

Repliche alla Lettera

Costituzionalizzare i neurodiritti?

di Maria Romana Allegri

Alcuni giorni fa, leggendo il giornale (ovviamente online dallo smartphone) mi sono imbattuta in un articolo riguardante una influencer americana di nome Caryn Marjorie che, utilizzando l'intelligenza artificiale, ha creato un avatar di se stessa (CarynAI), addestrato per replicare la sua personalità, il suo carattere e la sua voce, con cui i follower possono interagire pagando un dollaro al minuto.

La realizzazione di CarynAI è stata affidata a una società americana (Forever Voices) specializzata nel replicare personalità e voce di personaggi famosi con le ultime tecnologie AI e di sintesi vocale. L'idea è quella di creare una sorta di "fidanzata virtuale", con la quale sia possibile instaurare un rapporto di intima confidenza, se non addirittura di amore. Lo slogan che accompagna il lancio di questo nuovo prodotto è il seguente: «Hai bisogno di qualcuno che ti ami e si prenda cura di te o semplicemente di una persona che ascolti come è andata la tua giornata? CarynAI sarà lì per te, pronta ad ascoltarti. Una conversazione nella quale non sarai condizionato da alcun limite». Incuriosita, ho cercato altri articoli sul medesimo argomento, scoprendo che esistono vari sistemi di AI, più o meno evoluti e sofisticati, che permettono di relazionarsi con "partners virtuali". Fra questi, ad esempio, l'applicazione per smartphone Replika, che permette agli utenti di creare e personalizzare un avatar, per «avere un amico senza giudizi, problemi o ansia sociale» o «avere una connessione emotiva, condividere una risata o parlare di qualunque cosa». Oppure Xiaoice, un sistema di intelligenza artificiale sviluppato da Microsoft Asia, in grado di comprendere le emozioni delle persone con le quali si interfaccia, comportandosi di conseguenza: pare che per milioni di cinesi Xiaoice sia la "ragazza dei sogni", con cui poter parlare anche di temi molto intimi, come la depressione, la malattia, il suicidio, la fine di una relazione o il sesso.

Alla lettura di queste notizie è fin troppo ovvio reagire con un misto di scetticismo (ma quanto potrà mai essere realistica la conversazione con un avatar?) e preoccupazione per il trattamento che subiranno le personali – anzi personalissime – informazioni rivelate confidenzialmente all'avatar nel corso di qualche conversazione intima e finite nelle mani di imprese commerciali o – ancor più grave – di esponenti politici o decisori pubblici che mirano a parlare "alla pancia" del popolo. Nelle interazioni con questi avatar amorosi ed empatici, infatti, c'è da aspettarci che le difese psicologiche individuali siano particolarmente basse e che, conseguentemente, cresca di molto il rischio che le persone subiscano condizionamenti e manipolazioni, dato il rapporto di fiducia, amicizia o amore che si instaura nella relazione virtuale. Se, infatti, nella vita reale è piuttosto comune subire l'influenza

positiva o negativa del proprio partner, a maggior ragione nelle relazioni sentimentali virtuali, data la fragilità psicologica di molte persone coinvolte in questo tipo di rapporti, unita alle incredibili capacità manipolative e predittive dei sistemi di AI, si possono condizionare pesantemente i comportamenti umani e le scelte di vita. Ciò potrebbe, nel tempo, assumere le dimensioni di un fenomeno non episodico ma di vasta portata, che potrebbe ripercuotersi non solo sulle scelte di consumo (neuromarketing), ma anche su decisioni fondamentali per la vita di milioni di individui, non limitandosi alla sola sfera intima e personale, ma arrivando a toccare la sfera pubblica e persino politica.

Insomma, mi sembra che uno dei più gravi rischi determinati dal rapidissimo sviluppo dei sistemi di AI – di cui i partners virtuali non sono che un esempio – sia quello relativo all'integrità della sfera psichica individuale, cosa che può avere ripercussioni anche sul piano della collettività. E per fortuna non sono la sola ad avere questo timore.

