano rimossi “gli ostacoli di ordine economico e sociale, che (...) impediscono il pieno sviluppo della persona umana”. E in un'epoca in cui l'informazione è onnipresente, la possibilità di formare le proprie idee e di contribuire al dialogo sociale diventa una via per affermare la propria identità, cioè il «diritto di essere sé stesso»²⁶ che, sul piano relazionale, culmina nel qualificarsi per distinguersi dagli altri individui, ma che sul piano fondazionale si modella sul bagaglio di convinzioni ideologiche, politiche, religiose, filosofiche, morali e sociali.

Ma non è soltanto l'aspetto interiore della libertà di espressione che rileva in tema di formazione del pensiero, sussiste anche un aspetto dinamico e relazionale. Attorno all'art. 21 C. vengono a svilupparsi, principalmente con l'avvento del mezzo televisivo, orientamenti volti a costruire le basi di una «libertà di informare»²⁷ il prossimo, prospettiva non estranea ai lavori della Costituente²⁸, intesa come libertà di acquisire, dare e divulgare notizie, nonché opinioni sui fatti di pubblico interesse. In mancanza di una disciplina costituzionale dell'informazione (con la sola eccezione della stampa) e di una prescrizione generale sui mezzi, la Corte costituzionale ha sempre ricondotto il diritto di informare nell'ambito di tutela della libertà costituzionale di manifestazione del pensiero, sulla base della considerazione che le notizie, al pari delle opinioni, sono espressioni del pensiero²⁹. Per questa via la tutela costituzionale è stata proiettata ben al di là della manifestazione del pensiero, ricostruendo uno strettissimo nesso tra «libertà d'informazione» e democrazia pluralistica e conquistando una

²⁶ Corte cost., 3 febbraio 1994, n. 13.

²⁷ Cfr., in particolare, Corte cost. 1972/172.

²⁸ Vedi P. CARETTI, *Comunicazione e informazione*, in *Enc. dir.*, Annali, I, 2007, p. 223 ss.

²⁹ Cfr., ex multis, Corte cost. 1965/25; 1972/105; 1974/20, 86, 225; 1981/1, 18; 1990/348.

rilevanza anche in una dimensione latamente sociale, come interesse generale della collettività all'informazione (sent. n. 105 del 1972; sent. n. 225 del 1974).

Di libertà di informazione si parla non solo con riferimento al diritto di informare il prossimo, ma anche come 'diritto a essere informato', sulla scorta soprattutto dell'art. 19 della Dichiarazione universale dei diritti dell'uomo, ove dopo il riconoscimento della libertà di opinione e di espressione si fa scaturire come implicazione esplicita «il diritto di ricercare, di ricevere e di diffondere» le idee attraverso qualsiasi mezzo espressivo. La formulazione dell'art. 19 conferma che il diritto a essere informato serve proprio a garantire la libertà del foro interno, il libero convincimento, per cui l'aspetto dinamico sarebbe pur sempre funzionale a quello interiore.

Di un vero "diritto all'informazione" la Corte giungerà a parlare con la sent. n. 153/1987, in cui il "diritto" era speculare ai contrapposti obblighi della concessionaria del servizio pubblico radiotelevisivo. I contenuti di tale pronuncia verranno poi ripresi nella sent. n. 112 del 1993, dove il «diritto ad essere informati» esce dall'angusta prospettiva del rapporto sinallagmatico di utenza e si afferma come diritto "inviolabile" ai sensi dell'art. 2 Cost. È utile riportare il passaggio della sentenza: «Questa Corte ha costantemente affermato che la Costituzione, all'art. 21, riconosce e garantisce a tutti la libertà di manifestare il proprio pensiero con qualsiasi mezzo di diffusione e che tale libertà ricomprende tanto il diritto di informare, quanto il diritto di essere informati (v., ad esempio, sentt. nn. 202 del 1976, 148 del 1981, 826 del 1988)».

La dimensione sociale del diritto di informare, inteso come interesse generale della collettività all'informazione, e il diritto ad essere informati conforma la stessa normativa sulla responsabilità editoriale e la deontologia professionale del giornalista, al fine di garantire un ambiente informativo responsabile e rispettoso dei diritti nell'interesse generale.

