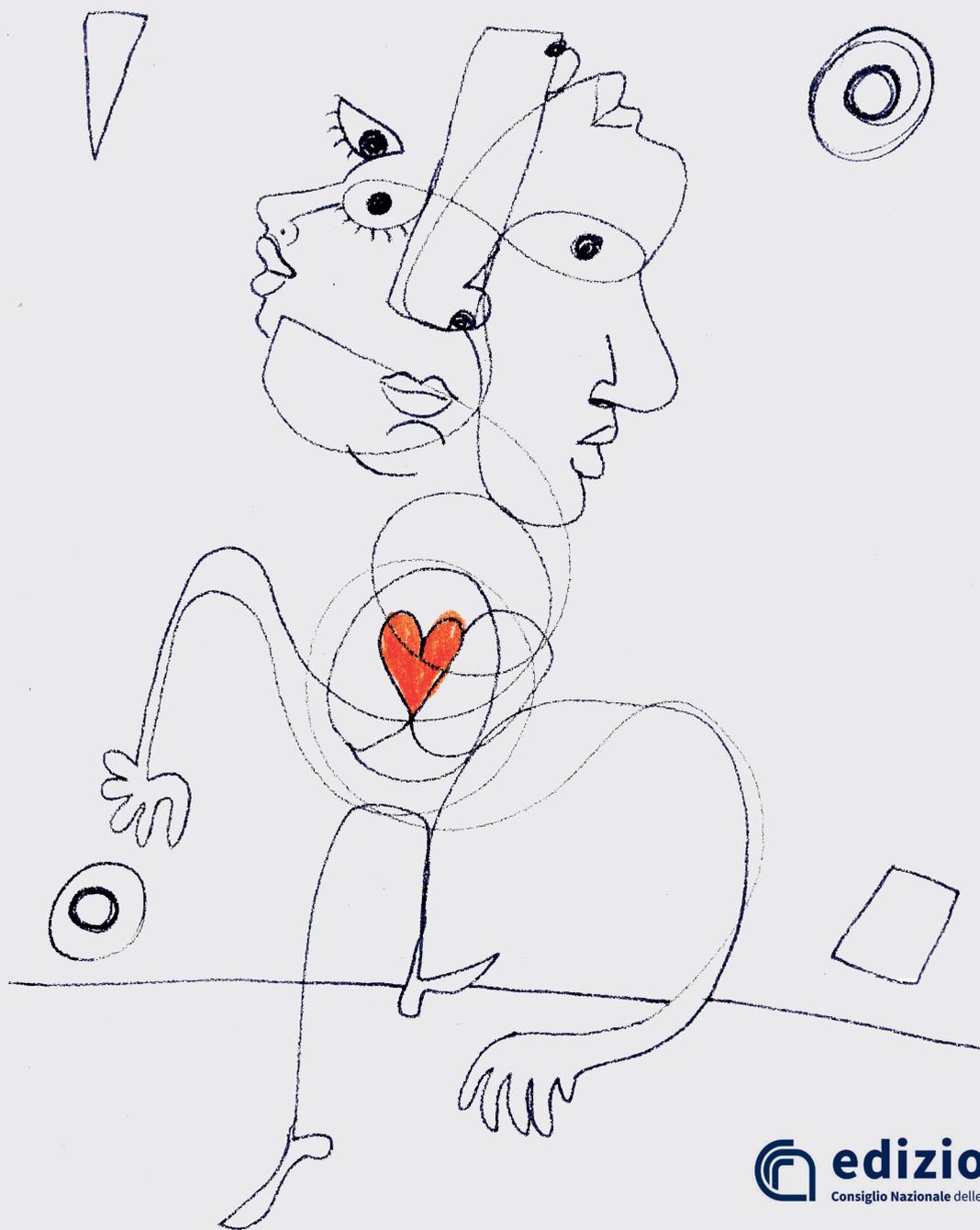


INTELLIGENZA ARTIFICIALE: DISTINGUE FREQUENTER

Uno sguardo interdisciplinare

Consulta Scientifica del Cortile dei Gentili

Prefazione di Giuliano Amato



INTELLIGENZA ARTIFICIALE: *DISTINGUE FREQUENTER*

Uno sguardo interdisciplinare

Consulta Scientifica del Cortile dei Gentili

Prefazione di
Giuliano Amato

A cura di
Cinzia Caporale e Laura Palazzani



www.cortiledeigentili.com

*Responsabile Comunicazione “Cortile dei Gentili”
(Pontificio Consiglio della Cultura)*
Giulia Tosana
g.tosana@cortile.va

© Cnr Edizioni 2023
P.le Aldo Moro, 7
00185 Roma
www.edizioni.cnr.it
bookshop@cnr.it

Redazione e progetto grafico
Marco Arizza

Segreteria di redazione
Giorgia Adamo e Tiziana Ciciotti

Disegno di copertina di Ginevra Saccà,
“Al cuore dell’Intelligenza Artificiale” (2023),
per gentile concessione dell’artista

Finito di stampare nel mese di maggio 2023
da Tipografia Facciotti S.r.l. - Roma

ISBN 978 88 8080 558 8
ISBN 978 88 8080 559 5 (digital version)
DOI <https://doi.org/10.48220/intelligenzaartificiale-2023>

INDICE DEL VOLUME

<i>Prefazione</i> Giuliano Amato	5
<i>Lettera alla Consulta Scientifica del Cortile dei Gentili</i> S.Em. Card. Gianfranco Ravasi	9
<i>Introduzione</i> Cinzia Caporale	11
<i>L'Intelligenza Artificiale: distingue frequenter.</i> <i>Come giungere a una comunanza etica nella società del pluralismo</i> Documento della Consulta Scientifica del Cortile dei Gentili	15
PARTE PRIMA: Per una storia dell'IA	
<i>Le stagioni dell'IA</i> Paolo Benanti	21
PARTE SECONDA: IA e società	
<i>Le scienze sociali di fronte all'IA</i> Carla Collicelli	45
PARTE TERZA: IA e filosofia	
<i>Contro l'infosfera. Salvare la presenza</i> Eugenio Mazzarella	61

PARTE QUARTA: IA e concezione dell'uomo

Cosa resta dell'umano nell'epoca dell'IA
Laura Palazzani 77

PARTE QUINTA: IA e teologia

Le domande della teologia cristiana sull'IA
S.E. Mons. Antonino Raspanti 87

PARTE SESTA: IA e etica

IA ed etica pubblica
Stefano Zamagni 95

Alla ricerca di un Habeas Corpus per l'IA
Luca Antonini e Antonella Sciarrone Alibrandi 105

PARTE SETTIMA: IA ed economia

La funzione degli algoritmi e il discernimento umano
Leonardo Becchetti 115

PARTE OTTAVA: IA in medicina

L'IA: sostituzione o sostegno del medico?
Giuseppe R. Gristina e Luciano Orsi 121

PARTE NONA: IA e umano: uno sguardo d'insieme

IA e umano: uno sguardo d'insieme
Jean-Pierre Darnis e Mons. Carlo Maria Polvani 127

**I membri della Consulta Scientifica del Cortile dei Gentili
e gli altri autori** 137

PREFAZIONE

Giuliano Amato

Nel mio frequente ruolo di prefatore ho usualmente ritenuto utile fornire al lettore anticipazioni e sintesi dei temi che avrebbe incontrato nelle pagine successive. Ebbene, in questo volumetto sull'Intelligenza Artificiale della nostra Consulta c'è un contributo finale – quello di Jean-Pierre Darnis e di Carlo Maria Piovani – che sintetizza con efficacia i contributi precedenti e ne trae il succo. Ne tenga conto chi legge e mi assolve se in questa prefazione mi asterrò dalle sintesi e utilizzerò parte dello spazio che mi è concesso per mettere in luce, del libro, un profilo, che è ormai una costante delle nostre pubblicazioni; un profilo particolarmente espressivo del lavoro che facciamo, in conformità alla missione a cui vuole assolvere il Cortile dei Gentili.

C'è all'inizio un documento comune, introdotto da chi si è assunto il compito di estrarre dalle precedenti discussioni sul tema le posizioni condivise che si erano raggiunte (in questo, come in altri casi, Cinzia Caporale, a cui siamo gratissimi); ci sono poi contributi individuali, nei quali diversi di noi mettono a fuoco sia le angolature del tema più rilevanti nelle rispettive discipline, sia le proprie, personali vedute, che accentuano ora l'una ora l'altra delle valutazioni emerse nel documento comune. Ecco, il Cortile è e vuole essere questo. Vuole raccogliere attorno a questioni controverse persone che in comune non hanno le stesse idee né le stesse premesse culturali. Hanno piuttosto una caratteristica fondamentale, che vorremmo fosse sempre meno rara nelle società del nostro tempo, troppo propense a dividersi su posizioni incompensabili e intolleranti delle posizioni altrui. È la caratteristica di ascoltare l'altro, non con agnosticismo, ma con la disponibilità a capirne le ragioni, anche qualora non sia facile comporre con le proprie verità. Quando questa disponibilità c'è in tutti, quando – ci dicono, ciascuno a suo modo, Jürgens Habermas e Charles Taylor – la stessa ragione impara ad essere dialogante con le verità religiose, allora si creano le premesse perché si trovino le

piattaforme comuni di cui le nostre società hanno bisogno, prima ancora per esistere che per essere governate.

Ebbene, noi, con la nostra Consulta e con l'intero Cortile di cui essa fa parte, siamo solo un frammento del mondo espanso di oggi. La nostra, perciò, non è nulla più che una testimonianza. Tuttavia, abbiamo l'orgoglio di poter dire che da anni, ormai, la pratichiamo su questioni le più diverse e da anni il miracolo, tutto terreno, si produce: ciascuno dice la sua, mettiamo a fronte ciò che ne esce, e non solo scopriamo che la composizione è possibile, ma sotto di essa vediamo noi stessi emergere principi e cornici valoriali che in realtà condividiamo e che, così come appaiono a noi, dimostrano di esistere e di valere al di là di noi. È dunque vero – ci dicono i nostri ripetuti esperimenti – che anche a Babele possiamo trovare, pur nelle nostre diversità, un'armonia sufficiente a vivere insieme.