Il 22 ottobre 2020 l'Assemblea parlamentare del Consiglio d'Europa ha adottato la Risoluzione n. 2344, riferita specificamente ai sistemi BCI (*brain-computer interfaces*), cioè quei sistemi in grado di creare «a fully symbiotic connection between the human brain and digital computing systems», con il conseguente rischio di una «profound violation of individual privacy and dignity, with the potential to subvert free will and breach the ultimate refuge of human freedom – the mind». Tutto ciò potrebbe addirittura arrivare a cambiare «the very nature of humanity and human societies», motivo per cui si raccomanda la definizione di un quadro etico per lo sviluppo e l'utilizzo delle tecnologie BCI e si invitano gli Stati membri a introdurre – come ha tentato di fare il Cile nella nuova Costituzione, respinta però dal referendum popolare di settembre 2022 – strumenti giuridici di protezione dell'emergente categoria dei *neurorights*, definibili generalmente in termini di «cognitive liberty, mental privacy, mental integrity and psychological continuity».

Il Parlamento europeo, nella sua Risoluzione del 20 ottobre 2020 relativa agli aspetti etici dell'intelligenza artificiale ha sottolineato, fra le varie cose, la necessità di evitare qualsiasi impiego dell'AI «che potrebbe comportare un'inammissibile coercizione diretta o indiretta, minacciare di compromettere l'autonomia psicologica e la salute mentale o portare a una sorveglianza ingiustificata, a inganni o a inammissibili manipolazioni» (par. 29). Successivamente, nella Risoluzione del 3 maggio 2022 sull'intelligenza artificiale in un'era digitale, il Parlamento europeo ha evidenziato i rischi di questo tipo di manipolazioni per la tenuta dei sistemi democratici.

Ancora più esplicitamente, il “considerando” n. 16 della proposta di Regolamento europeo sull'intelligenza artificiale (COM(2021) 206 del 21 aprile 2021) suggerisce di «vietare l'immissione sul mercato, la messa in servizio o l'uso di determinati sistemi di IA intesi a distorcere il comportamento umano e che possono provocare danni fisici o psicologici». Conseguentemente, nell'elenco delle pratiche di IA vietate, indicate all'art. 5 della proposta di Regolamento, figura proprio al primo posto «l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che utilizza tecniche subliminali che agiscono senza che una persona ne sia consapevole al fine di distorcerne materialmente il comportamento in un modo che provochi o possa provocare a tale persona o a un'altra persona un danno fisico o psicologico». Al di là del pregiudizio sul piano individuale, però, mi sembrano più gravi le ripercussioni dei condizionamenti psicologici sul piano

collettivo e pubblico. Purtroppo a questo secondo aspetto la proposta di Regolamento europeo non presta altrettanta attenzione. Il “considerando” n. 70, poi, è dedicato ai sistemi di IA destinati all’interazione con persone fisiche, che possono comportare rischi specifici di impersonificazione o inganno; il rimedio proposto – sancito dall’art. 52 par. 2 – è quello della trasparenza, tale per cui le persone che interagiscono con sistemi progettati per il riconoscimento delle emozioni dovrebbero riceverne apposita notifica. Ma basta davvero una notifica per proteggere le persone dalla manipolazione psicologica ad opera di sistemi automatizzati?

Venendo alla dimensione costituzionale, i Costituenti hanno evidentemente ritenuto la psiche una faccenda del tutto privata e quindi non “costituzionalizzabile”, tanto da ingenerare un lungo dibattito dottrinale sul possibile fondamento costituzionale della libertà morale. Anche la Corte costituzionale ha valorizzato il profilo della libertà psicologica solo in termini funzionali alla piena realizzazione della democrazia, in relazione ai profili della libertà del voto e di formazione di un’opinione pubblica avvertita e consapevole attraverso il pluralismo informativo. Eppure, la sfera dei sentimenti non è estranea al lessico giuridico penalistico (si pensi ai riferimenti testuali all’odio, al timore, all’ansia, al comune sentimento della morale o del pudore ecc. presenti nel codice penale) né sono irrilevanti, sempre in ambito penalistico, gli stati emotivi e i condizionamenti psicologici, per via del loro impatto sulla formazione della volontà e quindi sulla graduazione della colpevolezza.