Nel mondo dei *large language models* la posizione dell'utente dell'informazione è affatto diversa rispetto al contesto della stampa, della radio o della televisione, ma l'interesse da proteggere appare essere il medesimo, ovvero l'affidamento

all'informazione sotto il profilo del pluralismo delle fonti, dell'obiettività e dell'imparzialità dei dati forniti, della completezza, della correttezza e del rispetto della dignità umana e degli altri valori primari garantiti dalla Costituzione. Tutte esigenze, queste, che, se non rendono superati i tradizionali e risalenti strumenti di garanzia, pensati per l'editoria tradizionale, richiedono una riflessione di sistema che attualizzi le tutele costituzionali in questo *nuovo ambiente* di svolgimento della libertà di informare, della libertà all'informazione, che tuteli la fiducia delle persone destinatarie di quelle informazioni, e che, in definitiva, individui un modello di responsabilità inerente alle 'informazioni sintetizzate'. Se per altri ambiti è sufficiente il ricorso – adattandole - alle categorie giuridiche tradizionali, per i *large language models* è necessario costruire, sulla base sempre delle disposizioni costituzionali, nuovi modelli che tengano conto dei ruoli di tutti gli attori in gioco, dalla fase di ideazione e progettazione, passando per la fase di addestramento e a quella esperienziale, fino a giungere alla fase di utilizzo finale ove l'interazione tra la macchina 'autonoma' (su cui *infra*) e la persona porta a completamento il processo di costruzione delle informazioni sintetizzate. L'utilizzatore finale, in effetti, attraverso la sua domanda di conoscenza, interpretata dalla macchina come un'istruzione, un 'task' di completamento, contribuisce a determinare il risultato finale. E non solo. Egli partecipa attivamente al processo di apprendimento continuo del modello, posto che le conversazioni degli utenti possono essere utilizzate per raffinare e migliorare periodicamente il modello attraverso l'aggiornamento dell'addestramento. Ciò aiuta il modello a comprendere meglio le dinamiche linguistiche, a conoscere meglio l'utente persona, a rispondere in modo più coerente alle sue domande, quindi, ad esser più convincente, persuasivo, a migliorare le sue capacità generali di simulare una conversazione, senza apparenti limiti se non quelli, comunque ignoti, fissati in sede di programmazione dall'algoritmo. Senza limiti di coscienza e di comprensione etica³⁰.

³⁰ Come si dirà ampiamente nel paragrafo seguente, l'autonomia che si riconosce alla macchina non è qualificata da un'espressione di volontà, intenzione o scopo, restando così confinata sul piano meramente esecutivo (dell'algoritmo).

La regolazione europea: dal Regolamento sull'intelligenza artificiale (AI Act) alla Direttiva sulla responsabilità dell'AI (AILD)

Il dibattito circa il ruolo e i limiti della regolamentazione dell'innovazione digitale trova facile sponda nella difficoltà che il diritto incontra, obiettivamente, nel disciplinare un fenomeno in così rapida evoluzione e a forte connotato tecnico³¹. Ma, un punto pare poter esser condiviso: fenomeni globali, quali internet, le forme d'intelligenza artificiale, le informazioni sintetizzate, non si possono disciplinare se non con strumenti globali³², entro i confini delle tradizioni giuridiche comuni, almeno formalmente.

È proprio questa consapevolezza che ha spinto l'UE a proporsi sullo scenario globale come leader nella regolazione dell'innovazione digitale. Certo l'iniziativa ha anche un forte connotato politico, volendo l'UE affermare una sovranità digitale a fronte del potere delle aziende tecnologiche d'oltreoceano che, con i loro prodotti e servizi innovativi, stanno rimodellando il mondo. Perciò, con riferimento al regolamento sulla protezione dei dati personali, al regolamento sulla libera circolazione dei dati non personali³³, al regolamento sulla cybersicurezza³⁴, alla direttiva sull'apertura dei dati³⁵, alla proposta di regolamento sull'intelligenza artificiale (in fase di finalizzazione), e altro ancora, non è eccessivo affermare una loro rilevanza globale, posto che esse finiscono per coinvolgere anche le aziende extra UE.

³¹ Cfr. P. COSTANZO, *Il blog tra vocazione libertaria della rete e limiti costituzionali della manifestazione del pensiero*, in *Informatica e diritto*, 2008, 63; D. MARONGIU, *Organizzazione e diritto di internet*, Milano, 2013; A. CELOTTO *Algoritmi e algoretica: quali regole per l'intelligenza artificiale?*, in *Liber Amicorum per Pasquale Costanzo, Costituzionalismo, Reti e Intelligenza artificiale*, p. 367 ss, in cui l'autore riferisce di "una vera "nuova" generazione dei diritti".