Questo volumetto è solo l'ultima delle nostre testimonianze. Era per ciò stesso importante sottolineare qui le ragioni di fondo della sua stessa articolazione, conforme a quella dei precedenti (documento comune e contributi individuali). Ed è importante che il lettore ritrovi tali ragioni all'interno del tema dell'Intelligenza Artificiale, così come esso è svolto nelle pagine seguenti. Non gli sarà difficile farlo. Non troverà infatti posizioni luddiste, e nessun *vade retro*, nei confronti dell'innovazione, e quindi della stessa Intelligenza Artificiale. Al contrario, di essa vedrà unanimemente condiviso il formidabile contributo ai nostri progressi nei campi più diversi. Ma vedrà anche indicate, sempre unanimemente, le aree di limiti possibili ed anzi necessari, con il chiarimento, non meno necessario, che alcuni di tali limiti tocca al diritto stabilirli, mentre altri non possono che stare nella coscienza dei ricercatori.

È il diritto che deve stabilire in quali attività non c'è nulla di male che l'IA, oltre ad esserci di aiuto, arrivi anche a sostituirci – si tratti della raccolta dei precedenti, in medicina come nel diritto, si tratti dell'effettuazioni di operazioni complesse, che essa può fare in tempi molto rapidi – in quali attività invece deve esserci sempre alla fine la decisione umana – si tratti dell'uso di armi in un teatro di guerra, si tratti della condanna penale di un imputato. E le ragioni abbondano sull'uno e sull'altro versante. Per converso solo la coscienza dei ricercatori li può fermare in

quel “*playing God*” che è già iniziato e che porta, da un lato a produrre (o meglio, a cercar di produrre) ampliamenti senza limiti delle potenzialità umane, ben oltre, ad esempio, la neutralizzazione di malattie ereditarie, dall’altro a dare alla IA, via reti neurali, emozioni, sentimenti, pensieri sul futuro.

La coscienza. È questo il capitolo più delicato, anche per la messa alla prova della nostra aspettativa di incontro fra premesse valoriali diverse. Eppure l’incontro, ancora una volta, si è rivelato possibile e si legga, a riprova, il contributo di Antonino Raspanti dedicato alle domande della teologia cristiana sull’IA. Certo, vi sono in esso argomentazioni che sono solo ed esclusivamente teologiche. Ma le domande, alla fin fine, sono soltanto della teologia cristiana? Solo per i credenti vi è una ragione che impedisce di sfuggire alla nostra finitezza? Solo per Agostino e per chi ne condivide la fede la radice del male è porre l’io al di sopra di tutto? *L’hybris* è punita soltanto perché viola le leggi degli dèi, o perché contrasta con i principi universali della convivenza umana?

Lascio a chi legge le risposte (ritenendo peraltro trasparenti le mie). Posso solo concludere che, se tutto questo è sufficientemente condiviso, se da un lato la società con le sue leggi, dall’altro la ricerca con il suo codice morale sapranno tenere le barre loro affidate, potremo con tranquillità fidarci della IA: avvalerci dei suoi benefici, non drammatizzare le nostre difficoltà nel distinguere l’Incompiuta di Schubert da quella in parte compiuta dal computer, imparare a capire, parlandoci, quando uno studente ha scritto il suo paper, quando se lo è fatto scrivere da ChatGPT. Non sarà questo a portarci nelle acque dove, secondo padre Dante, finirebbe per richiudersi sopra di noi il mare. Quelle acque sono più avanti e tocca a noi, a tutti noi, evitare che l’umanità ci arrivi.

Città del Vaticano, 2 maggio 2023

Cari Amici,

nel ricordarVi con stima e amicizia, desidero esprimerVi il mio più sincero apprezzamento per la pubblicazione *Intelligenza artificiale: distingue frequenter*, che conferma nuovamente quanto l'esperienza della *Consulta scientifica* sia importante, significativa e preziosa per il "Cortile dei Gentili".

Il tema dell'intelligenza artificiale e della sua applicazione in numerosi ambiti della vita richiede una riflessione multidisciplinare continua, affinché le implicazioni etico-morali che il progresso scientifico porta con sé trovino urgentemente risposta. Per questo ho molto apprezzato la concretezza e la completezza del Vostro lavoro; la pluralità delle Vostre voci e delle Vostre competenze offre uno sguardo inedito sulla questione, che spero possa arricchirsi nel tempo di ulteriori contributi e pensieri. Vi invito, infatti, a continuare il Vostro lavoro di studio, dibattito e approfondimento, con l'intelligenza, la passione e la creatività che sempre vi contraddistinguono.

Nell'augurare il miglior successo a questo volume e a tutte le iniziative ad esso correlate, desidero rinnovarVi la mia ammirazione e più sincera gratitudine per la dedizione e l'impegno che riservate alla *Consulta scientifica*: siete per noi tutti non solo un esempio di competenza e professionalità, ma anche di grande sensibilità culturale e umana.

Con viva cordialità,



Gianfranco Card. RAVASI
Fondatore

INTRODUZIONE

Cinzia Caporale

I rischi dei sistemi di Intelligenza Artificiale (IA) includono: (i) *bias* e discriminazioni, essi possono infatti perpetuare e amplificare i pregiudizi esistenti nella società, portando a risultati arbitrariamente selettivi; (ii) soppressione di posti di lavoro e disparità di reddito, perché automazione e IA possono sostituire i lavoratori umani in alcune funzioni; (iii) problemi di sicurezza e di *privacy*, nonché di compromissione potenziale di informazioni sensibili, considerato che tali sistemi sono vulnerabili agli attacchi degli *hacker* e in generale alla violazione dei dati; (iv) *deficit* di responsabilità e di *accountability*, poiché l'IA potrebbe prendere decisioni senza l'intervento umano e dunque rendere difficile determinare chi è responsabile di eventuali errori o esiti dannosi; (v) controllo e manipolazione eccessivi, nel senso che l'IA può essere utilizzata per la sorveglianza e per attività ispettive o di verifica, con potenziali gravi ripercussioni sulla libertà delle persone e sulla riservatezza; (vi) conseguenze non volute anche molto serie, dato che i sistemi di IA possono produrre esiti inattesi ed effetti collaterali negativi non prevedibili durante il loro sviluppo; (vii) importanti questioni etiche, in particolare su come vengono prese le decisioni.

Si tratta evidentemente di rischi che devono essere considerati con attenzione e affrontati sia nella fase di sviluppo sia nell'implementazione dei sistemi di Intelligenza Artificiale, in modo da garantire che questi ultimi vengano utilizzati in modo eticamente sostenibile.

In generale, il fatto che i benefici superino i rischi, dipenderà dalle singole specifiche applicazioni e dalle modalità d'uso. Ad esempio, sebbene sia già piuttosto chiaro che l'IA può apportare vantaggi assai significativi come una maggiore efficienza, accuratezza e capacità decisionale, se il sistema di Intelligenza Artificiale non fosse adeguatamente convalidato o se vi fossero errori nei dati utilizzati per addestrarlo, i risultati potrebbero essere inaffidabili e danneggiare i fruitori di quella specifica applicazione. Un problema, questo, da non trascurare, anche in ragione del fatto che è

altamente improbabile che nel futuro saremo in grado di escludere anche solo parzialmente l'Intelligenza Artificiale dalle nostre vite visto che è già profondamente integrata in molti ambiti della società – dall'assistenza sanitaria alla finanza, dai trasporti al commercio al dettaglio, dalla formazione alla comunicazione – e che man mano che la tecnologia avanza, essa svolge un ruolo sempre più importante e pervasivo. Al riguardo, sarà naturalmente necessario regolamentare l'uso dei sistemi di IA, anche attraverso strumenti di *soft law* e linee guida.

Relativamente ai Paesi più poveri, l'impatto dell'IA sarà ancora più complesso perché a fronte degli indubbi benefici in termini di accesso alla sanità e all'istruzione, nonché in generale di efficienza, produttività e nuove opportunità economiche, vi sono rischi significativi legati alla mancanza delle infrastrutture, risorse e competenze necessarie per regolamentarla e controllarne efficacemente l'uso, il che potrebbe portare a conseguenze indesiderate e danni a popolazioni vulnerabili. Inoltre, si potrebbero esacerbare le disuguaglianze esistenti e crearne di nuove, poiché coloro che dispongono di maggiori risorse e più facile accesso alla tecnologia potrebbero trovarsi in una posizione molto migliore per raccogliere i benefici dell'IA.

Infine, vanno considerati, per tutti i Paesi, i rischi politici associati all'Intelligenza Artificiale, che evidenziano l'importanza di una cooperazione e regolamentazione che siano internazionali. Tra questi rischi: (i) la polarizzazione, poiché l'IA potrebbe essere utilizzata per diffondere disinformazione e manipolare l'opinione pubblica, alimentando le divisioni e i conflitti nella società; (ii) la concentrazione del potere nelle mani di poche aziende o organizzazioni, il che può avere conseguenze significative per la democrazia e la responsabilità politica; (iii) la sorveglianza e il controllo di massa, con serie preoccupazioni in merito alla privacy, alla libertà individuale e ai diritti umani fondamentali; (iv) la mancanza di responsabilità, cui si è già accennato, nel senso di non essere in grado di determinare chi ha causato cosa; (v) la perdita di posti di lavoro e i disordini politici conseguenti; (vi) l'incremento della competizione geopolitica e possibili nuove o maggiori tensioni internazionali; (vii) le questioni etiche relative al ruolo della tecnologia nella società e a come le decisioni prese dai sistemi di IA dovrebbero essere guidate da considerazioni morali.

Fin qui quello che l'IA, *id est* ChatGPT, pensava di se stessa il 3 febbraio 2023, letteralmente. *Letteralmente* a parte una revisione stilistica che chi scrive ha apportato traducendo in italiano il testo redatto da ChatGPT, nel tentativo, molto umano, di renderlo forse più gradevole, di certo meno impersonale; di scalfire la percezione di un prodotto confezionato in una specie di sistema fordista, utile perché realizzato con i criteri di un'affidabilità standard, ma non del tutto convincente per la stessa ragione.

I sistemi di Intelligenza Artificiale come ChatGPT si basano infatti fondamentalmente su tecniche di apprendimento automatico. Ciò significa che le soluzioni che essi propongono ai problemi posti dagli utilizzatori derivano dalla ricombinazione statistica di ciò che essi trovano nel web. Sebbene vi sia anche la presenza di tecniche di apprendimento supervisionato e per rinforzo, la struttura essenziale presuppone una 'orizzontalità' per cui non vi è distinzione tra quanto viene prodotto nel web, sugli stessi argomenti, da persone esperte e da persone *meno* esperte o addirittura *non* esperte.

