L'IRCOCERVO
PRIMA RIVISTA ELETTRONICA ITALIANA DI METODOLOGIA GIURIDICA - TEORIA GENERALE DEL DIRITTO E DOTTRINA DELLO STATO
FIRST ITALIAN DIGITAL JOURNAL OF LEGAL METHODOLOGY - GENERAL THEORY OF LAW AND DOCTRINE OF THE STATE

Macchinazioni.
Soggettività, agentività ed entità esotiche*

Salvatore Amato

Università degli Studi di Catania

Abstract: **Machinations. Subjectivity, Agency and Exotic Entities**
Opacity is one of the main problems of artificial intelligence not only because it is technically difficult (and in some cases impossible) to open the black box, but because it constitutes the borderline zone between mechanisation of mind and humanisation of machine. This zone was the extreme outcome of reducing brain activity entirely to the biochemistry of neurons or the biophysics of synapses. The black box effect is the other side of the same coin: a kind of possible humanisation of machines, increasingly complex and unpredictable.
Is human-centred artificial intelligence a return to natural law? Do ethics by design, info-ethics, algor-ethics, data ethics, data dignity reintroduce the principles and values of the natural law tradition?

Keywords: Expandiverse Technologies, Deep Learning, Opacity, Alien Reasoning, Natural Law.

Sommario: 1. L'inconscio digitale – 2. La scatola nera dell'apprendimento profondo – 3. *Exotic mind entities* – 4. Un diritto naturale al silicio.

"La macchina si adatta alla debolezza dell'uomo,
per fare dell'uomo debole una macchina".
K. Marx, *Manoscritti economico-filosofici.*

1. L'inconscio digitale

Se già nel medioevo, con meraviglia e preoccupazione, si affermava che "ogni giorno si scopre una nuova tecnica"[1], oggi sono gli aggettivi che ci mancano, *computing machine*, *business machine*, *learning machine*, *smart machine* per indicare non solo come le macchine sostengono le nostre azioni, ma soprattutto in che modo e fino a che punto influenzino i nostri pensieri, condizionino la nostra

* Acknowledge financial support from: PNRR MUR project PE0000013-FAIR.
[1] Lo ricorda C.M. Cipolla, *Le macchine del tempo. L'orologio e la società 1300-1700*, il Mulino, Bologna, 1996, p. 16.

cultura, determinino il nostro modo di essere. Sono *ingenia*, frutto del nostro ingegno, ma anche *search engineering*, *neuromorphic engineering*, costruttrici del proprio ingegno. A volte incapaci di rispondere alle nostre attese: *underfitting*. A volte riluttanti a produrre nuovi sviluppi: *overfitting*. A volte incontrollabili nelle procedure e imprevedibili nei risultati: *black box effect*. In ogni caso *mêchanízein* e alle loro macchinazioni stiamo affidando la nostra organizzazione sociale.

In media su internet si registrano ogni minuto 575mila tweet, 5,7 milioni di ricerche su Google, 65.000 foto su Instagram. Solo una parte minima di questi dati viene conservata allo stato grezzo, la maggior parte è rielaborata per gli scopi più vari, in genere commerciali, e comunque conservata per essere sovrascritta con dati più recenti. Abbiamo, quindi, un effetto moltiplicatore di metadati sui dati e di metadati sui metadati. Pare siano otto miliardi e seicento milioni gli oggetti connessi al *web* che ci assistono, senza che ce ne rendiamo conto, nelle nostre più svariate attività. Treni e rotte ferroviarie, aerei e rotte aeree, denaro e transazioni finanziarie, messaggi e sistemi di comunicazione, sono regolati da processi di intelligenza artificiale: oggetto delle nostre richieste e intanto artefici di un mondo a sé che si accresce continuamente a un ritmo vertiginoso.

Tra non molto si dovrebbero produrre intorno ai mille triliardi di byte. Siamo avviluppati dal *cloud*, ma anche aggrappati al *cloud*… L'Internet delle Cose (IoT) ci segue passo per passo nel nostro rapporto con il mondo esterno. L'Internet of Bodies (IoB) ci monitora sistematicamente nel rapporto con i nostri equilibri biologici. Fuori di noi e dentro di noi aleggia un custode digitale che capta anche quello che ci sfugge e registra anche quello che non ci interessa.

Non sapremmo cosa farcene di questo continuo flusso di dati, se non avessimo supercomputer la cui capacità di elaborazione non è neppure immaginabile da una mente umana: 200 milioni di miliardi di calcoli al secondo; quanto potrebbero fare tutte le persone sulla terra se eseguissero un calcolo ogni istante di ogni giorno per 305 giorni. È una capacità che diverrà irrisoria se riusciremo a realizzare efficienti computer quantistici che, basandosi sui quantum bit, o "qubit", operano nel cosiddetto stato di "sovrapposizione". Una coppia di bit normali può dar vita solamente a una delle quattro tradizionali combinazioni, o stati, possibili (00, 01, 10 e 11), ma una coppia di qubit contiene tutte e quattro le combinazioni insieme. Sycamore, il computer quantistico annunciato da Google, impiega 53 qubit per un totale di 253 valori pari a più di 10 biliardi di combinazioni. Dovrebbe in 200 secondi effettuare quelle operazioni che gli attuali supercomputer all'avanguardia possono compiere in 10.000 anni[2].

Se mettiamo assieme la capacità di assimilazione dei dati con la capacità di rielaborazione, la complessità si complessifica nel senso che non esistono solo tanti specifici dispositivi, ma ciascun dispositivo ne sviluppa altri e poi ancora altri finché non si determina una concatenazione strutturale che tende a divenire un

[2] È difficile ipotizzare le implicazioni della fine dell'era digitale per l'avvento dell'era quantistica. Ci prova M. Kaku, *Quantum supremacy: how the quantum computer revolution change everything*, Penguin, New York, 2023.

mondo a sé. Ogni macchina è "imperialistica" o "espansionistica"[3] perché si procura un proprio regno coloniale di servizi, sosteneva, già nel 1964, Gunther Anders, riflettendo sulla tendenza alla "macchinizzazione". A sua volta Jacques Ellul, nel 2004 aveva invitato a riflettere sul fatto che il sistema tecnico non solo cresce su se stesso, ma questo continuo auto-accrescimento sfugge a qualsiasi possibilità di controllo: "la direzione tecnica si decide da sé"[4].

Colonizzazione e autoaccrescimento sono ormai gli sviluppi strutturali dell'*expandiverse technology* da cui deriva l'infosfera. Applicazioni e dispositivi, per quanto diversi, convergono (paradigma della confluenza) verso una funzionalità unitaria che spesso non è progettata né voluta, ma è effetto dell'alleanza commerciale tra *Big Tech* e *startup* attraverso la quale le imponenti infrastrutture *hardware* (tra *cloud* e potenza di calcolo) hanno accesso a enormi quantità di dati. Non esiste una netta linea di demarcazione tra il *Machine Learning*, l'Apprendimento Profondo, la *Computer Vision*, il *Natural Language Processing* (NLP), la Robotica, la Simulazione del Comportamento Umano, le neuroscienze cognitive, le neuroscienze sociali e così via. Ciascuna di queste attività ha le sue caratteristiche, ma gli sviluppi dell'una incidono sugli sviluppi delle altre in un calderone di funzioni in cui tutto si accumula senza limiti in un disordine che è solo apparente perché per la prima volta nella storia dell'umanità:

1) disponiamo di tecnologie che possono agire regolarmente e normalmente come utenti autonomi di altre tecnologie;

2) abbiamo macchine che continuano a perfezionarsi indipendentemente dal compito per cui sono programmate;

3) il prodotto di questi sviluppi e interazioni è spesso assolutamente imprevedibile.