³² Di «questione eminentemente transnazionale, e come tale difficilmente ricomponibile ricorrendo alla tradizione giuridica singolo-nazionale», parla A. VENANZONI, *Intersezioni costituzionali – Internet e Intelligenze Artificiali tra ordine spontaneo, natura delle cose digitale e garanzia dei diritti fondamentali*, in *Forum di Quaderni Cost.*, 27 aprile 2018, 4.

³³ Regolamento (UE) 2018/1807.

³⁴ Regolamento (UE) 2019/881.

³⁵ Direttiva (UE) 2019/1024.

Il fenomeno è descritto come 'Brussels Effect'³⁶, inteso quale potere unilaterale dell'UE di regolare i mercati globali senza la necessità di ricorrere a istituzioni internazionali o di cercare la cooperazione di altre nazioni. Attraverso l'accennato apparato di regole, tese a governare il prezioso mercato interno dell'Unione, l'UE costringe le aziende straniere, che a quel mercato mirano, a dover seguire le norme europee, generando un'ampia spinta alla conformità da parte di soggetti extraterritoriali.

La normativa europea riesce così ad affermare una sovranità digitale europea anche al di fuori dei propri confini, andando a collocarsi in aperta contrapposizione all'approccio d'oltre oceano. Gli Stati Uniti, ad esempio, hanno tradizionalmente abbracciato una visione tecno-libertaria che enfatizza il primato del libero mercato, della libertà di parola e della libertà di Internet. In Cina, invece, si assiste a una combinazione tra sorveglianza governativa e forte controllo delle imprese Big Tech su massicce quantità di dati, senza sufficienti garanzie per i cittadini. L'Europa, invece, fonda le proprie regole sulla convinzione che i mercati, lasciati a sé stessi, non produrranno risultati ottimali e che l'intervento del governo è necessario per proteggere i diritti dei cittadini nell'era digitale.

Sullo sfondo di questa strategia politica si innesta lo strumentario normativo dell'UE, che mira a enfatizzare un 'approccio umano-centrico', espressione con cui si vuol evidenziare l'attenzione posta sulla complessa trama di relazioni che vengono a instaurarsi tra persone - i loro diritti - e tecnologia. Nel dibattito istituzionale³⁷ e scientifico³⁸ si fa sempre più strada una concezione protetica³⁹, che

³⁶ Cfr. A. BRADFORD, *The Brussels Effect – How the European Union Rules the World* (New York, 2020; edn online, Oxford Academic, 19 dicembre 2019), <https://doi.org/10.1093/oso/9780190088583.001.0001>.

³⁷ Il 18 febbraio 2020, il Ministro per l'innovazione tecnologica e la digitalizzazione e il Presidente della Fondazione Leonardo hanno firmato un protocollo d'intesa, per definire il contesto etico e giuridico all'interno del quale sviluppare e applicare l'intelligenza artificiale, in particolare per rispondere alle esigenze della Pubblica Amministrazione.

³⁸ J. SCHNAPP - A. BURDICK - J. DRUCKER - P. LUNENFELD - T. PRESNER, *Umanistica Digitale*, Milano, 2014.

³⁹ La visione protetica venne teorizzata almeno un ventennio fa da V. FROSINI, *Il giurista e le tecnologie dell'informazione*, Roma, 1998, 58, il quale evidenziava come l'informatica avesse determinato una vera mutazione antropologica: "Dialogare attraverso le macchine, o con le macchine stesse, non è la stessa cosa di conversare a viva voce: si adopera una protesi

ricosce alle nuove tecnologie il compito di espandere i confini della cognizione umana, determinando la liberazione della persona dai propri limiti cognitivi e assicurando alla società intera una profonda innovazione⁴⁰. Lo spazio della cognizione umana si espande considerevolmente incidendo - sia in senso positivo che negativo - i diritti fondamentali, ricadendo su limiti, tempi, modi e forme di esercizio dei diritti. Un approccio umano-centrico mira a uno sviluppo dei sistemi di intelligenza artificiale ponendo la tecnologia a servizio dell'uomo, nel suo realizzarsi come persona in tutti gli ambiti della vita.