Che ChatGPT produca risultati più accurati quando si tratta di argomenti molto specialistici e tecnici deriva dal fatto che di questi argomenti nel web "parlano" in misura preponderante persone qualificate, realmente esperte. Tuttavia, poiché evidentemente il maggiore impatto sociale di sistemi di IA come ChatGPT si ha relativamente al loro uso da parte degli utenti per questioni generali, che riguardano le loro vite, e non per questioni scientifiche, un'ovvia conseguenza del suo sistema di funzionamento è che tali sistemi danno una forma (apparentemente) logica a idee e visioni, sia fattuali sia valoriali, che non hanno un solido fondamento empirico o scientifico. In questo modo, si riproducono le condizioni che, baconianamente, generano i 'pregiudizi della mente', gli "Idòla", che si contrappongono alla verità e alla sua ricerca. Una ragione in più per riflettere sulla direzione che i sistemi di IA presumibilmente imprimeranno alle comunità su questioni della massima rilevanza per l'umanità e il suo futuro.

Dal febbraio scorso, da quando ChatGPT è stato interrogato al fine di riportare in questo volume, che in quel periodo prendeva forma, una

sorta di auto-valutazione etica dell'IA, nuove versioni della ChatBot si sono affastellate, nuovi competitori hanno guadagnato spazio nel mercato, nuove applicazioni sono state realizzate e si sono diffuse globalmente: le prospettive si sono evolute con una velocità di innovazione non riscontrabile in altri settori scientifici e tecnologici, con salti qualitativi che probabilmente non vi sono invece stati nell'analisi etico-giuridica, e antropologica, dell'impatto dell'IA, attuale e potenziale, sulle società e sulle economie del pianeta.

Il testo che segue, "L'Intelligenza Artificiale: *distingue frequenter*. Come giungere a una comunanza etica nella società del pluralismo", è una proposta della Consulta scientifica del Cortile dei Gentili, come sempre corale e polifonica, quale contributo alla riflessione su forme di intelligenza che sono una creazione umana e che hanno già raggiunto alcuni risultati comparabili a quelli ad oggi prodotti solo dall'umana ragione. Probabilmente, si tratta di un altro modo di comprendere la realtà, con rischi significativi e margini di miglioramento vertiginosi, da parte di sistemi straordinariamente sofisticati, rispetto ai quali il volume vuole offrire una guida che speriamo utile per l'interpretazione dei fenomeni in corso, *sine ira ac studio*.

L'INTELLIGENZA ARTIFICIALE: *DISTINGUE FREQUENTER*. COME
GIUNGERE A UNA COMUNANZA ETICA NELLA SOCIETÀ DEL
PLURALISMO

Consulta Scientifica del Cortile dei Gentili

L'Intelligenza Artificiale sta sempre di più assumendo lo stato di questione centrale dei nostri tempi. Lo è dal lato dei suoi sviluppi scientifici e delle applicazioni tecnologiche, e lo è dal lato delle valutazioni e delle azioni che nei suoi confronti vengono intraprese sia dalle istituzioni politiche sia dalle agenzie morali, e in primo luogo tra queste da parte delle religioni.

Va innanzitutto notato come l'espressione "Intelligenza Artificiale" non abbia una definizione sufficientemente rigorosa, ed è possibile che continuerà a non averla. L'espressione comprende in realtà cose molto diverse tra loro. Alcune riguardano la ricerca fondamentale nell'ambito della computazione. Altre riguardano i tentativi di simulazione del pensiero umano oppure applicazioni delle tecnologie derivanti dai nuovi strumenti di computazione alla razionalizzazione dei processi produttivi in tutti gli ambiti, da quelli dei servizi a quelli della manifattura, a quelli dell'agricoltura. Altre ancora le applicazioni delle tecnologie di computazione al controllo e all'organizzazione del comportamento umano, a livello sia individuale sia collettivo.

Di fronte alle domande fondamentali dal punto di vista etico sull'Intelligenza Artificiale, si mostra ancora una volta tutta la profondità del principio metodologico tradizionale della scuola gesuitica: "*Numquam nega raro adfirma distingue frequenter*".

È argomentabile che una parte non secondaria del basso grado di chiarezza che caratterizza le discussioni pubbliche sull'Intelligenza Artificiale sia dovuta proprio alla forza evocativa dell'espressione, alla quale non corrisponde necessariamente un contenuto descrittivo adeguato. Più di quarant'anni fa, il grande filosofo Karl Raimund Popper contestò fortemente che avesse un contenuto epistemologicamente valido dare

alla ricerca sulla computazione e sulla simulazione del pensiero umano la qualificazione di “intelligenza artificiale”. Gli algoritmi di computazione, come processi e come prodotti, possono rappresentare soltanto una parte, e non quella più significativa e importante, dell’intelligenza umana. La disponibilità oggi di strumenti di calcolo sempre più potenti non fa venir meno questa fondamentale verità.

Le questioni etiche che si pongono nei confronti dell’Intelligenza Artificiale difficilmente possono riguardare i progressi che vengono fatti sia nell’ambito della computazione sia in quello della simulazione del pensiero umano. Si tratta di progressi conoscitivi, non diversi da quelli che avvengono negli ambiti della fisica o della chimica o della biologia. Quando si considerano questi ambiti, si deve sempre tenere presente lo straordinario apporto che essi stanno dando a ogni settore della ricerca scientifica. Come è noto, ad esempio, sono stati i risultati del *deep learning* – che è una delle ramificazioni più avanzate dell’Intelligenza Artificiale – che hanno permesso di ottenere in tempi incomparabilmente più brevi del passato i vaccini contro SARS-CoV-2, il virus che causa Covid-19.

Ovviamente, anche questo aspetto dell’Intelligenza Artificiale è comunque soggetto a scrutinio etico, esattamente come avviene per ogni campo della conoscenza scientifica. Lo scrutinio riguarda sia la conformità delle procedure ai canoni etici fondamentali e alla deontologia, sia le potenziali conseguenze negative che possono derivare dall’utilizzo dei suoi risultati. Non vi è infatti nulla di assoluto nell’attività umana, nulla che non debba essere ricondotto al rispetto dei principi etici fondamentali.

L’Intelligenza Artificiale non nasce da una volontà negativa. Non nasce da una volontà di dominio. Nasce da una volontà conoscitiva. Ma sappiamo che la volontà conoscitiva può condurre a conseguenze lontane dai principi dell’etica. Come disse nel 1947 J. Robert Oppenheimer, tra gli inventori della bomba atomica, «in un senso crudo che non potrebbe essere cancellato da nessuna accezione volgare o umoristica, i fisici hanno conosciuto il peccato». D’altra parte, già nel 1950 Norbert Wiener, uno dei padri della cibernetica e dell’Intelligenza Artificiale, aveva ammonito circa l’utilizzazione inumana degli esseri umani ove si cedessero alle macchine gli umani poteri di scelta e di controllo.

Il “peccato” può non nascere da una volontà individuale. Esso può essere il risultato oggettivo di un concatenarsi di azioni individuali ognuna delle quali, singolarmente presa, è conforme alla morale, ma che combinate producono effetti moralmente inaccettabili. È in ciò il senso della nozione di struttura di peccato coniata da Giovanni Paolo II nell'enciclica *Sollicitudo rei socialis* (1987). Le scienze sociali hanno esplorato a fondo il fenomeno che David Hume definì “il risultato dell'azione umana ma non di un progetto umano”. Sappiamo che un'irrazionalità collettiva può derivare – e molto spesso deriva, come nel caso dell'inquinamento – dalla somma di innumerevoli azioni umane razionali. Soltanto l'azione di regole morali e giuridiche può evitare che questo si produca.

Oggi gli sviluppi dell'Intelligenza Artificiale, in ogni suo aspetto, presentano esattamente questa fenomenologia.

Vi sono aspetti dell'*applicazione* dell'Intelligenza Artificiale che sono chiaramente contrari ai principi etici: controllare i comportamenti ed ancor più le idee individuali sino a ridurre di fatto a finzione la libera scelta delle persone; trasformare la società in un mero aggregato di monadi, restringendo fortemente le comunità e le socialità naturali (per poi paradossalmente riproporle ma in mondi virtuali come il Metaverso); spingere la razionalizzazione dei processi produttivi sino al punto di togliere dignità alle lavoratrici ed ai lavoratori; imporre un modello algoritmico alle decisioni dei giudici e delle autorità di vigilanza. Opporsi a questo significa eminentemente volere regole giuridiche vincolanti che impediscano la reificazione della persona umana e, più in generale, l'artificializzazione dell'uomo. Il fondamento di queste regole è largamente contenuto nei valori costituzionali dei Paesi liberaldemocratici, e non solo. Il diritto positivo può dispiegare la sua azione secondo un modello che è conosciuto da sempre, riguardo sia i diritti individuali sia i diritti sociali. Un modello ovviamente da adeguare alle nuove circostanze, che si ispiri eticamente alla visione umano-centrica nel rapporto uomo-macchina. Una visione che riconosca alcune dimensioni antropologiche insostituibili e non simulabili quali l'autoconsapevolezza, la libertà e la relazionalità interpersonale, che affermi l'esigenza di mantenere un controllo umano significativo e che faccia propria l'esigenza di evitare eccessive deleghe tecnologiche mantenendo per un verso la sinergia complementare tra uomo e macchina, e

per un altro costruendo tecnologie dell'Intelligenza Artificiale affidabili, spiegabili (nei limiti del possibile), tracciabili, trasparenti e inclusive.

Certo è che la sfida maggiore della regolazione delle applicazioni dell'Intelligenza Artificiale dal punto di vista dei valori etico-costituzionali è quella di contrastare il fortissimo *driver* di stampo utilitaristico che esse stanno imprimendo a tutto il mondo della produzione economica. Se si vuole, una sorta di neo-taylorismo diffuso e generalizzato, che però non si limita più al settore della produzione industriale.