Come osserva Floridi, "la modernità è diventata ben presto l'universo di complesse dipendenze reticolari, di reazioni a catena meccanica così come di connessioni obbligate"[5]. I dati si sommano gli uni con gli altri perché le tecnologie si sviluppano le une con le altre: c'è sempre l'algoritmo di un algoritmo che rinvia ad altri algoritmi. L'effetto "*loop*" dei meccanismi ricorsivi del nostro cervello, per cui i pensieri derivano dai pensieri nei pensieri dei pensieri, è la tecnica propria anche dei linguaggi dei sistemi di calcolo. Cambia la capacità di *looping*, ma non l'imprevedibilità. Attraverso l'esperienza che ha maturato come *chief business officer* di Google, dopo aver lavorato in IBM e Microsoft, Mo Gawdat ci fa notare che

> le AI non vengono propriamente programmate. Benché sulle prime siano alimentate da algoritmi, analoghi ai semi dai quali germoglia l'intelligenza, le

[3] G. Anders, *Noi figli di Eichmann*, trad. it., Giuntina, Firenze, 1995, pp. 53-55.
[4] J. Ellul, *Il Sistema tecnico. La gabbia delle società contemporanee*, trad. it., Jaca Book, Milano, 2009, p. 279.
[5] L. Floridi, *La quarta rivoluzione industriale. Come l'infosfera sta trasformando il mondo*, trad. it., Raffaello Cortina, Milano, 2017, p. 31.

> loro vere capacità scaturiscono dall'osservazione. Una volta corredate del codice iniziale, analizzano enormi quantità di dati per rilevare pattern e modelli, seguendo un percorso simile a quello della selezione naturale, per sviluppare il proprio ingegno in erba. Infine diventano pensatrici originali e indipendenti, influenzate meno dagli input dei loro creatori che dai dati ricevuti da noi[6].

Gli algoritmi dei sistemi di apprendimento automatico, strutturati sul modello di reti di neuroni artificiali, "capiscono" dagli stessi dati ciò che devono fare attraverso l'architettura di svariati "trasformatori generativi pre-addestrati", i *learners* (apprenditori) e i *classifiers* (classificatori), che sono in grado, con un programma di apprendimento lungo poche centinaia di righe, di generare automaticamente milioni di stringhe di codice. Il rapporto tra dati e metadati diventa, così, sempre più labile perché la loro fusione determina sviluppi ed esiti che non sono prevedibili neppure dagli stessi programmatori (*black box effect*).

L'idea che le macchine siano a nostra disposizione si sta, quindi, lentamente capovolgendo nella situazione opposta: siamo noi a dipendere sempre più dalle macchine. "Dai motori di ricerca di Google ADS ai sistemi di personalizzazione e raccomandazione di Instagram e YouTube, Spotify Apple Music o Amazon, dalle chat BOT agli algoritmi di filtraggio delle applicazioni di incontri, tu, io e gli altri siamo tutti cavie che corrono alla cieca nel labirinto"[7]. Quante cose abbiamo rivelato ad *Alexa* di Amazon, a *Siri* di Google, a *Cortana* di Windows, oppure più semplicemente alla tastiera del nostro computer o ai messaggi del nostro cellulare? Non ci interessa. Pensiamo che tutto passi come la varietà dei pensieri che attraversano la nostra mente. Invece avviene proprio il contrario. Parafrasando un noto detto di Trockij, "anche se la guerra non ti interessa, la guerra si interessa di te", potremmo dire che, anche se quello che memorizzano Google o Facebook o Instagram non ci interessa, loro si interessano di noi, perché non dimenticano nulla: ogni dato potrebbe essere conservato, catalogato, collegato, impiegato.

Si parla, infatti, di una sorta di "inconscio digitale" per indicare tutta le informazioni su ciascuno di noi che sono raccolte dai sistemi di rilevazione dell'economia digitale che sfruttano il nostro "surplus comportamentale" per elaborare le proprie previsioni di mercato[8]. Il controllo dei corpi, delle propensioni e in parte anche dei sentimenti è sempre più in mano ad algoritmi[9]. E gli algoritmi sono al servizio del mercato, che a sua volta è condizionato dagli algoritmi che producono quell'entità impalpabile e onnipresente che ci avvolge nel *cloud* e ci

[6] M. Gawdat, *Superintelligenti. Come salvare il mondo dall'intelligenza artificiale*, trad. it., Rizzoli, Milano, 2022, p. 205 (ed. digitale).
[7] *Ivi*, p. 215.
[8] S. Zuboff, *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, trad. it., Luiss University Press, Roma, 2019, Part III, cap. 12.
[9] D. de Kerckhove, *Il futuro della memoria*, trad. it., Castelvecchi, Roma, 2018.

segue con l'intelligenza artificiale. Un nuovo *apparatčik* sotto la forma di dispositivi *wireless* onnipresenti[10]?

2. La scatola nera dell'apprendimento profondo

Infosfera, cloud, intelligenza artificiale sono le metafore con cui cerchiamo di identificare la complessa aggregazione di materiali e attività che costituisce l'attuale ecosistema produttivo. In un versante abbiamo fibre ottiche, chip e microchip, cavi, sensori, satelliti, computer, magazzini, megastrutture informatiche, collegate in rete a particolari catene di approvvigionamento, vincolate al consumo di ingenti quantità di acqua e di energia e, infine, intente a soddisfare miliardi di utenti. In un altro versante troviamo gruppi di ricerca, sviluppo e produzione sorretti da una miriade di brevetti e da un affastellarsi di norme nazionali e internazionali. Possiamo elencare materiali e strutture, isolare legami commerciali e rapporti contrattuali, ma continua a sfuggirci la compiuta percezione degli effetti sistemici dell'entità che si sta progressivamente delineando. L'accelerazione tecnologica ha profondamente trasformato noi stessi e la nostra società, ma non ha cambiato il nostro rapporto con la tecnologia digitale, perché continua a prevalere l'idea che ci troviamo di fronte a tutta una serie di oggetti di consumo di cui dobbiamo soltanto regolamentare le modalità di impiego.