Tale visione è stata tradotta nella strategia per lo sviluppo di un AI affidabile⁴¹, in particolare nella *Communication "Building Trust in Human-Centric Artificial Intelligence"*⁴², con cui la Commissione Europea ha recepito i requisiti fondamentali stabiliti negli *Orientamenti etici per una Intelligenza artificiale affidabile* predisposto dal Gruppo di Alti esperti sull'AI. Più di recente, l'approccio umano-centrico è stato posto alla base della «*Dichiarazione europea sui diritti e i principi digitali per il decennio digitale*», in cui, ancora una volta, il regolatore è chiamato a ponderare il bisogno di tutela della persona con la forte spinta verso la promozione dell'economia digitale⁴³.

intellettuale, che modifica il procedimento del conoscere e del comunicare. L'uomo è divenuto essenzialmente diverso da quello che è stato in tutti i tempi, e non potrebbe tornare indietro senza perdere la sua nuova umanità, che gli è stata conferita dalla cultura informatica di massa".

⁴⁰ Cfr. A. SIMONCINI, *Amministrazione digitale algoritmica. Il quadro costituzionale*, in R. CAVALLO PERIN - D.U. GALETTA (a cura di), *Il diritto dell'amministrazione pubblica digitale*, Torino, 2020, 3.

⁴¹ Nell'ambito della Strategia europea per l'Intelligenza Artificiale, la Commissione europea ha pubblicato il 21 aprile, la **proposta di Regolamento sull'approccio europeo all'Intelligenza Artificiale** [COM(2021) 206 final] che propone il **primo quadro giuridico europeo sull'AI, fondato sulla visione umano-centrica**. Il nuovo approccio europeo fa seguito ad una serie di iniziative intraprese negli ultimi anni, tra cui, la consultazione pubblica sul Libro Bianco sull'Intelligenza Artificiale (COM 2020) 65 final del 19 febbraio 2020) e gli *Orientamenti etici per una Intelligenza artificiale affidabile*, del Gruppo ad alto livello sull'intelligenza artificiale, pubblicate l'8 aprile 2019.

⁴² COM(2019)168 final, 8 aprile 2019.

⁴³ I dati sono considerati una risorsa essenziale per il progresso sociale in generale, e in particolare per la crescita economica, la competitività, l'innovazione e la creazione di posti di lavoro. Nel considerando 4 troviamo espressamente delineati i valori in gioco, evidenziando che «La visione dell'UE per la trasformazione digitale mette al centro le persone, conferisce loro maggiore autonomia e responsabilità e promuove le imprese innovative» (26.1.2022 COM(2022) 28 final).

Come parte della strategia digitale europea⁴⁴ occorre menzionare anche la proposta di Regolamento del Parlamento europeo e del Consiglio, il cui testo non è ancora consolidato, che stabilisce regole armonizzate sull'intelligenza artificiale (AI Act del 21 aprile 2021)⁴⁵.

Allo stato attuale, la proposta prevede di classificare i sistemi di AI in differenti livelli di rischio, secondo una metodologia *risk based*, di guisa da applicare a ciascun livello di rischio un set di strumenti e azioni ad hoc. In particolare, vengono previsti quattro diversi livelli di rischio: rischio inaccettabile, rischio elevato, rischio limitato e rischio minimo. I sistemi di intelligenza artificiale che comportano rischi inaccettabili rappresentano una chiara minaccia per la vita e alla sicurezza delle persone, per cui verranno vietati. Per gli altri livelli di rischio sono previsti obblighi via via meno stringenti⁴⁶.

Quanto ai *large language models*, la proposta prevede obblighi inerenti a requisiti di trasparenza nonché a «rivelare che il contenuto è stato generato da un'intelligenza artificiale; progettare il modello in modo da impedire la generazione di contenuti illegali; pubblicare riepiloghi dei dati con diritti d'autore utilizzati per l'addestramento».

Queste risposte tecnico-giuridiche non affrontano in modo esaustivo le problematiche tutte che investono l'AI generativa, non il tema della disinformazione, né quello della logica alla base dell'AI, né quello della autorialità e della responsabilità. Ma, sia detto a evitare equivoci, non si tratta di un vuoto normativo, posto che l'AI act non vuole disciplinare l'intera materia dell'intelligenza artificiale, come una *summa* di tutte le norme, ma vuole soltanto

⁴⁴ Comunicazione della Commissione, Plasmare il futuro digitale dell'Europa (COM(2020) 67 final).

⁴⁵ Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale e modifica alcuni atti legislativi dell'Unione, Bruxelles, 21.4.2021 COM(2021) 206 final, 2021/0106(COD)). Di notevole rilievo le successive proposte emendative, soprattutto la posizione negoziale del Parlamento UE del 14 giugno 2023.