Uno dei temi centrali per il prossimo futuro è quello delle regole e della governance delle tecnologie dell'Intelligenza Artificiale. In tutti i campi di applicazione la società ha trovato e può trovare nella nuova strumentazione tecnologica un supporto, a volte assai utile, alla propria sussistenza e sostenibilità, ma nulla ancora nel panorama dello sviluppo tecnologico ha dato vita a forme soddisfacenti di riproduzione dei processi umani e sociali più complessi, di sostituzione delle forme di aggregazione e relazione tra individui, dei rapporti terapeutici ed educativi, delle attività lavorative. Oltre che sui rischi di un annullamento dell'umano, o di una sua radicale trasformazione in altro da sé, le scienze sociali richiamano quindi l'attenzione sugli strumenti di governo, controllo e orientamento dello sviluppo tecnologico, e sulla riflessione etica rispetto all'armonizzazione dei percorsi di ricerca tecnologica con i principi della dignità umana e della giustizia allocativa.

Una questione importante che si pone è come far sì che le regole giuridiche non blocchino o limitino fortemente un giusto progresso tecnologico, ovvero un progresso che migliori le prospettive di vita dei cittadini. Non si tratta solamente di un miglioramento della produttività. Le tecnologie dell'Intelligenza Artificiale già oggi hanno apportato progressi notevoli nel campo della medicina, della gestione del territorio, della lotta all'inquinamento, dell'uso razionale delle risorse scarse come l'acqua. Esse possono avere un ruolo fondamentale nel migliorare la qualità della vita, e aiutare ad avere un ambiente migliore.

Regolare l'Intelligenza Artificiale temperando le diverse esigenze è probabilmente la sfida maggiore che i nostri sistemi si trovano ad affrontare. Con il corollario che l'evoluzione delle tecnologie "invasive" appare essere molto più veloce ed imprevedibile degli strumenti del dirit-

to positivo a disposizione dei nostri sistemi politici. D'altro canto, la globalizzazione del sistema economico espone i singoli Paesi ad asimmetrie fortissime nei costi e benefici della regolazione delle applicazioni dell'Intelligenza Artificiale in molti ambiti sociali. Come per la gran parte delle altre regole dell'economia – ad esempio, quelle che riguardano i diritti dei lavoratori – è difficilmente immaginabile, in ogni prospettiva storica realistica, che si arriverà ad avere regole globali, che riguardino se non tutti i Paesi del mondo almeno quelli più importanti e popolosi. In questo, tuttavia, l'Europa potrebbe avere una funzione propulsiva ed essere la prima, globalmente parlando, a indicare un modello cui riferirsi.

Distingue frequenter. Le regole e gli obblighi giuridici relativi alle *applicazioni* sono distinti da quelli che devono riguardare la corretta direzione della *ricerca* sull'Intelligenza Artificiale. Specialmente di quella parte di essa che mira a riprodurre attraverso gli *artifici* – come si sarebbe detto nel Rinascimento – il pensiero umano e, secondo certe linee di ricerca, anche le emozioni e i sentimenti umani. Non solo conoscere, ma appunto *riprodurre*.

Qui la dimensione esclusivamente, o anche prevalentemente, giuridica, è largamente insufficiente. Lo è per la natura stessa dello strumento giuridico, che elabora norme volte a evitare danni certi o probabili che riguardino singole persone o l'intera società. Ciò può avvenire, appunto, quando si tratti di evitare controlli pervasivi sulle vite dei cittadini. Ma qui ci troviamo di fronte a ciò che i giuristi definiscono “interessi adepti”, come sono quelli delle generazioni future. Difficilmente le norme giuridiche possono essere di guida o di contenimento di comportamenti le cui conseguenze negative non sono né certe né determinabili al presente e tantomeno al futuro, e che non consentono quindi quel giudizio di valore o disvalore proprio della norma giuridica.

Qui si dispiega lo spazio proprio e specifico dell'etica. Spetta alla coscienza dei ricercatori, come singole persone e come comunità, svolgere il ruolo di guida nel decidere su cosa fare ricerca, e come farla. È affar di coscienza valutare i rischi per le generazioni presenti e per quelle future che possano derivare dallo spingere la frontiera delle conoscenze oltre un limite tale che le ricadute tecnologiche potrebbero entrare in contrasto con i valori fondamentali della persona umana.

In questa attenzione costante, però, la coscienza dei ricercatori non deve rimanere sola. Perché essa si nutre della riflessione comune che deriva dalle concezioni che vengono elaborate e mantenute dalle comunità morali, a partire dalle comunità spirituali e religiose. È in queste comunità morali che si elaborano e si rivedono continuamente le visioni antropologiche che stanno alla base della determinazione non dei mezzi – ché questi spettano alla razionalità calcolante – ma dei fini propri della persona e della società. Senza una antropologia robusta, le regole morali sono affermazioni di un mero “dover essere” privo di un fondamento universale, e quindi soggetto al flusso del relativismo storico.

Questa determinazione deve unire gli sforzi tanto di chi condivide una morale derivante da una fede religiosa e quindi da valori trascendenti, quanto di chi condivide una morale basata su valori immanenti ma che non per questo siano destinati a dissolversi nel relativismo. Il problema serio, ancora in attesa di soluzione, è dunque quello di come giungere a una comunanza etica (la *Koinotes* aristotelica) nella società del pluralismo.

Il Documento è stato elaborato con il contributo di tutti i componenti della Consulta scientifica del Cortile dei Gentili (www.cortiledeigentili.com) ed è stato approvato il 28 ottobre 2022. Ha curato le successive versioni Cinzia Caporale. Osservazioni e commenti possono essere inviati all'indirizzo di posta elettronica segreteria@cortile.va.

LE STAGIONI DELL'IA

Paolo Benanti

INFORMAZIONE O CONTROLLO?

Viviamo in una società e in un tempo caratterizzati dal digitale, la *Digital Age*, un periodo complesso a causa dei profondi cambiamenti che queste tecnologie stanno producendo. La pandemia di Covid-19 ha accelerato una serie di processi che stavano già da tempo cambiando radicalmente la società perché è stato possibile sganciare il contenuto, la conoscenza, dal suo supporto¹. Il cambio d'epoca che stiamo attraversando è prodotto dalla tecnologia digitale e dal suo impatto sul nostro modo di comprendere noi stessi e la realtà che ci circonda.

Per capire questa sfida dobbiamo tornare all'inizio di questa trasformazione. In un documentario granuloso girato ai *Bell Laboratories* nel 1952, il matematico e ricercatore dei *Bell Labs* Claude Shannon, si trova accanto a una macchina di sua costruzione. Costruita nel 1950, è stata uno dei primi esempi al mondo di apprendimento automatico: un topo robotico che risolve labirinti noto come *Theseus*. Il Teseo dell'antica mitologia greca ha navigato nel labirinto di un minotauro ed è fuggito seguendo un filo che aveva usato per segnare il proprio percorso. Ma il giocattolo elettromeccanico di Shannon è stato in grado di "ricordare" il percorso con l'aiuto di interruttori di relè telefonici.

Nel 1948, Shannon aveva introdotto il concetto di teoria dell'informazione in *A Mathematical Theory of Communication*, un documento che fornisce la prova matematica che tutta la comunicazione può esse-

¹ Si pensi a fenomeni come le cosiddette *fake news*, la comparsa dello *sharp power*, i fatti di Capitol Hill o la Brexit nell'ambito pubblico o come il digitale plasmi le aspettative e i modi delle relazioni romantiche con piattaforme e modalità prima inedite, solo per citare alcuni esempi.

re espressa digitalmente. Claude Shannon ha mostrato che i messaggi potevano essere trattati puramente come una questione di ingegneria. La teoria matematica e non semantica della comunicazione di Shannon astrae dal significato di un messaggio e dalla presenza di un mittente o di un destinatario umano; un messaggio, da questo punto di vista, è una serie di fenomeni trasmissibili ai quali si può applicare una certa metrica².

Queste sue intuizioni diedero vita a una visione della realtà nuova e di matrice trans-disciplinare: la cibernetica di Norbert Wiener. Per Wiener, la teoria dell'informazione è un modo potente di concepire la natura stessa. Mentre l'universo sta guadagnando entropia in accordo con la seconda legge della termodinamica – cioè, la sua distribuzione di energia sta diventando meno differenziata e più uniforme – ci sono sistemi locali contro-entropici. Questi sistemi sono gli organismi viventi e le macchine elaboratrici di informazioni che costruiamo. Tali sistemi si differenziano e si organizzano: generano informazioni³. Il privilegio di questo approccio è ciò che permette alla cibernetica di esercitare un sicuro controllo nell'ambito interdisciplinare che genera e di cui si occupa: «la cibernetica può esser già sicura della sua “cosa”, vale a dire di calcolare tutto ciò che è nei termini di un processo controllato»⁴.

A partire dal decennio precedente la Seconda guerra mondiale, e accelerando durante la guerra e dopo, gli scienziati hanno progettato sistemi meccanici ed elettrici sempre più sofisticati che permettevano alle loro macchine di agire come se avessero uno scopo. Questo lavoro intersecò altri lavori sulla cognizione negli animali e i primi lavori sull'informatica. Ciò che emerse fu un nuovo modo di vedere i sistemi, non solo meccanici ed elettrici, ma anche i sistemi biologici e sociali: una teoria unificante dei sistemi e della loro relazione con il loro ambiente. Questo passaggio verso “interi sistemi” e “pensiero di sistema” divenne

² Cfr. Polt R., *A Heideggerian Critique of Cyberbeing*, in Pedersen H., Altman M. (a cura di), *Horizons of Authenticity in Phenomenology, Existentialism, and Moral Psychology*, Dordrecht, 2015: 181.

³ Cfr. *Ibidem*.

⁴ Heidegger M., Fabris A. (curatore), *Filosofia e cibernetica*, Pisa, 1988: 34-35.

noto come cibernetica. La cibernetica inquadra il mondo in termini di sistemi e dei loro obiettivi.

Secondo la cibernetica i sistemi raggiungono gli obiettivi attraverso processi iterativi o cicli di *feedback*. Improvvisamente, i maggiori scienziati del dopoguerra stavano parlando seriamente di causalità circolare (A causa B, B causa C e, infine, C causa A). Guardando più da vicino, gli scienziati videro la difficoltà di separare l'osservatore dal sistema. In effetti, il sistema sembrava essere una costruzione dell'osservatore. Il ruolo dell'osservatore è quello di fornire una descrizione del sistema, che viene fornita ad un altro osservatore. La descrizione richiede un linguaggio. E il processo di osservazione, creazione di linguaggio e condivisione delle descrizioni crea una società. Dalla fine degli anni Quaranta del secolo scorso il mondo della ricerca più avanzato ha iniziato a guardare alla soggettività – del linguaggio, della conversazione e dell'etica – e alla sua relazione con i sistemi e col *design*. Diverse discipline stavano collaborando per studiare la “collaborazione” come categoria di controllo.