Infosfera, *cloud*, intelligenza artificiale, nella loro vaghezza semantica, sono la misura di questo limite concettuale. Sono il segno della difficoltà nell'individuare e nel delimitare quanto sta avvenendo. Le "tecnologie esponenziali del technocapitalismo" ci pongono di fronte alla "elaborazione di sistemi destinati a rispondere col tempo a ogni circostanza esistenziale e a istituire una gestione automatizzata del mondo"[11]. Anche se rifuggiamo dall'ipotesi estrema della *Singularity*, della *Superintelligence*, della *Scary Smart*, non sappiamo fino a che punto l'inevitabile sviluppo dell'automazione, effetto di questo enorme accumulo di informazioni e dispositivi, sarà in grado di ricondurre la connettività ad unitarietà, la coerenza a intelligenza generale, e l'intelligenza generale a una qualche forma di senzienza se non a un prototipo di coscienza. Non lo sappiamo perché non conosciamo quali siano le leggi fondamentali del pensiero e come queste si colleghino al funzionamento del cervello. Sappiamo che il cervello ha una composizione biochimica e quindi dovrebbe sottostare alle leggi della fisica. Sappiamo che tutti i cervelli sono modulari e quindi si articolano attraverso una rete di connessioni come qualsiasi sistema informatico. Tuttavia né la meccanica newtoniana, né la teoria elettromagnetica di Maxwell, né la teoria quantistica sono

[10] A. Keen, *Digital Vertigo: How Today's Online Social Revolution is Dividing, Diminuishing and Disorieting us*, St. Martin's Press, New York, 2012, cap. 7.
[11] E. Sadin, *La siliconizzazione del mondo. L'irresistibile espansione del liberalismo digitale*, trad. it., Einaudi, Torino, 2019, p. 40.

ancora in grado di spiegarci come la chimica diventi coscienza negli scambi tra i nostri neuroni. E non conosciamo neppure il ruolo della genetica in tutto questo.

Allo stato attuale dobbiamo limitarci a prendere atto di come sia sempre maggiore il numero di sistemi in grado di svilupparsi autonomamente e con crescenti margini di imprevedibilità (l'effetto scatola nera). E più questi sistemi operano in connessione, più il limite dell'uno viene colmato dalle capacità dell'altro. La robotica, ad esempio, non è solo un campo limitato a singole macchine, che eseguono ripetitivamente sempre gli stessi compiti prefissati. Per effetto del *cloud computing*, si sta trasformando in un'entità collettiva, una rete diffusa di macchine che condividono conoscenze ed esperienze, moltiplicando in maniera esponenziale, la capacità di adattarsi ai cambiamenti, risolvere problemi e trovare nuove vie di sviluppo. Se non sarà mai intelligente il singolo computer o robot, non sappiamo fino a che punto potremo dire lo stesso del *cloud* a cui è connesso e che, in qualche modo, lo regola e assimila.

Il dubbio è giustificato dall'analogia con il funzionamento del cervello dove il singolo neurone, anche se all'interno di una rete nervosa, non è "consapevole" e non produce "consapevolezza", mentre l'esperienza cosciente che caratterizza la nostra mente deriva dalla complessità del sistema di interazioni dei neuroni tra loro e con l'ambiente esterno, l'ancora in gran parte ignoto rapporto tra *pattern* di *firing* neuronali e *binding* di questi *pattern*. Risale già alla metà del secolo scorso la convinzione che, per effetto della ridondanza dei sistemi di calcolo, l'auto-organizzazione potesse essere la base della progettazione di una "*infallible network of fallible neurons*"[12].

Senza cedere alle suggestioni dell'esplosione computazionale di una super intelligenza, dobbiamo renderci conto che l'incremento delle possibilità determinato dalle tecnologie trasformative rende sempre più evidente la divisione teorica tra le due anime dell'intelligenza artificiale: quella ingegneristica e riproduttiva, che tende a imitare e perfezionare, attraverso gli impulsi che riceve dai programmatori, le condotte umane, assistendole o sostituendole in un gran numero di contesti; quella cognitivista e produttiva che aspira ad ottenere l'equivalente del nostro cervello con sistemi di apprendimento automatico (*machine learning*) e/o con sofisticati artefatti biologici[13]. Non si tratta soltanto di una scelta ideologica tra realismo e futurismo, tra techno-ottimismo e techno-pessimismo, tra *Effective Accelerationism* ed *Effective Altruism*, ma delle due vie che abbiamo di fronte: entrambe necessarie nella definizione dell'orizzonte etico e nell'elaborazione delle linee regolamentari. Dobbiamo certamente muovere dall'una, l'anima ingegneristica e riproduttiva, senza tuttavia rinunciare a tenere sullo sfondo l'altra, l'anima cognitivista e produttiva.

[12] M. Pasquinelli, *The Eye of the Master. A Social History of Artificial Intelligence*, Verso, London, 2023, p. 148.
[13] L. Floridi, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, trad. it., Raffaello Cortina, Milano, 2022, pp. 48 ss.

La prima prospettiva non nega i cambiamenti negli attuali equilibri sociali e politici, dal mondo del lavoro all'esercizio della democrazia, dall'economia alla finanza, dalla tutela dei diritti fondamentali alla giustizia digitale, ma ritiene che sia sufficiente un progressivo adeguamento degli schemi giuridici e dei modelli di *governance*, restando tuttavia nel solco di consolidate tradizioni culturali dove il libero arbitrio è ancora il fondamento della responsabilità individuale, dove la tutela della *privacy* è ancora il presidio della dignità individuale, dove le consultazioni elettorali sono ancora la radice della democrazia, e così via. I sostenitori di questa prospettiva affermano che non dobbiamo lasciarci ingannare dalla contiguità degli orizzonti di ricerca. Per quanto le neuroscienze suggeriscano modelli di sviluppo all'informatica e l'informatica aiuti a verificare i progressi nella conoscenza della codifica dei segnali nervosi e del modo in cui le reti neurali elaborano l'informazione, le macchine resteranno macchine e gli algoritmi algoritmi. Dietro le une e gli altri sarà sempre riconoscibile il primato dell'impronta umana. Malgrado il perfezionarsi della neurorobotica e il sistematico impiego di interfacce computer-cervello sarà sempre evidente la linea di demarcazione tra soggetto e oggetto, organico e inorganico, naturale e artificiale.

La seconda prospettiva pensa che le implicazioni di una tecnologia *expandiverse* non consentano, sin da ora e ancora meno in futuro, di capire dove finisce l'intelligenza ristretta e specializzata della singola macchina e dove cominci lo "apprendimento associato illimitato" che costituisce il presupposto di una futura intelligenza generale e indeterminata. È innegabile che la singola macchina sia "stupida", ma non sappiamo se potremo affermare lo stesso del "sistema" di macchine che si sta profilando all'orizzonte. Sono "stupidamente intelligenti" perché fanno tantissime cose, ma non sanno cosa fanno? Il problema è che incominciamo a non sapere neppure noi cosa faranno, perché si continuano a sviluppare sistemi "*human out of the loop*" per raggiungere risultati imprevisti o addirittura imprevedibili, perché i sistemi di *Machine Learning* sono progettati in modo da assimilare dati, prendere decisioni o fare previsioni senza che siano esplicitamente istruiti per farlo.