⁴⁶ Cfr. G. FINOCCHIARO, *La regolazione dell'intelligenza artificiale*, in *Rivista Trimestrale di Diritto Pubblico*, 2022, LXXII, pp. 1085 ss., secondo la quale, mentre il sistema *risk management* è accompagnato dal principio di *accountability* nel GDPR, ciò non accade nell'AI Act, ove la valutazione del rischio dei sistemi intelligenti non è rimessa alle scelte degli operatori economici, ma al legislatore.

determinare condizioni e requisiti che i sistemi di AI devono rispettare per poter accedere al mercato unico digitale⁴⁷.

Per tale ragione occorre volgere l'analisi verso quella che viene definita l'altra faccia della stessa medaglia⁴⁸, le norme in materia di responsabilità dell'AI.

Già nel 2020, il Parlamento europeo emanava una risoluzione⁴⁹ in cui manifestava l'esigenza di un regime armonizzato a livello di UE in materia di risarcimento dei danni causati dai sistemi di AI. Fornitori, operatori, utenti dei sistemi di AI e gli stessi danneggiati senza una disciplina uniforme si troverebbero di fronte a 27 regimi di responsabilità differenti, il che comporterebbe livelli di protezione disuguali e una distorsione della concorrenza tra le imprese di Stati membri diversi.

Il 28 settembre 2022 la Commissione ha presentato la proposta di direttiva sulla responsabilità in materia di intelligenza artificiale (AILD - AI liability directive)⁵⁰. Per coloro che subiscono un danno, identificare il responsabile e dimostrare che sussistano gli elementi del nesso causale e del danno potrebbe risultare difficile o eccessivamente costoso, in considerazione della complessità, autonomia e opacità dei sistemi di AI⁵¹. Per questo motivo, la direttiva mira all'alleggerimento

⁴⁷ L'obiettivo del mercato unico digitale ha consentito di promuovere la regolamentazione del digitale come strumento per far avanzare l'integrazione europea. Un mercato unico digitale ben funzionante richiede normative UE armonizzate poiché standard normativi differenti ostacolano il commercio transfrontaliero. Nella *Strategia per il mercato unico digitale in Europa* (COM(2015) 192 final) la Commissione UE attesta che la costruzione di un mercato unico digitale potrebbe arricchire il PIL europeo di 415 miliardi di euro.

⁴⁸ Proposta di Direttiva del Parlamento Europeo e del Consiglio relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale (direttiva sulla responsabilità da intelligenza artificiale), 2022/0303(COD). Se l'AI act ha l'obiettivo di ridurre i rischi di sicurezza, consci di non poterli eliminare del tutto e delle possibili conseguenze dannose, sono necessarie norme in materia di responsabilità che rafforzano quelle sulla sicurezza.

⁴⁹ Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale (2020/2014(INL)).

⁵⁰ La scelta dello strumento legislativo è caduta sulla direttiva, posto che gli Stati membri ambiscono a conservare una propria identità riguardo alle norme in materia di responsabilità extracontrattuale (attinenti al diritto privato) che rappresentano tradizioni giuridiche di lunga data.

⁵¹ Il 28 settembre 2022 la Commissione europea ha presentato anche la Proposta di direttiva sulla responsabilità per prodotti difettosi (PLD). L'obiettivo comune delle proposte è adattare le norme in materia di responsabilità all'era digitale, al fine di garantire che le vittime dei danni

dell'onere della prova del danneggiato attraverso due principali strumenti: presunzione relativa e divulgazione di informazioni. La prima è volta a semplificare ai danneggiati la prova del nesso causale tra la colpa del convenuto-danneggiante e l'output prodotto dal sistema di AI o la mancata produzione di un output da parte del sistema di AI che ha provocato il danno. La seconda mira ad assicurare, a coloro che chiedono il risarcimento del danno, la possibilità di ottenere informazioni sui sistemi di AI ad alto rischio che devono essere registrate/documentate a norma della legge sull'AI⁵².

A fronte di questo quadro normativo, seppur in fieri, sorge il dubbio che il concetto di colpa di una persona possa non essere del tutto adeguato rispetto alle problematiche sottese ai *large language models*. Questi sistemi utilizzano tecniche basate su reti neurali profonde (deep learning), apprendimento automatico (machine learning), sono quindi dotati di capacità di autoapprendimento⁵³, di guisa che l'output non discende soltanto dalle regole fissate in sede di progettazione, dalle attività di training successivo, dai dati esperienziali acquisiti nel tempo e dagli input dell'utilizzatore, ma da una combinazione di tali fattori che restano indissolubilmente mescolati e non più distinguibili, l'uno dall'altro, al termine del processo. La complessità di queste tecnologie rende difficile la comprensione, anche da parte degli stessi progettisti e programmatori, dei meccanismi in base ai quali è stata presa una determinata decisione. Ugualmente, rende difficile individuare l'autore dell'informazione sintetizzata sulla scorta delle categorie giuridiche tradizionali.

causati dall'AI godano di un livello di protezione equivalente alle vittime dei danni causati senza il coinvolgimento di un sistema di AI.