Fino a quel momento i fisici avevano descritto il mondo in termini di materia ed energia. La comunità cibernetica propose una nuova visione del mondo attraverso la lente delle informazioni, dei canali di comunicazione e della loro organizzazione. In questo modo, la cibernetica è nata agli albori dell'era dell'informazione, nelle comunicazioni pre-digitali e nei media, colmando il modo in cui gli esseri umani interagiscono con macchine, sistemi e l'uno con l'altro. La cibernetica si concentra sull'uso del feedback per correggere gli errori e raggiungere gli obiettivi: la cibernetica fa della macchina e dell'uomo una sorta di topo di Shannon.

È a questo livello che dobbiamo guardare con maggiore attenzione gli effetti che tutto questo può avere sul capire e capirsi dell'uomo e sulla libertà. Con la maturità delle discussioni, gli obiettivi della comunità cibernetica si sono espansi. Nel 1968, Margaret Mead stava contemplando l'applicazione della cibernetica ai problemi sociali: «Con l'allargamento della scena mondiale, vi è la continua possibilità di utilizzare la cibernetica come forma di comunicazione in un mondo di crescente specializzazione scientifica. ... dovremmo considerare molto seriamente l'attuale situazione della società americana, all'interno della quale speriamo di poter sviluppare questi modi molto sofisticati di gestire sistemi che, in

effetti, hanno un disperato bisogno di attenzione. Problemi delle aree metropolitane... Le interrelazioni tra i diversi livelli di governo, la ridistribuzione del reddito, ... i collegamenti necessari tra parti di grandi complessi industriali...»⁵.

L'approccio cibernetico come sottolineerà Martin Heidegger rileggendo Wiener e l'opera dei cibernetici, "riduce" la stessa attività umana, nella pluralità delle sue configurazioni, a un qualcosa di funzionante e controllabile dalla macchina: «l'uomo stesso diviene "qualcosa di pianificato, cioè di controllabile" e, posto che una tale riduzione non sia possibile viene messo fra parentesi quale "fattore di disturbo" nel calcolo cibernetico»⁶. Nota infatti Fabris che:

Nella sua analisi del fenomeno cibernetico Heidegger tiene costantemente presente la matrice greca del vocabolo e privilegia questo aspetto, piuttosto che – ad esempio – la nozione centrale di *feedback*, quale filo conduttore per comprendere e spiegare le caratteristiche di una tale «disciplina non disciplina». Nella lettura heideggeriana la cibernetica indica l'avvento di un processo di controllo e d'informazione all'interno delle differenti sfere tematiche delle varie scienze. Il comando e il controllo (la *Steuerung*) sono intesi anzitutto, da un punto di vista ermeneutico, come quella prospettiva all'interno della quale vengono regolati i rapporti dell'uomo col mondo⁷.

Nel cuore dei ciberneti, cioè di quegli studiosi che sono i padri della società informatica, delle intelligenze artificiali e di tutti questi impres-

⁵ La traduzione è mia ed è tratta da: Mead M., *Cybernetics of Cybernetics*, Foerster von H. et alia (a cura di), *Purposive Systems. Proceedings of the First Annual Symposium of the American Society for Cybernetics*, New York, 1968: 4-5.

⁶ Heidegger M., Fabris A. (curatore), *Filosofia e cibernetica*, cit.: 10.

⁷ Ivi, 11. Annota Fabris che «la cibernetica è vista da Heidegger come il momento più avanzato, l'esito più evidente di quel dominio della tecnica in cui sfocia l'intera metafisica occidentale. La storia dell'essere – come emerge dai corsi universitari su Nietzsche degli anni Trenta – ha infatti il suo punto d'arrivo nell'evento della tecnica, in cui trova piena manifestazione la volontà di potenza (volontà di volontà) che determina l'azione dell'uomo e si estende a ogni ambito del reale. All'interno di un tale processo di autoriferimento della volontà il progetto cibernetico riceve la propria giustificazione e definisce i propri rapporti con la filosofia, facendosi carico di alcuni dei suoi compiti e assumendo le sue prerogative tradizionali» (*Ibidem*).

sionanti sviluppi che il digitale sta realizzando nel nostro vivere, però, potrebbe esserci stata la promessa di uno scopo ancora più grande.

Gregory Bateson, primo marito di Margaret Mead, in una celebre intervista affermò che ciò che lo entusiasmava nelle discussioni sulla cibernetica era che: «Era una soluzione al problema dello scopo. Da Aristotele in poi, la causa finale è sempre stata il mistero. Questo venne fuori allora. Non ci rendevamo conto allora (almeno io non me ne rendevo conto, anche se McCulloch potrebbe averlo fatto) che l'intera logica avrebbe dovuto essere ricostruita per la ricorsività»⁸.

INTELLIGENZE ARTIFICIALI, ALGORITMI E LORO IMPATTO SOCIALE

Il mondo della cibernetica ha dato vita e nutrito l'idea di poter realizzare in silico una macchina che replicasse le funzioni cognitive umane: l'intelligenza artificiale (IA). L'intelligenza artificiale è una disciplina giovane, ha appena sessant'anni, che nasce da un insieme di scienze, teorie e tecniche e che mira a imitare le capacità cognitive di un essere umano. All'interno delle IA troviamo nozioni che sono mutate da: logica matematica, statistica, probabilità, neurobiologia computazionale e informatica. Questi studi si sono avviati nel finire della Seconda guerra mondiale e i suoi sviluppi sono intimamente legati a quelli dell'informatica: le IA hanno portato i computer a svolgere compiti sempre più complessi, che prima potevano essere delegati solo a un essere umano. Vogliamo provare a ripercorrere le tappe salienti di questo sviluppo⁹.

⁸ La traduzione è mia ed è tratta da: Brand S., *For God's Sake, Margaret a conversation with Margaret Mead and Gregory Bateson*, *CoEvolutionary Quarterly*, 10-21 giugno 1976: 32-44. La teoria delle cause che Aristotele ha introdotto in *Fisica* II 3-7, in *Metafisica* Δ 2, in *Metafisica* A 3-10 e in *Analitici Posteriori* II 111, ha destato fin dagli esordi diverse discussioni. L'importanza della teoria aristotelica delle cause è dovuta soprattutto al fatto che, a partire da Aristotele in poi, possiamo parlare di conoscenza quando possiamo dar conto dei principi e delle cause che hanno avuto un ruolo nel realizzarsi di un certo evento.

⁹ Per la ricostruzione di questa breve panoramica storica ci rifacciamo a un articolo molto importante in questo campo: Buchanan B.G., *A (Very) Brief History of Artificial Intelligence*, *AI Magazine* 26-4, 2005: 53-60. Bruce G. Buchanan è stato membro fondatore dell'*American Association for Artificial Intelligence* (AAAI) e ne ha ricoperto

Il termine intelligenza artificiale fu coniato durante un evento, il *Dartmouth Workshop*, del 1956 finanziato dal *Rockefeller Institute*. L'evento fu voluto da Marvin Minsky, John McCarthy, Claude Shannon e Nathan Rochester di IBM¹⁰. Il rationale della conferenza si proponeva di

cariche apicali dal 1986 fino al 2001. È considerato una delle voci più autorevoli dell'AI. Sono consapevole che una storia abbreviata lascia necessariamente fuori molti autori chiave e pietre miliari nello sviluppo dei sistemi. L'idea è quella di offrire una panoramica delle prospettive più che quella di comporre una vera e propria storia.

¹⁰ Ci occupiamo in questa ricostruzione dei sistemi *software* e *hardware* che hanno dato vita all'AI. Tuttavia, abbiamo ben presente che: «i filosofi da molti secoli hanno ventilato la possibilità di macchine intelligenti come un dispositivo letterario per aiutarci a definire cosa significa essere umani. René Descartes, per esempio, sembra essere stato più interessato all'«uomo meccanico» come metafora che come possibilità. Gottfried Wilhelm Leibniz, d'altra parte, sembrava vedere la possibilità di dispositivi di ragionamento meccanico che utilizzano regole di logica per risolvere le controversie. Sia Leibniz che Blaise Pascal progettaron macchine calcolatrici che meccanizzavano l'aritmetica, che fino ad allora era stata la provincia di uomini dotti chiamati «calcolatori», ma non hanno mai affermato che i dispositivi potessero pensare. Etienne Bonnot, Abbé de Condillac ha usato la metafora di una statua nella cui testa abbiamo versato pepite di conoscenza, chiedendo a che punto avrebbe saputo abbastanza per sembrare intelligente. Gli scrittori di fantascienza hanno usato la possibilità di macchine intelligenti per avanzare la fantasia di non-umani intelligenti, così come per farci riflettere sulle nostre caratteristiche umane. Jules Verne nel XIX secolo e Isaac Asimov nel XX sono i più noti, ma ce ne sono stati molti altri, compreso L. Frank Baum, che ci ha dato il Mago di Oz. Baum scrisse di diversi robot e descrisse l'uomo meccanico Tiktok nel 1907, per esempio, come un «Uomo Meccanico Extra-Responsabile, Creatore di Pensieri, Parlante Perfetto... Pensa, Parla, Agisce, e Fa Tutto tranne Vivere». Questi scrittori hanno ispirato molti ricercatori di AI. I robot, e gli esseri creati artificialmente come il Golem nella tradizione ebraica e Frankenstein di Mary Shelly, hanno sempre catturato l'immaginazione del pubblico, in parte giocando sulle nostre paure. Animali meccanici e bambole – tra cui un trombettista meccanico per il quale Ludwig van Beethoven scrisse una fanfara – furono effettivamente costruiti con meccanismi a orologeria nel XVII secolo. Sebbene fossero ovviamente limitati nelle loro prestazioni e fossero intesi più come curiosità che come dimostrazioni di pensiero, fornirono una certa credibilità iniziale alle visioni meccanicistiche del comportamento e all'idea che tale comportamento non dovesse essere temuto. Man mano che il mondo industriale si meccanizzava, le macchine diventavano più sofisticate e più comuni. Ma era ancora essenzialmente un meccanismo ad orologeria. Gli scacchi sono ovviamente un'impresa che richiede pensiero. Non è troppo sorprendente, quindi, che le macchine giocatrici di scacchi del diciottesimo e diciannovesimo secolo, in particolare «il Turco», furono esibite