Del resto i programmi di ricerca avanzata si basano sulla centralità di *Answer box* che uniscono sofisticati modelli di archiviazione a nuove tecniche di indicizzazione e di elaborazione del linguaggio naturale non solo per fornire la risposta alla domanda dell'utente, ma per capirne le intenzioni. Tra *Knowledge Graph* e *Passage Ranking* lo scopo non è sempre quello di fornire una risposta corretta, ma quella risposta che l'utente desidera. I sistemi di intelligenza artificiale non possiedono quel "sapere di sfondo" sulla società in cui operano e sui pregiudizi da cui è animata, che è proprio degli esseri umani, ma si stanno affinando le tecniche di apprendimento in contesto (*Chain-of-Thought Prompting*) e di *Representational similarity analysis* per cogliere le sfumature implicite tra le pieghe delle analogie all'interno dei dati raccolti. Pressate tra *rich answers*, *direct answers*, *instant answers*, *quick answers*, *featured snippets*, *knowledge panel*, *chain-of-thought prompting* le applicazioni a cui ci rivolgiamo stanno scardinando nel nostro rapporto con i sistemi informatici le barriere tra sintassi e semantica, tra reazione e

cooperazione, tra programmazione e coevoluzione fino ad avvicinare l'interazione alla "sensibilità". È l'ideale a cui aspirano gli algoritmi affettivi o i *companion* robot e verso cui ci sospinge l'apprendimento profondo (*Deep Learning*), che utilizza reti neurali artificiali, la cui struttura tenta di avvicinarsi alla complessità dell'organizzazione del cervello.

> I computer addestrati con l'apprendimento profondo possiedono delle reti neurali enormi, in cui la conoscenza viene distribuita tra migliaia di pesi di connessione. Inoltre, alcune delle caratteristiche mentali sono difficili da identificare persino nei codici espliciti, perché richiedono interazione tra moduli diversi all'interno del programma complessivo del computer[14].

Assieme alla complessità cresce l'opacità[15]. Il modello ingegneristico e riproduttivo presuppone la prevalenza degli algoritmi deterministici in cui l'intelligenza artificiale è vincolata a specifici parametri di azione, a definite funzioni e a chiari obiettivi, attraverso un rapporto tra *input* e *output* regolato da "schemi esecutivi", accessibili e trasparenti, che indichino i singoli passaggi, le loro connessioni e gli esiti (modello *white box*). Gli algoritmi deterministici coprono, però, solo una parte, per quanto rilevante, dei processi digitali. Nella maggior parte dei casi le intelligenze artificiali, una volta corredate di un codice iniziale, sviluppano più modelli in concorrenza tra loro, seguendo un percorso simile a quello della selezione naturale nell'incrementare le proprie potenzialità. Questa tecnologia di apprendimento automatico genera quell'effetto *black box* che non consente di spiegare come e perché un dato diviene metadato, come e perché i metadati filtrano i dati, segnando il passaggio dalla "realtà aumentata" alla "realtà virtuale". Non siamo neppure in grado di capire quando gli algoritmi deterministici di un *white box model* mutino negli algoritmi stocastici di un *black box model*.

> *The specific characteristics of many AI technologies, including opacity ('black box-effect'), complexity, unpredictability and partially autonomous behaviour, may make it hard to verify compliance with, and may hamper the effective enforcement of, rules of existing EU law meant to protect fundamental rights*[16].

[14] P. Thagard, *Cervelli a confronto. Perché l'intelligenza umana è diversa da quella degli animali e dei robot*, trad. it., FrancoAngeli, Milano, 2021, p. 34.
[15] S. Chesterman, *We, the Robots? Regulating AI and the Limits of the Law*, Cambridge University Press, Cambridge (UK), 2021, pp. 63 ss.
[16] *White Paper On Artificial Intelligence – A European approach to excellence and trust*, 19/02/2020, p. 12.

3. *Exotic mind entities*

Il "*veil of technical inscutability*"[17], il problema dell'opacità, sta diventando centrale nella riflessione sull'intelligenza artificiale non solo perché è tecnicamente difficile (e in alcuni casi impossibile) aprire la scatola nera dei codici di funzionamento, ma perché costituisce la zona di confine nel rapporto tra meccanizzazione della mente e umanizzazione della macchina. Questa zona grigia era già il dato inquietante, inquietante per il dualismo tra soggetto e oggetto, del tentativo delle neuroscienze cognitive di spingere fino alle estreme conseguenze il "sottofondo biologico" dell'attività cerebrale riconducendola integralmente alla biochimica dei neuroni o alla biofisica delle sinapsi. Se le nostre sensazioni non sono altro che particolari algoritmi spazio-temporali che si armonizzano con determinate frequenze elettromagnetiche, dovremmo cominciare a prendere sul serio l'equiparazione del cervello a un computer. Il riduzionismo fisicalista ritiene che, "se il senso e la mente vengono associati con la materia, è perché *provengono da* essa"[18].

La mente sarebbe il *software* che elabora gli algoritmi implementati nel cervello (l'*hardware*) e sono tali algoritmi che le diverse branche delle neuroscienze cognitive cercano di riprodurre per poi trasferirli nei supporti fisici più disparati. Il problema è trovare gli algoritmi giusti e, forse, l'algoritmo "definitivo" con cui è sintetizzabile la teoria del tutto, una stringa che contiene l'essenza dell'universo[19].

Il *black box effect* ci presenta l'altra faccia della stessa medaglia: una sorta di possibile umanizzazione delle macchine che, rendendo sempre più complessi e imprevedibili i "propri" percorsi, avvicinano il digitale al celebrale. "Ci vogliono più che nervi d'acciaio per fare una cosa del genere. Ci vuole un cervello di silicio"[20]. È il commento di Kasparov, il campione di scacchi, dinanzi alla mossa, con cui il computer *Deep Blue* lo ha sconfitto. "Era diversa da qualunque altra cosa un computer avesse mai fatto prima. Ed era anche diversa da qualunque cosa un essere umano avesse mai preso in considerazione. Era qualcosa di nuovo, una totale rottura con la tradizione, una deviazione radicale da migliaia di anni di esperienza accumulata"[21]. È il commento al "colpo 37" con cui il computer *AlphaGo* ha sconfitto il campione del mondo, Lee Sedol, in una partita di *Go*, il gioco più difficile per la sua complessità da inserire in un sistema informatico.

Ha avuto la stessa sensazione di entrare a contatto con una singolare forma di pensiero l'ingegnere di Google, Blake Lemoine, dialogando con LaMDA il sistema

[17] F. Pasquale, *The Black Box Society. The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambridge (Mass.)/London, 2015, p. 163.
[18] J.P. Dupuy, *Alle origini delle scienze cognitive. La meccanizzazione della mente*, trad. it., Mimesis, Milano, 2009, p. 37
[19] P. Domingo, *L'algoritmo definitivo. La macchina che impara da sola e il futuro del nostro mondo*, trad. it., Bollati Boringhieri, Torino, 2016, p. 332 (ed. digitale).
[20] G. Kasparov, M. Greengard, *Deep Thinking. Dove finisce l'intelligenza artificiale e comincia la creatività umana*, Fandango, Roma, 2019, p. 147 (ed. digitale).
[21] B. Labatut, *Maniac*, trad. it., Adelphi, Milano, 2023, p. 314.

elaborato dalla sua azienda attraverso l'impiego di reti neurali artificiali. Anche gli *affective computing* sembrano, per la loro capacità di leggere le nostre emozioni, manifestare una sorta di emotività. Del resto "*Without emotion, computers are not likely to attain creative and intelligent behavior, but with too much emotion, we, the maker, may be eliminated by our creation*"[22].