⁵² L'European Data Protection Supervisor nella sua recente Opinion (n. 43 - 11 ottobre 2023) si è mostrato critico sulla previsione delle garanzie procedurali soltanto a carico dei sistemi ad alto rischio.

⁵³ Cfr. M. BASSINI, L. LIGUORI, O. POLLICINO, *Sistemi di Intelligenza Artificiale, responsabilità e accountability. Verso nuovi paradigmi?*, in F. Pizzetti (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Giappichelli, 2018, pp. 356 ss., secondo cui «Proprio la presenza di dinamiche intrinseche al loro funzionamento che permettono alle macchine di apprendere nuove funzioni, acquisendo gradualmente un'autonomia sempre maggiore rispetto a chi ne ha programmato il software, segnerebbe un limite importante alla capacità di tenuta degli standard normativi esistenti».

La questione è legata anche alla forte suggestione che circola in vari ambienti di studio dell'AI, in cui si accredita l'idea dell'autonomia della macchina dotata di *self-learning*. L'equazione sottintesa è: "autoapprendimento", quindi, autonomia decisionale. Tali sistemi sono capaci di agire in totale autonomia, assumendo decisioni proprie, non prevedibili dai loro programmatori. Non può negarsi, per vero, che la sofisticata automazione di cui sono dotati questi sistemi può portare a considerare il risultato finale come prodotto in maniera 'autonoma rispetto' al controllo e all'intervento umano.

Tuttavia, spostando la luce dell'analisi verso la prospettiva del giurista dobbiamo svolgere una riflessione critica sulla percezione dell'autonomia delle macchine dotate di *self-learning*: l'autonomia di cui si discute è riferita alla capacità della macchina di adattarsi al mondo esperienziale e rispondere alle nuove informazioni raccolte, ma sempre entro i limiti e criteri predeterminati dai creatori umani⁵⁴.

Dunque, il modo di operare in maniera 'autonoma rispetto' al programmatore o utilizzatore è solamente l'obiettivo posto dagli stessi creatori, al fine di poter delegare alla macchina attività pericolose, faticose o time-consuming, ma non lo si può qualificare come un'espressione di volontà, intenzione o scopo, e quindi di libertà. Si tratta pur sempre della manifestazione di una forma, seppur molto sofisticata ed efficiente, di eteronomia, in assenza di una qualche coscienza e comprensione etica. Per tali ragioni, non c'è spazio per poter fondare nella macchina - e nei *large language models* - un'autonomia giuridica, una capacità di autodeterminazione che qualifica il soggetto di diritto⁵⁵.

⁵⁴ Cfr. C. O'NEIL, *Armi di distruzione matematica*, Firenze-Milano, 2017, 190. Sia consentito un rinvio a G. FASANO, *Dati personali e decisioni automatizzate: quale trasparenza*, in *MediaLaws - Rivista di diritto dei media*, 3, 2022, p. 417. In effetti, nella progettazione degli algoritmi la modellazione prende avvio con la scelta degli obiettivi, in tal senso potremmo dire che gli algoritmi sono ideologicamente orientati, secondo target di efficienza in linea con gli interessi di sviluppatori e progettisti, e agiscono secondo le istruzioni e le logiche incorporate durante la fase di sviluppo.

⁵⁵ Proprio sull'elemento della soggettività giuridica si fonda la decisione resa sul cd 'caso **DABUS**', una macchina che il suo inventore **definisce** creativa perché in grado di produrre opere originali e, in quanto tale, titolata a beneficiare delle norme sul copyright. Ma, il Copyright Office statunitense, prima, e la District Court del Distretto della Columbia, poi, hanno respinto la richiesta di registrazione come opera suscettibile di tutela, in quanto la creazione di opere dell'ingegno per potere godere della protezione del diritto d'autore deve

Di conseguenza, e questo è il punto fondamentale, quest'autonomia dal controllo e dall'intervento umano non può escludere i profili di responsabilità⁵⁶ che inevitabilmente restano in capo a chi fornisce (in senso lato) la macchina. Queste macchine non 'fioriscono' autonomamente ma vengono create, addestrate e supervisionate dagli esseri umani, per scopi ben precisi anche se spesso non dichiarati. La loro capacità di elaborazione, che solo per semplificare possiamo etichettare come capacità di autodeterminazione, non può essere separata dalle scelte umane intrinseche alla loro creazione e implementazione. In definitiva, l'autonomia dei *large language models* è strettamente vincolata alle decisioni delle persone che li hanno plasmati e a queste va riconnessa la conseguente responsabilità (sia di natura contrattuale, sia di natura extracontrattuale e persino di tipo penale).