Tutto ciò considerato, sono sempre più convinta che il rapido sviluppo dei sistemi BCI imponga di considerare l’integrità della dimensione psichica ed emozionale come un bene giuridico meritevole di tutela costituzionale, seguendo l’esempio del Cile, primo Paese al mondo ad aver tentato di costituzionalizzare i neurodiritti. Altrimenti, lo strumento del diritto rappresenterebbe un argine assai debole dinanzi alla manipolazione psicologica realizzata attraverso il cosiddetto *affective computing*. E, comunque, dovremmo rassegnarci all’inevitabile integrale cessione al legislatore sovranazionale del potere di determinare l’estensione e i limiti della tutela della personalità individuale in ambiente Internet. Il ché, peraltro, sta avvenendo anche per la scelta – secondo me miope – di lasciare Internet al di fuori della Costituzione, nella convinzione che i principi costituzionali a tutela dei diritti individuali possano applicarsi tanto offline quanto online con qualche adattamento interpretativo. Invece, lo sviluppo della IA impone di ripensare la persona umana nella sua interezza, di ricollocarla in una diversa dimensione caratterizzata dall’inevitabile relazione con le “macchine sociali” dotate di funzioni cognitive (come nota A. Simoncini su Rivista AIC) e dalla traslazione di molte attività nel “metaverso”: continuare pervicacemente ad adattare principi tradizionali a pratiche radicalmente nuove significa, nel tempo, condannare la Costituzione alla marginalità.

Alcune questioni di metodo e di merito

di Francesco Cirillo

Vorrei sottolineare due questioni di metodo, relative alla definizione dell'oggetto e alle condizioni della regolamentazione, e due questioni di merito, relative all'impatto delle tecnologie. Quanto alle prime, mi riferisco (1.1) al problema definitorio dell'IA e (1.2) alle condizioni a cui la Costituzione possa rivolgersi a tale orizzonte. Quanto al merito, ai rischi delle tecnologie (2.1) in una dimensione collettiva e (2.2) individuale.

1.1. È tuttora in corso nelle scienze cognitive un dibattito su cosa sia un'IA e su cosa sia lecito attendersi che faccia. Non è chiaro, cioè, se l'IA sia solo il *nomen* di un programma di ricerche riconducibili al proposito di replicare alcune funzioni cognitive (soprattutto quelle *computazionali*) in una macchina, o se essa possa ambire ad essere un risultato finale (e non un mero processo). Tale risultato potrebbe, p.e., essere raggiunto superando il test di Turing; e, però, per alcuni ciò non sarebbe sufficiente (cfr., per tutti, la stanza cinese di Searle).

Seppur vero che *à la* Carnelutti il nostro sguardo debba spingersi oltre il diritto, lì non si danno risposte pronte, ed è precluso al giurista – perché esterno ai suoi metodi e al suo oggetto – di assumere posizioni arbitrarie.

Nessuna tecnologia, per quanto straordinaria, consente in modo incontestabile di trattare di una “intelligenza”. Siamo di fronte a dispositivi che, se rivolti a problemi *semplici*, non ci appaiono *intelligenti* e, se rivolti a problemi *difficili* (p.e. le reti neurali profonde), ci sembrano insondabili e diversi dall'essere umano. Allora, non avrebbe senso dire che le macchine *pensano*, così come non ha senso – se non figurato – dire che le automobili *corrono*. Ci rivolgiamo piuttosto a un insieme articolato di applicazioni, dirette a replicare capacità sensoriali, agire nell'ambiente, imitare funzioni cognitive, o a raggiungere risultati computazionali (senza neppure imitare il pensiero). Un insieme per cui sarebbe impensabile una sola regolamentazione, non meno di quanto lo sarebbe per la tecnologia *tout court*.