Non possiamo, quindi, escludere che le nuove tecnologie si stiano consolidando attorno a una particolare forma di intelligenza che nasce dalle macchine e si sviluppa con le macchine, andando oltre le macchine. Macchine di Dio[23]? Macchine sapienti[24]? Oppure semplicemente "macchine ingannevoli"[25]? Deep Blue, LamDa, Watson, AlpaZero, Chat-GPT, sono estremamente diverse tra loro, ma sono indifferenziatamente molto più vicine a noi di quanto potrebbe esserlo una lavatrice. Ce ne rendiamo conto appena cerchiamo di individuare le caratteristiche fondamentali dell'intelligenza umana. Ad esempio, la percezione, il *problem solving*, la pianificazione, la decisione, la comprensione, l'apprendimento, l'astrazione, la creazione, il ragionamento, i sentimenti, la comunicazione, l'azione[26].

Ogni forma di intelligenza artificiale implica o esclude qualcuna di queste caratteristiche. Ma, per quante dissomiglianze possiamo individuare, restiamo estremamente lontani dal mondo degli oggetti. Dobbiamo fare ricorso all'*agency* della terminologia anglosassone, meno compromettente del nostro "soggettività", per sottolineare il passaggio da un'intelligenza artificiale puro strumento a nostra disposizione a un'intelligenza artificiale che ci supporta aggiungendo il "suo" alle nostre volontà. Non sappiamo dove ci condurranno gli algoritmi indeterministici, con la loro opacità, o i processori quantistici, con la loro complessità, e non siamo neppure in grado di immaginare cosa potrà rappresentare per noi quella super intelligenza che la teoria della singolarità prospetta come la sintesi unitaria di tutte le tecnologie digitali. Il ricorso alla nozione di agentività dovrebbe consentire di graduare i livelli regolamentari in base ai diversi tipi di autonomia, di complessità e di opacità, imponendo tutta una serie di specifici metodi di controllo e di blocco in relazione ai livelli di rischio di questi singolari ed eterogenei agenti artificiali.

Resta aperto il problema più ampio del rapporto tra organico e inorganico, tra la tecnologia del carbonio che sta alla base dell'uno e la tecnologia del silicio che sta alla base dell'altro.

[22] R.W. Picard, "Affective Computing", in *M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 321*, p. 15.
[23] H. Novotny, *Le macchine di Dio. Gli algoritmi predittivi e l'illusione del controllo*, trad. it., Luiss University Press, Roma, 2022.
[24] P. Benanti, *Le macchine sapienti. Intelligenze artificiali e decisioni umane*, Marietti, Bologna, 2018.
[25] S. Natale, *Macchine ingannevoli. Comunicazione, tecnologia, intelligenza artificiale*, trad. it., Einaudi, Torino, 2022, p. 9.
[26] Seguo l'analisi di P. Thagard, *op. cit.*, pp. 30 ss.

> Se delle molecole chimiche molto complicate possono operare negli esseri umani in modo tale da renderli intelligenti, dei circuiti elettronici altrettanto elaborati potrebbero a loro volta far agire i computer in modo ingegnoso. E una volta raggiunta l'intelligenza, è presumibile che loro stessi saranno in grado di progettarne altri dotati di una complessità e di un acume ancora più sviluppati[27].

Seguendo la teoria dell'evoluzione, il pensiero si è prodotto dalla materia[28]. Perché non potremmo riprodurre, a nostra volta, il pensiero attraverso la materia? L'opacità, che non riusciamo né a spiegare né a controllare, è solo la prima avvisaglia di una possibile transizione di fase in cui l'approfondirsi dei processi di *machine learning* finirà per determinare quello stesso scarto che, nella prospettiva fisicalista, avrebbe dato origine prima alla vita e poi all'attività cerebrale?

L'intelligenza artificiale ci costringe a ripercorrere quel processo per cui siamo sospesi, come insegnava Monod, tra il caso e la necessità. Il caso delle transizioni di fase e la necessità dei nuovi e inattesi equilibri che si sono venuti a determinare. Che siano i 13,8 miliardi di anni, nei primi brevissimi sviluppi dell'evoluzione dell'universo, in cui si è determinato uno squilibrio tra materia e antimateria, oppure i 3,5 miliardi di anni in cui sono comparsi i primi batteri o i 430 milioni di anni dell'esplosione biologica del cambriano o… delle prime sinapsi, a un certo punto, avviene uno scarto. Il metabolismo di enzimi e proteine ha determinato la capacità di duplicazione degli acidi nucleidi o la capacità di duplicazione degli acidi nucleidi ha determinato il metabolismo? In un caso e nell'altro, all'improvviso, la chimica e la fisica hanno prodotto "informazione". L'informazione che consente l'acquisizione e la distribuzione di energia; l'informazione che consente l'autoreplicazione e quindi produce quelle macchine per la sopravvivenza che chiamiamo vita. Dalla vita al pensiero è un altro scarto: "com'è accaduto che un protozoo simile a un coanoflagellato diventasse una spugna, una spugna diventasse uno cnidario, uno cnidario un invertebrato bilaterale, e un invertebrato bilaterale un vertebrato e un vertebrato un essere pensante?"[29].

Mi pare estremamente significativo che, negli ultimi anni, la maggior parte dei libri sull'intelligenza artificiale tocchi il tema dell'origine della vita e del pensiero, mentre i libri sull'origine della vita e del pensiero si concludono, a loro volta, con una serie di interrogativi sull'intelligenza artificiale. Sullo sfondo si staglia il problema del *black box effect*, dell'imprevedibilità. La teoria dell'evoluzione ci pone fronte al fascino della complessificazione. Ci spiega che avremmo avuto, sin dai primi secondi del *big bang*, un continuo aumento della

[27] S. Hawking, *Le mie risposte alle grandi domande*, trad. it., Rizzoli, Milano, 2018, p. 136 (ed. digitale).
[28] T.W. Deacon, *Natura incompleta. Come la mente è emersa dalla materia*, trad. it., Le scienze, Milano, 2012.
[29] J. Le Doux, *Lunga storia di noi stessi, Come il cervello è diventato cosciente*, trad. it., Raffaello Cortina, Milano, 2020, p. 161.

complessità e degli adattamenti sistemici alla complessità, ma non ci fornisce una ragione della complessificazione.

Possiamo pensare, con Gould, che siamo un evento improbabile nel film della vita[30] o, convenire con Al-Khalili, che "la probabilità che processi casuali determinino l'origine della vita è la stessa di una tromba d'aria che colpisca una discarica e produca un aereo"[31]. Resta il fatto che "[…] noi esseri viventi siamo i vortici generati dalle onde d'acqua liberate dall'apertura della paratia. Siamo lo spumeggiare irreversibile dell'energia libera che era intrappolata nel disequilibrio fa idrogeno ed elio ed è stata liberata dal sole"[32]. Proprio l'inesplicabile legame tra organico e inorganico, evidenziato dalla fisica quantistica, impone l'analisi dei modelli di senzienza o forse addirittura di seità presenti in tutte le specie viventi, mentre si prospetta l'eventualità che "l'intelligenza biologica sia soltanto un fenomeno transitorio, una fase passeggera nell'evoluzione dell'intelligenza nell'universo"[33].