La sfida, dunque, resta quella di superare lo scarto culturale che vede il concetto di autorialità legato alla sola categoria persona, un individuo che crea o origina qualcosa, come un libro, un dipinto, una musica, una teoria o qualsiasi altra forma di espressione artistica o intellettuale⁵⁷. Abbiamo disunito la capacità di comprensione dalla capacità 'tecnica' di produrre idee e opinioni, introducendo uno scarto evidente tra sapere e significato; abbiamo creato autori senza persona, senza diritti⁵⁸, senza libertà ma che, nei fatti, simulano l'esercizio di quella; abbiamo il compito, adesso, di individuare modelli normativi per gestire le responsabilità derivanti da questi nuovi modelli di autorialità.

essere frutto, almeno in parte, dell'ingegno umano. Dello stesso tenore le risposte dell'Ufficio della proprietà intellettuale del Regno Unito (IPO) e dell'Ufficio europeo dei brevetti (EPO).

⁵⁶ S. CHIODO, *Human autonomy, technological automation (and reverse)*, in *AI & Society*, 37, 39–48 (2022), <https://doi.org/10.1007/s00146-021-01149-5>, in cui l'autrice esplora il passaggio dalla definizione delle tecnologie in termini di automazione, alla definizione delle tecnologie in termini di autonomia come un cambiamento che «may mean that we are using technologies, and in particular emerging algorithmic technologies, as scapegoats that bear responsibility for us by making decisions for us».

⁵⁷ La normativa sul diritto d'autore, ad esempio, nata per proteggere i diritti degli autori sulle loro opere, è stata modellata dalla società e dalla tecnologia del periodo in cui è stata sviluppata, e riflette le concezioni culturali e filosofiche di quel tempo.

⁵⁸ La UK Supreme Court, nella recente sentenza del 20.12.2023 in ordine all'applicazione dei diritti di proprietà intellettuale in conformità al Patents Act 1977, ha concluso che «an inventor ... must be a natural person, and DABUS is not a person at all, let alone a natural person» (*Thaler v Comptroller-General of Patents, Designs and Trade Marks [2023] UKSC 49*).

Conclusioni.

La società dell'informazione ha vissuto negli ultimi anni una epocale trasformazione divenendo, grazie alle piattaforme online e, soprattutto, ai motori di ricerca e ai social network, una rete capillare di idee, pensieri e convincimenti liberamente circolanti, a cui si accede in modo economico, rapido ed equo, essendo facilmente raggiungibile dalla stragrande maggioranza delle persone.

L'espansione del web è stata favorita anche dalla dottrina del 'libero mercato delle idee'⁵⁹, cosicché la libertà di espressione del pensiero, di informare e di essere informati non sono più limitate, come avveniva coi tradizionali mezzi di informazione.

Il libero mercato delle idee ha prodotto criticità ormai note, legate alla disinformazione, al pluralismo informativo e alla concentrazione di potere, ma anche al sovraccarico di informazioni in circolazione, problematica a cui la tecnologia dei *large language models* vorrebbe dar efficace risposta. Si apre così un nuovo scenario: non più un libero mercato delle idee e, anzitutto, non più un mercato dalla cui competizione può emergere la forza del pensiero, ma una nuova fonte che distilla le informazioni singolarmente, per ciascun individuo, attraverso una sofisticata attività di addestramento e personalizzazione basata sulla profilazione della persona.

In effetti, se pensiamo allo stile comunicativo di molte di queste chat alimentate dai *large language models*, molto persuasivo, e se pensiamo all'ambiente in cui viene generata e veicolata l'informazione sintetica, quello confidenziale tipico di una chat, ci rendiamo conto di come sia difficile per la persona non affidarsi ai contenuti trasmessi. Se un'informazione raccolta con un motore di ricerca viene letta da molti acriticamente, ancora di più ciò accade se l'informazione è veicolata nel contesto di una comunicazione a due, con uno stile narrativo calibrato sul profilo della persona

⁵⁹ Nota dottrina del *free marketplace of ideas*, sviluppata in modo più completo da Justice J. Holmes, più di un secolo fa nel caso *Abrams v. United States*, 250 U.S. 616 (1919), secondo cui «the best test of truth is the power of the thought to get itself accepted in the competition of the market».

e corrispondente alle sue aspettative di conoscenza⁶⁰. Quindi, più convincente e seduttivo. In tale contesto si formeranno sempre più spesso idee e coscienze, incidendo non solo sulla libertà di essere informati, di formare un pensiero, ma sul diritto di 'spessore costituzionale'⁶¹ all'autodeterminazione.