Questa difficoltà si traduce spesso, nel campo dell'etica dell'IA e nelle proposte regolatorie, nell'enunciazione di un imperativo generico: che la tecnologia si sviluppi nel rispetto dei diritti umani, che essa sia *human centered*, come nelle leggi di Asimov. Ciò appare condivisibile per ogni attività umana e, però, insoddisfacente per il giurista.

1.2. L'impatto delle tecnologie interessa anche la riflessione costituzionalistica, ma – come prima – il ventaglio dei problemi è ampio. Ci si può soffermare su due profili critici delle garanzie costituzionali, sia in ordine all'ambito di applicazione che alla loro *natura*.

In alcuni casi le tecnologie sembrano compromettere la concezione del territorio, ma ciò non è sempre vero: il problema non è identico nel caso dei *chatbot* o della robotica. Possiamo sempre

immaginare un intervento regolatorio solo nazionale, ma nell'ipotesi di attività per lo più dislocate nell'*ambiente digitale*, questo sarebbe agevolmente aggirato; diversamente non accadrebbe quanto ai dispositivi dell'*ambiente analogico*.

Per altro verso, molte tecnologie sono adottate nell'erogazione di servizi o prodotti tra privati, sia pure tra *stakeholder* e consumatori. Sul punto, è indiscusso che le garanzie costituzionali si estendano anche ai rapporti privati: la costituzionalizzazione del diritto privato, l'emersione dei diritti della personalità, l'ipotesi della *Drittwirkung*, ecc.

Molto meno chiaro, però, è fino a che punto ciò accada. Si pensi all'indisponibilità della libertà morale donde il divieto di tecniche di *lie detection* nel processo (art. 188 c.p.p.), cui si oppone un'accettazione delle stesse tecnologie in un contesto *extraprocessuale* (sono in commercio macchine della verità con finalità ludiche; con usi nei programmi televisivi, ecc.). Ancora, alla preoccupazione per il trattamento dei dati sanitari nella pandemia cui non corrisponde un'analogia attenzione per il trattamento disinvolto da parte di altre applicazioni fruibili dall'utente.

L'aspetto più problematico dell'efficacia dei diritti costituzionali tra i privati consiste proprio nella perdita progressiva della loro indisponibilità a vantaggio di un regime più flessibile. La stessa autodeterminazione informativa incontra criticità rilevanti, se si osserva che la gestione del dato da parte del titolare del trattamento possa tradursi nei fatti in una perdita del controllo da parte dell'interessato. E tutto ciò in un contesto in cui il consenso è condizione di accesso ai servizi, senza che però si dia un particolare impegno cognitivo nella sua manifestazione.

La dimensione costituzionale, allora, va ricostruita tenendo conto della tensione verticale verso lo spazio sovranazionale che può regolare (soprattutto) alcuni fenomeni, e della tensione orizzontale che impone di riflettere sulle condizioni a cui i diritti si possano garantire tra privati. Nessuno dei due fronti può eludere il problema del rapporto tra diritti e interessi opposti (e del bilanciamento); una questione spesso aggirata dal ricorso a una visione irenica dei diritti fondamentali, che trascura il dato teorico e fattuale del loro frequente conflitto.

2.1. Venendo all'impatto delle tecnologie, alcuni rischi sono osservabili soprattutto in un orizzonte collettivo.

Nel provvedimento del Garante di limitazione del noto *chatbot*, si osserva che «le informazioni fornite [...] non sempre corrispondono al *dato reale*»: un profilo che non attiene al solo diritto soggettivo all'informazione, quanto piuttosto ai rischi per una concezione *pubblica* della verità. Analogamente, si pensi allo sciopero della Writers Guild of America, o al dibattito sui *chatbot* nelle scuole. La questione è collettiva proprio perché non si dubita del legittimo ricorso a un *software* che scriva sceneggiature, risolva equazioni o risponda a domande di geografia; il problema sorge soltanto se molti adottano tale tecnologia, con possibili conseguenze *sociali*.