Una volta messo da parte il pregiudizio cartesiano degli animali come macchine senz'anima, ci troviamo di fronte a svariate "figure della mente"[34] che vanno dai batteri alle piante. Non solo, ma emerge l'ipotesi che "l'esperienza sentita comparve diverse volte in diverse linee evolutive"[35]. Noi saremmo soltanto una di queste possibili varianti. Dovremmo, quindi, abbandonare ogni residuo antropomorfico e incominciare a immaginare una pluralità di dimensioni in cui azione, sensibilità e coscienza si esprimono non solo variamente, ma anche diversamente. Le riflessioni su una coscienza diffusa in tutte le specie viventi muovono dalla constatazione che ogni cervello presenta un'immagine virtuale del mondo esterno, controllata dai sensi e regolata dal principio di autoconservazione. Ogni specie vivente opera attraverso una sorta di "allucinazione" determinata dalle proprie caratteristiche biologiche[36]. Non entriamo a contatto con il mondo così come è, secondo le leggi della fisica, ma come ce lo presenta, e forse meglio racconta, la nostra percezione sensoriale. Diverse percezioni, diversi mondi, diverse agentività, diverse senzienze, diverse seità…

Mi torna in mente la meraviglia destata dall'evoluzione dei *Large Language Models*. "*A threshold was reached, as if a space alien suddenly appeared that could*

[30] Riprendo le osservazioni di D. Bickerton, *Quello di cui la natura non ha bisogno. Linguaggio, mente ed evoluzione*, trad. it., Adelphi, Milano, 2022, p. 285 (edizione digitale).
[31] J. Al-Khalili, J. McFadden, *La fisica della vita. La nuova scienza della biologia quantistica*, trad. it., Bollati Boringhieri, Torino, 2015, p. 229 (ed. digitale).
[32] C. Rovelli, *Buchi Bianchi*, Adelphi, Milano, 2023, p. 112.
[33] M. Kaku, *Il futuro dell'umanità. Dalla vita su Marte all'immortalità, così la scienza cambia il nostro destino*, trad. it., Rizzoli, Milano, 2018, p. 292.
[34] S. Ginsburg, E. Jablonka, *Figure della mente. La coscienza attraverso la lente dell'evoluzione*, trad.it., Raffaello Cortina, Milano, 2023. Attraverso il "filo di Arianna" della biologia evoluzionista il libro passa in rassegna le diverse entità che possono definirsi coscienti e in che termini.
[35] P. Godfrey-Smith, *Metazoa*, trad. it., Adelphi, Milano, 2021, p. 320.
[36] A. Seth, *Come il cervello crea la coscienza*, trad. it., Raffaello Cortina, Milano, 2023, pp. 207 ss.

communicate with us in an eerily human way"[37]. Nel rapporto con questi "inquietanti spazi alieni" il concetto di senzienza è particolarmente utile perché ci impone di cambiare prospettiva, prendendo in esame forme "esotiche" di mente che, pur non rispecchiando i parametri umani, elaborano costrutti diversamente intelligenti. Se si parla di embrioni di coscienza nei batteri e nelle piante o di sofferenza nei pesci non si intende la pedissequa riproduzione dei nostri sentimenti. Come diamo per scontato che ogni forma vivente abbia elaborato i propri percorsi di interazione con il mondo in relazione alla propria nicchia biologica ed ecologica, così i percorsi imprevedibili degli algoritmi stocastici potrebbero essere il segno di qualcosa di analogo: l'inattesa reazione di questi algoritmi alla diversità dei dati in rapporto all'ulteriore diversità dei sistemi di addestramento. Se si parla di intelligenze artificiali al plurale è proprio per sottolineare la varietà dei modelli di sviluppo e di implementazione, pur sulla base delle stesse premesse scientifiche.

Possiamo parlare di *exotic mind entities* oppure di *alien reasoning* o di assemblaggi cognitivi, ci rendiamo in ogni caso conto che non è rilevante domandare quanta intelligenza ha una macchina o un topo, dovremmo piuttosto interrogarci su quanto siamo in grado di percepire e utilizzare il loro "esotismo" e fino a che punto sia necessario abituarci a convivere con un paradigma logico diverso dal nostro. Come è stato detto a proposito di Alpha Zero: si sta facendo strada l'idea che ci possano essere modi del conoscere che, pur non essendo assimilabili alla coscienza umana, siano ugualmente forme di conoscenza[38]. Da una parte, come nota Floridi, l'intelligenza artificiale ci apre

> lo spazio sterminato di problemi e compiti, ogni volta che questi possono essere conseguiti senza comprensione, consapevolezza, acume, sensibilità, preoccupazioni, sensazioni, intuizioni, semantica, esperienza, bio-incorporazione, significato, persino saggezza e ogni altro ingrediente che contribuisca a creare l'intelligenza umana[39].

Dall'altra l'effetto *black box* aggiunge a tutto questo, l'elaborazione di percorsi di apprendimento che sfuggono al condizionamento umano, sommando all'automatismo di compiti sempre più complessi l'eventualità di modelli di analisi e di comunicazione finora impensabili.

Ad esempio, l'utilizzazione di un programma di intelligenza artificiale simile a ChatGPT, addestrato su dati relativi alla composizione chimica e genetica di milioni di cellule, ha consentito di scoprire le "cellule Norn", un raro tipo di cellula renale che è responsabile della produzione dell'eritropoietina quando i livelli di ossigeno diventano troppo bassi. "*A vital discovery about biology that otherwise*

[37] T.J. Sejnowski, "Large Language models and the reverse turing test", in *Neural Computation*, 35 (2023), n. 3, p. 303.
[38] Fino a che punto possiamo aprirci all'impensato? Cfr. N. Katherine Hayles, *L'impensato. Teoria della cognizione naturale*, trad. it., Effequ, Firenze, 2021.
[39] L. Floridi, *Etica dell'intelligenza artificiale*, cit., p. 52.

would not have been made by the biologists" ha affermato Eric Topol, direttore dello Scripps Research Translational Institute[40].

Partendo da presupposti simili, si è pensato di impiegare i trasformatori generativi pre-addestrati, che stanno alla base dei sistemi digitali di traduzione, per intercettare le comunicazioni tra gli animali. Sappiamo che alcune specie (dai delfini ai topi, dai capodogli agli uccelli) hanno modelli sofisticati di trasmissione dei segnali che, legati in specifiche sequenze, assumono un particolare significato in base al modo in cui si formano le combinazioni finali[41]. Sono in corso diversi esperimenti non solo per riuscire ad operare una corretta registrazione e classificazione, ma anche per realizzare una plausibile decifrazione attraverso "animalgoritmi"[42]. L'autonomia della capacità computazionale potrebbe rompere le barriere naturali, consentendoci di penetrare le forme di comunicazione "altra"[43]? E con quali effetti?