mettere a tema: «ogni aspetto dell'apprendimento o qualsiasi altra caratteristica dell'intelligenza che possa essere descritta in modo così preciso in modo da poter costruire una macchina per simularla»¹¹. I partecipanti includevano Ray Solomonoff, Oliver Selfridge, Trenchard More, Arthur Samuel, Allen Newell e Herbert A. Simon: questi furono i padri della disciplina e gli autori dei più importanti programmi di ricerca sull'IA nei decenni successivi. Fu John McCarthy a convincere i partecipanti ad accettare "Artificial Intelligence" come nome del nuovo campo di studi. Il nome fu scelto da McCarthy per evitare associazioni con la cibernetica e collegamenti con l'influente cibernetista Norbert Wiener¹². Infatti, nel periodo tra il 1940 e il 1955 si sono registrate forti congiunzioni negli sviluppi tecnologici, in parte grazie anche all'accelerazione impressa dalla Seconda guerra mondiale, accompagnate a un genuino desiderio di capire come far convivere il funzionamento delle macchine e degli esseri organici. Il contributo di Norbert Wiener, un pioniere della cibernetica, è evidente. Ha lavorato per unificare la teoria matematica, l'elettronica e l'automazione sviluppando un'intera teoria del controllo e della comunicazione, sia negli animali che nelle macchine. L'IA di fatto è stata influenzata da idee provenienti da molte discipline: dall'ingegneria (come

come macchine intelligenti e persino ingannarono alcune persone a credere che le macchine stessero giocando autonomamente. Samuel L. Clemens ("Mark Twain") scrisse in una rubrica di giornale, per esempio, che il Turco doveva essere una macchina perché giocava così bene! Gli scacchi sono stati ampiamente utilizzati come veicolo per studiare i meccanismi di inferenza e di rappresentazione nei primi decenni di lavoro dell'AI» (Buchanan B.G., *A (Very) Brief History of Artificial Intelligence*, *AI Magazine* 26-4, 2005: 53-54, la traduzione è mia).

¹¹ McCarthy J., Minsky M., Rochester N., Shannon C., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 31 August 1955 (<https://web.archive.org/web/20080930164306/http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html> accesso: 29.01.2022).

¹² Cfr. McCarthy J., *Review of The Question of Artificial Intelligence*, in Id., *Defending AI Research. A Collection of Essays and Reviews*, The Center for the Study of Language and Information Publications, Stanford, 1996: 73. Nel saggio di McCarthy si legge testualmente: «[O]ne of the reasons for inventing the term "artificial intelligence" was to escape association with "cybernetics". Its concentration on analog feedback seemed misguided, and I wished to avoid having either to accept Norbert (not Robert) Wiener as a guru or having to argue with him».

il lavoro di Norbert Wiener sulla cibernetica, che include il *feedback* e il controllo); dalla biologia (per esempio, W. Ross Ashby e il lavoro di Warren McCulloch e Walter Pitts sulle reti neurali in organismi semplici); dalla psicologia sperimentale (come Newell e Simon); dalla teoria della comunicazione (per esempio, il lavoro teorico di Claude Shannon); dalla teoria dei giochi (in particolare di John Von Neumann e Oskar Morgenstern); dalla matematica e dalla statistica (per esempio, Irving J. Good); dalla logica e dalla filosofia (per esempio, Alan Turing, Alonzo Church e Carl Hempel), e dalla linguistica (come il lavoro di Noam Chomsky sulla grammatica)¹³. Queste radici hanno lasciato una profonda impronta nella struttura di questo campo e continuano a segnare notevolmente le ricerche attuali. È interessante notare come l'IA oggi sia cresciuta andando anche al di là delle conoscenze che l'avevano vista formarsi e, in diverse circostanze, i suoi risultati hanno influenzato lo sviluppo di queste discipline.

All'inizio del 1950 John Von Neumann e Alan Turing realizzarono il passaggio da macchine basate su logica decimale del XIX secolo alle macchine basate su logica binaria basate su algebra booleana. I due ricercatori, che lavorarono prima che il termine IA fosse formalizzato, diedero però vita all'architettura che ancora oggi caratterizza i nostri computer. La loro opera ha dimostrato che il neonato *computer* era una macchina universale, in grado di eseguire quanto programmato algoritmicamente. Turing, inoltre, ha sollevato per la prima volta la questione della possibile intelligenza di una macchina in un famoso articolo del 1950¹⁴ e l'ha posta all'interno di un ipotetico "gioco di imitazione", dove la soglia dell'intelligenza della macchina si situa nel non far riuscire a un essere umano a distinguere in un dialogo telescrivente se sta parlando con un uomo o una macchina. Per quanto controverso possa essere questo articolo a partire dalla sua pubblicazione sarà spesso citato come l'origine della messa in discussione del confine tra l'uomo e la macchina.

¹³ Cfr. Buchanan B.G., *A (Very) Brief History of Artificial Intelligence*, cit.: 53–60.

¹⁴ Cfr. Turing A. M., *Macchine calcolatrici e intelligenza*, Somenzi V., Cordeschi R. (a cura di), *La Filosofia degli automi. Origini dell'intelligenza artificiale*, Torino, 1986: 157–183.

Le macchine pensate da Turing sono, seguendo il testo del suo articolo, macchine a stati finiti – computer – in grado di simulare altre macchine a stati discreti – l'uomo nel suo atto comunicativo¹⁵. Una macchina per sostenere il test doveva essere programmata considerando la descrizione di un uomo in termini discreti: stati interni, segnali, simboli. Dalla complessità del *software*, si legge tra le righe dell'articolo, emergeranno le funzioni intellettuali. Su questa aspettativa si fonda una disciplina nota come intelligenza artificiale il cui scopo è la costruzione di una macchina in grado di riprodurre le funzioni cognitive umane.

A partire dal 1956 l'IA ha conosciuto periodi alterni: a momenti di forte innovazioni sono seguiti dei periodi di vuoto che i membri della comunità scientifica chiamano allegoricamente “inverno”. Proviamo a vedere le stagioni principali dell'AI.

1940-1960: NASCITA DELL'IA SULLA SCIA DELLA CIBERNETICA

Le principali realizzazioni di questi anni furono: *Unimate* il primo robot industriale che nel 1961 entrò in funzione presso gli stabilimenti della *General Motor* rimpiazzando i primi uomini nelle catene di montaggio, *Eliza* il primo *chatbot* sviluppato da Joseph Weizenbaum al MIT in grado di sostenere delle conversazioni con gli umani e *Shakey* il primo robot mobile *general-purpose* realizzato a Stanford in grado di “ragionare” sulle proprie azioni. Sebbene la tecnologia sia rimasta affascinante e promettente, la popolarità di questo nuovo campo tecnologico è pian piano diminuito nei primi anni Sessanta. Le macchine avevano pochissima memoria, rendendo difficile l'uso di un linguaggio informatico per programmare le funzioni complesse. Tuttavia, c'erano già alcune basi ancora oggi presenti come gli alberi delle soluzioni per risolvere i problemi: l'IPL, linguaggio di elaborazione delle informazioni, aveva così permesso di scrivere già nel 1956 il programma LTM (*logic theorist machine*) che mirava a dimostrare teoremi. I linguaggi di manipolazione dei simboli come Lisp, IPL e POP e i sistemi di condivisione del tempo – in cima ai progressi dell'hardware sia nei processori che nella memoria – die-

¹⁵ Cfr. *Idem*.

dero ai programmatori nuovo potere negli anni '50 e '60. Nonostante i limiti tecnologici, ci furono numerose dimostrazioni impressionanti di programmi che risolvevano effettivamente problemi che solo persone intelligenti erano state precedentemente in grado di risolvere. Il programma di Arthur Samuel per giocare a dama, scritto negli anni '50, era impressionante date sia le limitazioni dell'*hardware* IBM 704 per il quale il programma era stato scritto come test di verifica, sia le limitazioni del linguaggio *assembly* in cui era scritto. Il programma di Samuel è impressionante perché il programma ha imparato attraverso l'esperienza a migliorare la propria capacità di giocare a dama – giocando contro avversari umani e contro altri computer¹⁶. In questi anni si ottennero anche le prime dimostrazioni del potere di un approccio euristico al tema dell'AI e si stabilì così un principio fondamentale a cui Herbert Simon, economista e sociologo, diede il nome di "satisficing" e che possiamo riassumere così: in assenza di un metodo efficace che garantisca la soluzione di un problema in un tempo ragionevole, l'euristica può guidare un decisore verso una soluzione molto soddisfacente, anche se non necessariamente ottimale. Minsky ha riassunto gran parte del lavoro nella prima decade circa dopo il 1950:

L'idea centrale del periodo precedente al 1962 era quella di trovare dispositivi euristici per controllare l'ampiezza di una ricerca per tentativi ed errori. Una seconda preoccupazione era quella di trovare tecniche efficaci per l'apprendimento. Nell'era post-1962 la preoccupazione divenne meno sull'"apprendimento" e più sul problema della rappresentazione della conoscenza (comunque acquisita) e sul problema correlato di rompere la formalità e la ristrettezza dei vecchi sistemi. Il problema dell'efficienza della ricerca euristica rimane come vincolo di fondo, ma non è più il problema a cui si pensa, perché ora siamo immersi in sottoproblemi più sofisticati, per esempio, la rappresentazione e la modifica dei piani¹⁷.

Herbert Simon profetizzò nel 1957 che l'IA sarebbe riuscita a battere un essere umano a scacchi nei successivi 10 anni, ma poi accadde

¹⁶ Cfr. Buchanan B.G., *A (Very) Brief History of Artificial Intelligence*, cit.: 53-60.