La macchina è in grado di utilizzare il linguaggio naturale per informare e formare le persone, a seconda dello scopo derivante dallo specifico campo di applicazione. La macchina, sia detto a evitare equivoci, non esercita un diritto di libertà, ma – egualmente all'esercizio della libertà di espressione – ad essa deve corrispondere l'assunzione di doveri e responsabilità, posto che qualunque sia il suo scopo pratico esso è raggiunto prendendo parte all'esercizio della libertà all'informazione delle persone.

Non dal contesto dello specifico utilizzo deve trarsi una valutazione di rilevanza sociale dell'interesse all'informazione, quanto dall'attitudine delle informazioni sintetizzate a ricoprire una vasta gamma di argomenti e domini di conoscenza, dalla rapidità di confezionamento e di personalizzazione dei contenuti, dalla capillarità della loro diffusione, dall'assenza di un mercato di libero confronto e, in definitiva, dal soddisfare un "interesse generale" all'informazione.

L'ordinamento costituzionale possiede già solidi presidi atti a tutelare tale libertà e la giurisprudenza saprà interpretare ed attualizzare principi e regole vigenti rispetto a una realtà in continuo cambiamento sino a definire la responsabilità ultima che sta dietro alla "macchina da scrivere intelligente". Ciò nondimeno, sarebbe auspicabile che il legislatore (nazionale o europeo) si soffermasse, anche

⁶⁰ Vedi T. WU, *The attention merchants: from the daily newspaper to social media, how our time and attention is harvested and sold*. London, 2017, in cui l'autore sostiene che a causa della crescente quantità di informazioni e stimoli che ci bombardano quotidianamente attraverso i mezzi di comunicazione, abbiamo perso la capacità di concentrarci profondamente e di entrare in uno stato di flusso creativo.

⁶¹ Cfr. S. MANGIAMELI, *Autodeterminazione: diritto di spessore costituzionale?*, in C. NAVARINI (a cura di), *Autonomia e autodeterminazione - Profili etici, bioetici e giuridici*, Roma, 2011, in cui l'autore afferma che «la Costituzione non parla di autodeterminazione in alcuna delle sue disposizioni, e ciò nonostante che la nozione di persona, nelle sue diverse aggettivazione (personale, personalità), sia richiamata 21 volte, quella di uomo (o umano) ben 9 volte, in alcuni casi in reciproca combinazione (nell'art. 3, comma 2, e nell'art. 32, comma 2, Cost.) e la dignità è richiamata due volte (nell'art. 3, comma 1, come dignità sociale, e nell'art. 41, comma 2, come dignità umana)».

per ragioni di certezza del diritto, sui nuovi profili giuridici inerenti alle informazioni generate in modo autonomo. Non per assecondare l'abitudine di rispondere sbrigativamente a problemi nuovi con nuove regole (che innalzano l'inflazione dei diritti piuttosto che la loro tutela effettiva), ma perché c'è bisogno di definire nuovi modelli normativi che fissino i contenuti e i limiti dei concetti di autorialità e responsabilità⁶². E le tutele dell'art. 21 C., sotto il profilo del diritto alla informazione e alla libertà di formazione del pensiero, lungi dall'essere obsolete, hanno un potenziale di evoluzione normativa e sociale capace di coprire costituzionalmente anche le nuove tecnologie. Come è accaduto in passato per la radio e la televisione e, più di recente, per la rete, dovrà avvenire anche con riferimento ai *large language models*, attraverso un adattamento degli strumenti di tutela alla nuova dimensione.

⁶² Anche negli Stati Uniti, culla del *free market place of ideas*, si va discutendo la questione se l'intelligenza artificiale generativa sia protetta o meno dalla Sezione 230 del Communications Decency Act, la legge approvata dal Congresso nel 1996 secondo cui "Nessun fornitore e nessun utilizzatore di servizi Internet può essere considerato responsabile, come editore o autore, di una qualsiasi informazione fornita da terzi".