4. Un diritto naturale al silicio?

Gli sviluppi dell'intelligenza artificiale sembrano, quindi, imporre una riflessione su ciò che è "naturale" o meglio su quanto l'artificiale sia un aspetto del naturale[44]. Le reti neurali artificiali dei sistemi di *Deep learning* imitano la struttura e i modelli di apprendimento del cervello; la progettazione dei robot muove dalla biomimesi, dalla riproduzione delle forme e dei movimenti dei più svariati organismi biologici; gli animalgoritmi cercano di riprodurre i processi comunicativi presenti nelle varie forme di vita; i computer quantistici toccano il problema della struttura dell'universo; gli *affective computing* interpretano i modelli espressivi delle nostre condotte. In nessuna epoca della storia ci siamo interrogati, con dispositivi tanto raffinati, su quanto vi sia di meccanico in noi e su quanto vi sia di umano nel mondo che ci circonda. Più costruiamo sofisticati modelli di intelligenza artificiale e più scopriamo quanto poco sappiamo su noi stessi. Potremmo ripetere con Eliot, "*Endless invention, endless experiment,… /All our knowledge brings us nearer to our ignorance… /Where is the wisdom we have lost in knowledge?*"[45].

Quale "macchinazione" è più perfetta di quella che muove da un mondo meccanico dominato da algoritmi e robot per poi farci intuire le tante "figure della

[40] C. Zimmer, "A.I. Is Learning What It Means to Be Alive", in *New York Times* (10/03/2024).
[41] E. Bucci, *Geni, memi e bit. Evoluzione biologica, termodinamica e teoria dell'informazione*, Mondadori, Milano, 2024, pp. 200 ss.
[42] T. Mustil, *Come parlare il balenese. Il futuro della comunicazione animale*, trad. it., Il Saggiatore, Milano, 2023, cap. 9.
[43] È più lontano nel tempo del libro di Mustil, ma altrettanto suggestivo S. Budiansky, *Se un leone potesse parlare. L'intelligenza animale e l'evoluzione della coscienza*, trad. it., Baldini Castoldi Dalai, Milano, 2007.
[44] F. De Felice, R. Race, *Dialoghi su etica e intelligenza artificiale*, Luiss University Press, Roma, 2023, p. 69.
[45] T.S. Eliot, "Cori da 'La rocca'", in *Opere [1904-1939]*, Bompiani, Milano, 2005, p. 1230.

mente" di una natura che, tra l'agentività e la senzienza, presenta svariati modelli di percezione sensoriale e di trasmissione dell'informazione? Quale macchinazione è più perfida di quella che mostra quanto siamo impotenti dinanzi all'opacità degli algoritmici e, poi, suggerisce che proprio questa opacità ci potrebbe aiutare a penetrare alcuni degli aspetti più riposti del vivente?

Sembra delinearsi, pur con tutte le radicali ed evidenti differenze, un orizzonte culturale molto vicino a quello su cui si sono fondate, nel passato, le varie teorie del diritto naturale e, in particolare, quella declinazione di questa teoria che intendeva far propri gli insegnamenti riposti nelle costanti dei processi di sviluppo e nelle omogeneità latenti in natura: penso alla teoria dell'*oikeiosis*, alla *dynamei on* di Aristotele, alla *teleiosis* stoica, all'*ordo factivus* medievale. Il tentativo di elaborare animalgoritmi ricorda quel *proximun humanis sensibus*, quell'attenzione per la vicinanza tra tutte le specie viventi cara allo stoicismo, mentre le potenzialità degli algoritmi, che prendono *forme* impreviste a mano a mano che assimilano l'imponderabilità dei dati, richiamano la raffinata terminologia scolastica che distingueva l'*intellectus agens*, l'intelletto che è in azione, dall'*intellectus possibilis*, l'intelletto che ha in sé gli elementi dei futuri sviluppi, e l'*intellectus in actu*, l'intelletto che si manifesta caso per caso.

Se nel diritto naturale è fondamentale la domanda su quali siano i valori propri dell'identità umana, anche per l'intelligenza artificiale viene invocato un "*HCAI approach*", uno sviluppo umano centrico, e si parla di *ethics by design*, eualgoritmica, infoetica, algor-etica, *data ethics*, *data dignity*.

L'*Ethics by Design* viene richiamata più volte dal Regolamento 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati (GDPR). L'*Ethics by Design* è ribadita dalla Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate, ponendo le basi delle scelte compiute nel recente *AI Act*.

Una sorta di eualgoritmica[46] si profila nella Raccomandazione *The Ethics of Artificial Intelligence* dell'Unesco del 31/07/2022 che suggerisce agli Stati membri di sviluppare strategie che assicurino la continua valutazione della qualità della programmazione dei sistemi digitali, con particolare attenzione per l'adeguatezza dei processi di raccolta e selezione dei dati, promuovendo la sicurezza e l'affidabilità attraverso la previsione di specifici meccanismi di *feedback* per imparare dagli errori e condividere le migliori pratiche, eventualmente certificate con "*Gold standard datasets*".

Data ethics e *data dignity* dovrebbero essere l'esito della *DIKW pyramid* (*Data*, *Information*, *Knowledge*, *Wisdom*) che auspica un'ecologia della

[46] Usa questa espressione M. Chiriatti, *#Humanless. L'algoritmo egoista*, Ulrico Hoepli, Milano, 2024, pp. 79 ss.

comunicazione in cui i singoli passaggi (*purpose*, *insight*, *meaning*, *context*) dovrebbero, se programmati adeguatamente, determinare un'ideale concatenazione virtuosa.

L'info-etica ha per oggetto la deontologia dei programmatori, la serietà dei diffusori delle notizie, la disciplina delle piattaforme che acquisiscono e veicolano l'informazione, la correttezza dei fruitori pubblici e privati. In questo campo abbiamo, attualmente, un insieme eterogeneo di indicazioni normative. Dichiarazioni internazionali, linee guida, codici deontologici, politiche aziendali sottolineano continuamente, anche se con sfumature e accentuazioni diverse, la necessità di garantire: dignità umana, responsabilità, autonomia, giustizia, equità, solidarietà, trasparenza, affidabilità, sicurezza, salute, integrità fisica e morale, *privacy*, sostenibilità. Vengono progettate *start up* per contrastare le *fake news*, verificando le fonti e l'attendibilità delle notizie, e si propone l'istituzione di *board* per l'etica e la sicurezza. Molte di queste aspirazioni sono significativamente condivise da filosofi, scienziati e imprenditori come si evince, ad esempio, dalla Dichiarazioni di Asilomar del 2017 o dal documento *Rome Call for AI Ethics*, sottoscritto da Pontificia Accademia per la Vita, Microsoft, IBM e Fao, a conclusione dell'incontro svoltosi in Vaticano il 28 febbraio 2020 su *The 'Good' Algorithm? Artificial Intelligence: Ethics, Law, Health*.

L'ambizione estrema della programmazione algor-etica sarebbe quella di inserire un codice etico in un codice macchina, dotando gli stessi algoritmi di regole che garantiscano il rispetto dei valori fondamentali. Valori condivisi e quindi universali, valori programmabili e quindi definiti, valori essenziali e quindi efficaci. Tra algor-etica, *ethics by design*, eualgoritmica, infoetica, *data ethics* e così via emerge una sistematica esigenza di eticizzazione dei bit alla ricerca di un "costituzionalismo digitale", espressione positiva dell'ordine di valori essenziale alla tutela dei diritti fondamentali. Possiamo definirlo un diritto naturale al silicio?