¹⁷ Minsky M., *Semantic Information Processing*, Cambridge, 1968: 9 (la traduzione è mia).

quello che si chiama il primo inverno dell'IA e tutto rimase fermo. Oggi sappiamo che la profezia di Simon si è rivelata giusta solo trenta anni dopo. Nel 1968 Stanley Kubrick diresse il film *2001 Space Odyssey* dove un computer – HAL 9000 (una sigla in cui sostituendo ogni lettera con la seguente nell'alfabeto si ottiene IBM) riassume in sé tutta la somma delle questioni poste dall'IA: rappresenterà un alto livello di sofisticazione, un bene per l'umanità o un pericolo? L'impatto del film non sarà naturalmente scientifico ma contribuirà a rendere popolare il tema, proprio come l'autore di fantascienza Philip K. Dick, che non smetterà mai di chiedersi se, un giorno, le macchine vivranno emozioni. L'IA inizia a contaminare la cultura pop.

1980-1990: SISTEMI ESPERTI

Fu con l'avvento dei primi microprocessori alla fine del 1970 che si uscì dall'inverno in cui l'IA era caduta entrando in una nuova stagione di fioritura di soluzioni chiamate sistemi esperti. Questo percorso conosce due predecessori importanti: DENDRAL del MIT, un sistema esperto specializzato in chimica molecolare sviluppato nel 1965, e MYCIN della Stanford University, un sistema specializzato nella diagnosi delle malattie del sangue e dei farmaci da prescrivere sviluppato nel 1972. Entrambi erano basati su un "motore di inferenza", programmato per essere uno specchio logico del ragionamento umano. Inserendo i dati, il motore forniva risposte classificate come di alto livello di competenza. Ira Goldstein e Seymour Papert hanno descritto le dimostrazioni del programma DENDRAL a metà degli anni Sessanta come un "cambiamento di paradigma" per le AI. MYCIN e le migliaia di sistemi esperti che lo seguirono divennero dimostrazioni visibili del potere di piccole quantità di conoscenza di abilitare programmi di decisione intelligente in numerose aree importanti. Anche se di portata limitata, in parte a causa dello sforzo per accumulare la conoscenza richiesta, il loro successo nel fornire assistenza a livello di esperti rafforza il vecchio adagio che la conoscenza è potere¹⁸. La programmazione di tale conoscenza ha richiesto in realtà un grande sforzo e

¹⁸ Cfr. Buchanan B.G., *A (Very) Brief History of Artificial Intelligence*, cit.: 53-60.

da 200 a 300 regole note si è verificato un effetto “scatola nera” dove non era più chiaro come ragionasse la macchina. Lo sviluppo e la manutenzione diventavano così estremamente problematici e, soprattutto, più rapidi e in molti altri modi meno complessi e meno costosi.

Un altro campo che vide realizzazioni notevoli fu il linguaggio. Gli sviluppi provengono dal lavoro di ricerca fatto al MIT nei primi anni Sessanta da Danny Bobrow, Bert Raphael, Ross Quillian e Fischer Black: la comprensione e la traduzione del linguaggio erano all’inizio ritenute semplici, dato il potere dei computer di immagazzinare e recuperare parole e frasi in enormi dizionari. Alcuni esempi comici di fallimenti dell’approccio alla traduzione basato sulla ricerca di tabelle hanno fornito ai critici dell’AI abbastanza argomenti per fermare i finanziamenti sulla traduzione automatica per molti anni. Il lavoro di Danny Bobrow dimostrò che i computer potevano usare il contesto limitato dei problemi di parole dell’algebra per capirli abbastanza bene da risolvere problemi che avrebbero sfidato molti adulti. Ulteriori lavori di Robert F. Simmons, Robert Lindsay, Roger Schank e altri mostrarono in modo simile che la comprensione – anche una certa traduzione – era possibile solo in domini limitati.

Le promesse dell’IA in questa seconda primavera prevedevano uno sviluppo massiccio del settore ma di nuovo si è assistito a un inverno decennale tra la fine del 1980 e l’inizio del 1990. Va ricordato che negli anni Novanta il termine intelligenza artificiale era quasi diventato un tabù e variazioni più modeste erano addirittura entrate nel linguaggio universitario, come l’espressione “*advanced computing*”.

Il successo nel maggio 1997 di *Deep Blue*, il sistema esperto di IBM, alla partita a scacchi contro Garry Kasparov ha soddisfatto la profezia di Herbert Simon del 1957 a trenta anni di distanza. Ma questo successo non ha prodotto un ritorno economico in grado di ripagare il finanziamento e lo sviluppo di questa forma di IA. Il funzionamento di *Deep Blue* si basava su un algoritmo sistematico di forza bruta, in cui tutte le possibili mosse venivano valutate e ponderate. La sconfitta dell’umano è rimasta molto simbolica nella storia ma *Deep Blue* in realtà era riuscito solo a trattare un perimetro molto limitato (quello delle regole del gioco degli scacchi), molto lontano dalla capacità di modellare la complessità

del mondo. Tuttavia, dopo *Deep Blue*, l'ultima parte degli anni Novanta assistettero ad alcune interessanti realizzazioni: nel 1998 Cynthia Breazeal al MIT ha introdotto *KISmet* un robot dotato di una sorta di intelligenza emotiva in grado di riconoscere e interagire coerentemente con i sentimenti delle persone; nel 1999 la Sony ha lanciato sul mercato *AiBO*, un acronimo per IA robot, il primo cane robot destinato al consumo di massa con abilità e personalità che si sviluppavano nel tempo; nel 2002 iRobot ha introdotto *Roomba* il primo sistema autonomo di aspirazione e pulizia casalinga destinato al mercato di massa. A cavallo del nuovo secolo i sistemi dotati di IA iniziano ad uscire dalla cinematografia e iniziano ad abitare nei primi prodotti di consumo. Tuttavia un altro inverno attendeva le AI: tutta la prima decade di questo secolo non ha visto realizzare alcuna promessa di quelle che sembravano pronte a realizzarsi nella seconda fioritura dei sistemi esperti.

DAL 2010: DATI MASSICCI E NUOVA POTENZA DI CALCOLO

Due fattori spiegano il nuovo *boom* delle AI cominciato intorno al 2010. In primo luogo, è diventato possibile l'accesso a enormi volumi di dati. Per poter utilizzare algoritmi per la classificazione delle immagini e il riconoscimento dei gatti, ad esempio, in precedenza era necessario eseguire personalmente il campionamento delle foto, oggi, una semplice ricerca su Google può trovare milioni di foto già correttamente identificate. In secondo luogo, la scoperta che i processori delle schede grafiche per computer, fino ad allora destinate ai *videogames*, erano ad altissima efficienza per accelerare il calcolo degli algoritmi di apprendimento automatico. Poiché l'addestramento di un modello è un processo molto iterativo, prima del 2010 sarebbero state necessarie settimane per elaborare l'intero campione. La potenza di calcolo di queste sofisticatissime schede, capaci di oltre mille miliardi di transazioni al secondo, sviluppate grazie al nuovo mercato dei videogiochi che è passato dal 2012 al 2021 da 52 a 138 miliardi di dollari, ha consentito notevoli progressi a un costo finanziario contenuto – meno di 1000 euro a scheda.

Questa nuova generazione di *hardware* ha consentito alcuni significativi successi che hanno avuto una grande risonanza mediatica e ha

aumentato esponenzialmente i finanziamenti: nel 2011, *Watson*, l'AI di IBM vinto contro i campioni umani di *Jeopardy!*, un popolare quiz in cui data la risposta bisogna indovinare la domanda. Nel 2012, *Google X* – il laboratorio di ricerca di Google per le AI – è stato in grado di fare in modo che un'intelligenza artificiale insegnasse a se stessa a riconoscere i gatti in un video. Per quest'ultimo compito sono stati utilizzati più di 16.000 processori, ma il potenziale è straordinario: una macchina per la prima volta passa dalla visione alla percezione e impara a distinguere qualcosa attribuendole un nome. Nel 2016 *AlphaGo* – l'AI di Google, sviluppata dai laboratori di Deep Mind, batterà Fan Hui, la campionessa europea di Go e subito dopo Lee Sedol il campione del mondo. L'anno seguente, una nuova versione, *AlphaGo Zero*, ha battuto *AlphaGo*. È importante ricordare che il gioco del Go ha una combinazione di possibili mosse molto più grande degli scacchi, di fatto le mosse sono più del numero di particelle che compongono l'universo: per il gioco del Go non è possibile avere risultati così significativi utilizzando la forza grezza del calcolo come per *Deep Blue* nel 1997.

Tutto questo si associa ed è prodotto da un completo cambio di paradigma rispetto ai sistemi esperti. L'approccio è diventato induttivo: non si tratta più di codificare regole come per i sistemi esperti, ma di farle scoprire ai computer da soli per correlazione e classificazione, sulla base di un'enorme mole di dati. Oggi tra le tecniche di apprendimento automatico, il *deep learning* sembra essere la più promettente per numerose applicazioni, tra cui il riconoscimento vocale o delle immagini.

Il metodo ereditato dai sistemi esperti veniva detto *Symbolic AI*. Il modo con cui la *Symbolic AI* trattava la conoscenza è descrivibile in termini di "costanti" e "predicati". Le costanti sono semplicemente oggetti in un mondo di riferimento e i predicati sono relazioni tra gli oggetti. Un eccellente esempio di questa AI simbolica è un albero genealogico. Utilizzando un ragionamento deduttivo sui progenitori nell'albero genealogico, cioè partendo da un elenco iniziale di relazioni, possiamo dedurre nuove relazioni tra oggetti, cioè le relazioni di parentela sui discendenti. Questo modo, come abbiamo visto, fornisce una rappresentazione interpretabile del mondo di riferimento e del processo decisionale ed era il grande vantaggio dei sistemi esperti. Gli algoritmi generalizzano gli