Allo stesso modo, le bolle informative indotte dalla profilazione delle piattaforme rincorrono e rinforzano i *bias* dell'individuo: nessuno nega il diritto di ciascuno di leggere un sol genere di quotidiani o di libri; nondimeno, sembra che la tecnologia sia poco trasparente nel far percepire le operazioni di profilazione e le scelte dei contenuti che essa destina all'utente; quindi, che l'esercizio di un diritto (indiscutibile per il singolo) provochi un effetto negativo, percepibile in una dimensione collettiva.

Non si vuole sostenere che siano questioni prive di ricadute sui diritti fondamentali, ma che la dimensione collettiva è più idonea a rilevare alcuni rischi e a delineare strategie di regolamentazione.

2.2. Infine, ci si interroga su possibili cataloghi di “nuovi” diritti. Si prenda in considerazione, per tutti, l’interferenza della tecnologia nella *psiche*, che indurrebbe a prospettare nuovi neurodiritti. Il tema, chiaramente, in parte sconfinava rispetto al perimetro dell’AI (perché abbraccia anche l’intervento sul corpo).

Si possono distinguere nelle attuali proposte alcune classi di diritti: una prima si rivolge all’integrità (psichica, della mente, ecc.); una seconda alla *privacy* (della mente, del cervello, ecc.); una terza alla libertà (della mente, cognitiva, o al libero arbitrio, ecc.); una quarta all’identità e alla continuità (psichica, psicologica, ecc.); una quinta all’accesso al potenziamento (neurale, cognitivo, ecc.).

Il giudizio complessivo, qui forse apodittico, è che alcuni diritti siano ricompresi in quelli riconosciuti (in certi ordinamenti): l’integrità psichica, l’identità, l’equo accesso alle prestazioni sanitarie (se il neuropotenziamento è ricondotto a tale *genus*); perciò non si richiederebbe un ulteriore riconoscimento.

Invece, altri neurodiritti farebbero rinvio a “oggetti” molto eterogenei: libertà della *psiche*, libertà *cognitiva*, diritto al *libero arbitrio*, diritto alla *continuità psicologica*, *privacy* della *mente*, ecc. Sembra inconferente il riferimento al solo livello *neurale*, perché si ricorre a una semantica che esclude l’approccio neuroriduzionistico. Nel caso del *libero arbitrio*, della *continuità psicologica* o della *mente*, il bene oggetto del diritto sarebbe perciò inafferrabile: massima parte delle neuroscienze “scommette” sulla riduzione fisicalista della dimensione cognitiva e sull’inconsistenza di tali concetti.

Allo stato, comunque, non è possibile acquisire soluzioni al dilemma del libero arbitrio o della coscienza. Tuttavia, se si assume un approccio *fisicalista*, la mente e il libero arbitrio non possono poi figurare come bene oggetto di un diritto; se, invece, si accedesse a una prospettiva *mentalista*, l’interferenza delle tecnologie non dovrebbe rappresentare un rischio reale per la *psiche*.

Inoltre, la definizione dei neurodiritti investe la stessa dogmatica del diritto soggettivo, ponendo la questione se il libero arbitrio sia il fondamento dei diritti di libertà (un postulato necessario e indimostrabilmente presente in ciascuno) o se la libertà giuridica non dipenda (o non coincida) con la libertà *filosofica* del volere. Si tratta di questioni aperte, soprattutto perché le scienze cognitive non offrono risposte e, anzi, consegnano un acceso dibattito: d’altronde, la regolamentazione delle neurotecnologie non può attendere la soluzione di simili dilemmi.

Il problema potrebbe essere affrontato, forse, facendo ricorso al concetto (provvisorio) di *neuroprivacy*, che avrebbe il pregio di abbracciare – nella sua genericità – lo sviluppo delle ricerche: postulando l’intuitiva esigenza di riservatezza e di protezione da interferenze esterne; e facendo riferimento, in mancanza di altre e più precise distinzioni, al livello neurale, che sembra più interessato dalle neurotecnologie.

Sono alcune delle questioni con cui confrontarsi, che ricordano, come osservato in apertura del dibattito, «le fatiche di Sisifo»: potrebbero essere necessarie risposte rapide, ma non perciò frettolose; soluzioni precarie e sperimentali, ma non anche sbrigative.