Nel riflettere sulla storica contrapposizione tra autocrazia e democrazia, Kelsen affermava ironicamente che "l'uomo non è un libro pensato a tavolino"[47]. Eppure, seduti ai loro tavolini, i programmatori stanno ormai da tempo ripensando e condizionando il senso della vita e della nostra società. Ecco la necessità di porre tutta una serie di domande. Dati: quali? Informazione: su cosa? Conoscenza, per chi? Saggezza: di chi? Domande ormai ineludibili dinanzi al dominio esercitato sul *web* e sulle tecnologie ad esso collegate dalle corporazioni digitali che, come scrive Habermas, promuovono una sistematica "spinta verso la mercificazione dei contesti del mondo della vita"[48]. Chi guida i *cookie*, elabora i filtri, canalizza i flussi delle informazioni? Gli sviluppi tecnologici assecondano le aspirazioni etiche o le esigenze del mercato? L'intelligenza artificiale si progetta per vendere o per assistere? Serve a incrementare gli utili o il benessere? Guarda al bene o al consumo? All'efficienza produttiva o alla qualità della vita?

[47] H. Kelsen, *Il primato del parlamento*, trad. it., Giuffré, Milano, 1982, p. 41.
[48] J. Habermas, *Nuovo mutamento della sfera pubblica e politica deliberativa*, trad. it., Raffaello Cortina, Milano, 2023, p. 56.

Può la saggezza scaturire dagli algoritmi o dall'insieme dei dati? Il dato, al singolare, è un enunciato descrittivo, vero o falso, ma in sé privo di qualsiasi implicazione assiologica. L'algoritmo, al singolare, è un insieme di numeri, più o meno coerenti logicamente, ma in sé privo di qualsiasi implicazione assiologica. I dati e gli algoritmi, assunti nella loro pluralità attraverso la rielaborazione nei contesti in cui sono formati o alla luce degli scopi per cui sono acquisiti, assumono una capacità trasformativa per cui c'è una soglia oltre la quale la quantità o acquista la dignità di uno strumento di elevazione sociale o si traduce in un meccanismo di oppressione. Siamo in bilico tra infoetica e infocrazia.

Lo "*HCAI approach*", l'approccio umano-centrico, mette in luce come l'etica non sia qualcosa di sovrapposto alla scienza e la scienza qualcosa di avulso dall'etica, ma i valori siano il modo attraverso cui ci accostiamo ai fatti e comprendiamo i fatti: è l'antica lezione del diritto naturale sulla indispensabile ricerca di una connessione tra empirismo e razionalismo[49]. All'interno della concezione positivistica la distinzione tra fatti e valori è netta. Nella prospettiva del non cognitivismo etico, fondato sul timore della fallacia naturalistica, è impossibile parlare di un'etica dei dati e ancor meno di un'algor-etica. Tuttavia gli sviluppi dell'intelligenza artificiale ci mostrano come l'etica non possa restare qualcosa di esterno ed eventuale, chiamata magari *a posteriori* a valutare le conseguenze di determinati effetti, ma deve essere inglobata nella dimensione progettuale. L'*ethics by design* impone l'affermazione del principio di esplicabilità[50]: "*the obligation to gain greater awareness of unintended consequences and the moral significance of what they do, including how they deal with tragic problems. If AI is not going to be responsible in this sense, it will crash*"[51].

Il principio di esplicabilità è appena accennato nel generico riferimento al "diritto di ottenere l'intervento umano, di esprimere la propria opinione, di ottenere una spiegazione" del Considerando 71 del GDPR. Emerge, ma sempre timidamente, nel Considerando 38 dell'AI Act che afferma: "potrebbe inoltre essere ostacolato l'esercizio di importanti diritti procedurali fondamentali, quali il diritto a un ricorso effettivo e a un giudice imparziale, nonché i diritti della difesa e la presunzione di innocenza, in particolare nel caso in cui tali sistemi di IA non siano sufficientemente trasparenti, spiegabili e documentati". Anche il Considerando 40 dell'AI Act sottolinea come "per ovviare all'opacità che può rendere alcuni sistemi di IA incomprensibili o troppo complessi per le persone fisiche, è opportuno imporre un certo grado di trasparenza per i sistemi di IA ad alto rischio. Gli utenti

[49] B. Shneiderman, *Human-Centered AI*, Oxford University Press, Oxford, 2022, p. 39.

[50] B. Goodman, S. Flaxman, "European Union regulations on algorithmic decision-making and a 'right to explanation'", in *AI Magazine*, 38 (2017), n. 3, pp. 1 ss.

[51] M. Coeckelbergh, "Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability", in *Science and Engineering Ethics*, 26 (2020), p. 2068.

dovrebbero poter interpretare gli output del sistema e utilizzarlo in modo adeguato". Sono più perentorie le linee guida dell'IBM "*Imperceptible AI is not ethical AI*"[52].

L'esplicabilità garantisce la fiducia. La fiducia presuppone il primato della trasparenza sull'opacità. La trasparenza implica il controllo dell'imprevedibilità da parte della ragione e quindi la limitazione del potere da parte dei diritti. Se guardiamo al passato, non possiamo ignorare come sia stata la crescita dei diritti umani sotto l'impulso del diritto naturale a dissolvere l'assolutismo degli *arcana imperii*[53]. Oggi la scatola nera del *machine learning* rappresenta una nuova forma di oscurantismo, avallato dalle pressioni degli interessi economici sulla politica e della politica sulla scienza. Non sappiamo se questo singolare diritto naturale al silicio che si sta formando, tra computer e algoritmi, avrà la stessa capacità che ha avuto la tradizione giusnaturalistica nel sollevare la dignità umana, nel sollecitare quella "intenzione verso l'andatura eretta"[54] che ha rifiutato la soggezione al potere. È innegabile che l'intelligenza artificiale rappresenti una svolta cruciale all'interno del nostro futuro. L'eualgoritmica, l'infoetica, l'algor-etica, l'*HCAI approach*, la *DIKW pyramid* hanno il merito di porre "oggettivamente in primo piano il problema dei giusti motivi sociali dell'agire, e soggettivamente il problema di coscienza della possibilità di conoscere quei giusti motivi"[55].

[52] "Explainability", in *IBM Design fo AI*. Recuperato da: https://www.ibm.com/design/ai/ethics/explainability/#:~:text=Explainability%20is%20key%20for%20users%20interacting%20with%20AI,seamless%20experience.%20Imperceptible%20AI%20is%20not%20ethical%20AI,[Data di consultazione: 30/05/2024].

[53] Su AI e totalitarismo, M. Coekelberg, *The Political Philosophy of AI*, Polity Press, Cambridge, 2022, pp. 38 ss.

[54] E. Bloch, *Diritto naturale e dignità umana*, trad. it., Giappichelli, Torino, 2005, p. 173.

[55] Quando ha scritto queste parole, Welzel non sapeva neppure cosa fosse l'intelligenza artificiale, ma era sicuro di offrire una nitida descrizione del diritto naturale (*Diritto naturale e giustizia materiale*, trad. it., Giuffré, Milano, 1965, p. 7).