aspetti concettuali a un livello superiore attraverso il ragionamento e la tracciabilità dei risultati. Il framework dell'AI simbolica fornisce al programmatore e al programma i blocchi costitutivi – sotto forma di costanti, simboli, relazioni –, che consentono di comprendere e dare ragione formale delle decisioni che il sistema suggerisce o mette in atto. Proprio la potenza di questo sistema ha rappresentato il suo più grande limite quando si è tentato di implementarlo nei sistemi di intelligenza artificiale. Fondamentalmente, la conoscenza del sistema deve essere ereditata manualmente affinché questo programma funzioni. Tornando all'esempio dell'albero genealogico questo significa che le relazioni possono essere organizzate e identificate con facilità, soprattutto perché il lettore dell'albero è in grado di leggere e comprendere le linee tracciate per stabilire connessioni chiare da un oggetto all'altro. Di fatto il cervello umano è capace di raccogliere grandi quantità di dati apparentemente neutri e stabilire schemi e connessioni attraverso la memoria sia a breve che a lungo termine estraendo dal contesto un senso che si fa portatore eventualmente di alcuni significati. Un computer non lo può fare e nel caso del mondo reale toccherebbe ai programmatori stabilire le connessioni tra tutti gli oggetti differenti che vengono datificati codificando tutto il significato apparentemente implicito nel sistema per rendere la macchina capace di estrarre un significato dai dati. I ricercatori in questa nuova primavera delle IA hanno iniziato ad utilizzare nuove famiglie di algoritmi che possano da soli fare parte di questo lavoro, le reti neurali, ad esempio, sono parte integrante di questo processo. Quello che si è generato è una nuova forma di IA definita *inexplicit*. Queste tecniche che di fatto partono dal voler emulare il funzionamento degli strati di neuroni che compongono il cervello umano hanno portato enormi risultati pratici ma hanno ulteriormente complessizzato il campo rendendo il meccanismo decisionale delle IA assolutamente inesplicabile in alcuni casi: le nuove IA spesso sono delle vere e proprie *black box* cognitive. Oltre alle tecniche di apprendimento per rinforzo sono stati introdotti anche paradigmi mutuati dalla teoria evolutiva in biologia applicata alla modalità euristica che caratterizzava la prima ondata delle IA. Un algoritmo evolutivo è sostanzialmente un algoritmo euristico che si ispira al principio di evoluzione degli esseri viventi. Volendo semplificare possia-

mo dire che un algoritmo evolutivo parte da una soluzione valida per un problema e tenta di modificarla cercando di farla evolvere con una serie di modifiche casuali fino a giungere ad una soluzione migliore o più performante di quella di partenza. John Haugeland ha coniato una espressione, *Good Old-Fashioned Artificial Intelligence* (GOFAI) per indicare il paradigma simbolico contribuendo così a diffondere l'idea di un processo che per quanto affascinante è destinato a scomparire perché non capace di portare soluzioni utili al mercato¹⁹.

Il passaggio a queste nuove forme di approccio al mondo della IA è senz'altro legato alla grandissima disponibilità di dati e alla nuova capacità di calcolo ma questi due fattori non bastano a spiegare quanto sta avvenendo. A partire dal 2010 intelligenze artificiali come *SIRI* di Apple e *Alexa* di Amazon hanno iniziato a diventare l'oggetto di grandi finanziamenti e forti investimenti da parte dei *Venture Capitalist* (VC). Le analisi degli investimenti globali dei VC in società private incentrate sulle IA rilevano che gli investimenti VC nell'IA sono cresciuti nell'ultimo decennio con ritmi impressionanti. Sono gli Stati Uniti e la Repubblica Popolare Cinese a guidare questa ondata di investimenti che tendono a concentrarsi su pochi settori chiave. L'Unione europea, il Regno Unito e il Giappone hanno aumentato gli investimenti, ma sono molto indietro rispetto ai due attori dominanti²⁰. I VC di fatto sono il volano che ha sottratto l'IA a discussioni filosofiche e scientifiche che toccavano la natura di questa forma di conoscenza, facendola focalizzare su soluzioni che potessero funzionare ed essere scalabili per un utilizzo massivo ed economicamente favorevole.

Nel 2003 Geoffrey Hinton dell'Università di Toronto, Yoshua Bengio dell'Università di Montreal e Yann LeCun dell'Università di New York hanno deciso di avviare un programma di ricerca per aggiornare le reti neurali. Esperimenti condotti simultaneamente presso Microsoft, Google e IBM con l'aiuto del laboratorio di Toronto di Hinton hanno di-

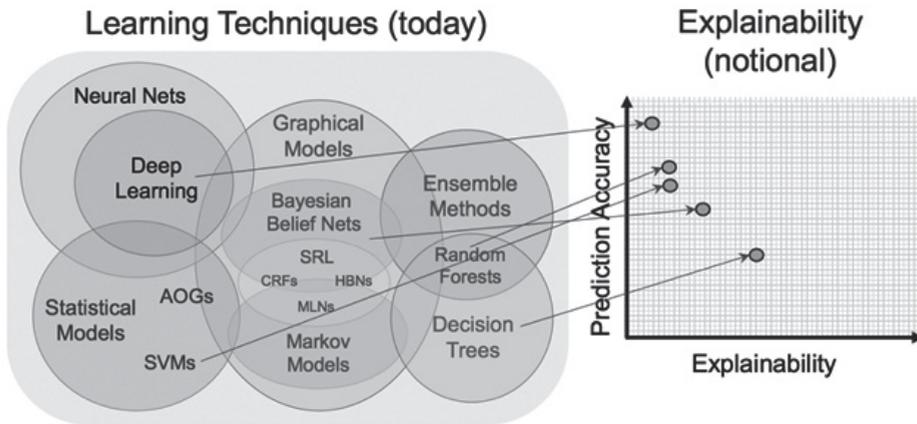
¹⁹ Cfr. Haugeland J., *Artificial Intelligence: The Very Idea*, Boston, 1985.

²⁰ Cfr. R. Tricot, *Venture capital investments in artificial intelligence: Analysing trends in VC in AI companies from 2012 through 2020*, *OECD Digital Economy Papers*, n. 319, OECD Publishing, Paris, 2021.

mostrato che questo tipo di apprendimento riesce a dimezzare i tassi di errore per il riconoscimento vocale. Risultati simili sono stati raggiunti dal *team* di riconoscimento delle immagini di Hinton.

La grande maggioranza dei gruppi di ricerca oggi usa questa tecnologia con vantaggi indiscutibili ma, secondo esperti come Yann LeCun, c'è ancora molta strada da fare per produrre sistemi di comprensione del testo. I *bot* conversazionali illustrano bene questa sfida: i nostri *smartphone* sanno già come trascrivere un'istruzione ma non riescono a contestualizzarla completamente e ad analizzare le nostre intenzioni.

Questo il panorama dei sistemi che oggi abbiamo sviluppato in termini di capacità predittiva e spiegabilità nozionale da parte del sistema²¹:



Possiamo dire quindi, a conclusione di questo percorso che l'IA come è stata definita da John McCarthy è «la scienza e l'ingegneria per creare macchine intelligenti, in particolare programmi per computer intelligenti. È correlata al compito simile di utilizzare i computer per comprendere l'intelligenza umana, ma l'IA non deve limitarsi a metodi biologicamente osservabili»²². L'apprendimento automatico o *machine learning* è un sot-

²¹ Immagine fonte DARPA (https://www.darpa.mil/attachments/XAIIndustryDay_Final.pptx accesso: 29.01.2022).

²² Cfr. McCarthy J., *What Is Artificial Intelligence?*, Stanford, 2004 (<https://homes>).

tocampo dell'IA che funziona secondo modelli di apprendimento automatico classici – detti *non-deep* – che richiedono un maggiore intervento umano per segmentare i dati in categorie. Il *deep learning* è a sua volta un sottocampo del *machine learning* che tenta di imitare l'interconnessione del cervello umano utilizzando le reti neurali. Le sue reti neurali artificiali sono costituite da strati di modelli, che identificano i modelli all'interno di un determinato set di dati. Questi sfruttano un volume elevato di dati di addestramento per apprendere in modo accurato, che di conseguenza richiede *hardware* più potente. Per quanto gli algoritmi di *deep learning* siano oggi fortemente associati con la realizzazione di IA di livello umano presentano enormi problemi per quanto riguarda la spiegabilità dei processi decisionali che mettono in pratica.

Per superare questa impasse si teorizza la realizzazione di una intelligenza artificiale forte o *Strong AI*, nota anche come *Intelligenza Artificiale Generale* (GAI). L'idea dei ricercatori è di sviluppare una GAI mediante la quale la macchina potrebbe avere un'intelligenza pari a quella umana, sviluppare l'equivalente di una coscienza ed avere la capacità di risolvere problemi, imparare e a pianificare azioni future. Di fatto la *Strong AI* mira a creare macchine intelligenti indistinguibili dalla mente umana che proprio come un bambino imparerebbero attraverso input ed esperienze progredendo nel tempo. I ricercatori, parte del mondo accademico e il settore privato stanno investendo molto nella creazione della GAI. Tuttavia, allo stato attuale dello sviluppo questa è solo un concetto teorico e nulla di tangibile che si possa avvicinare a questa è ancora stato realizzato. Il dibattito è molto acceso e alcuni citano Minsky ricordando le sue previsioni troppo ottimistiche sullo sviluppo delle AI, altri sono più radicali e arrivano a dire che i sistemi di AI forte non possono nemmeno essere sviluppati. Altri ancora, come Judea Pearl pensano che sia possibile invece realizzare una macchina capace di inferenza causale e quindi possiamo realizzare il primo passo verso la realizzazione di una GAI²³.

di.unimi.it/borghese/Teaching/AdvancedIntelligentSystems/Old/IntelligentSystems_2008_2009/Old/IntelligentSystems_2005_2006/Documents/Symbolic/04_McCarthy_whatissai.pdf accesso: 30.01.2022, la traduzione è mia).

²³ Cfr. Pearl J., *The Book of Why*, New York, 2019.

Questa rivoluzione digitale è stata possibile grazie ad alcuni artefatti tecnologici specifici: gli algoritmi informatici. Questi strumenti sono il motore della digitalizzazione. I computer grazie agli algoritmi di fatto fabbricano la nostra realtà, la organizzano e la orientano producendo convenzioni e sistemi di equivalenza che selezionano e gerarchizzano secondo quadri di valori generando schemi cognitivi e culturali contemporanei²⁴.

Dobbiamo essere consapevoli che parlare di algoritmi non è immediatamente comprensibile e probabilmente il lemma stesso algoritmo è praticamente assente dal linguaggio comune fino alla seconda metà del secolo scorso. È stata l'informatica e la teoria computazionale che ha prodotto con la diffusione dei suoi prodotti nella cultura di massa, anche la relativa diffusione sui *media* del termine algoritmo.

Il termine algoritmo deriva dal latino *algorithmus* o *algorismus*²⁵ e deve il suo nome al momento della storia in cui in Europa si passò dal sistema di numerazione romano a quello indo-arabo. L'adozione in Europa dei numeri indo-arabi avvenne tramite gli scritti di Abū Ja'far Muhammad ibn Mūsā al-Khwārizmī, uno scienziato musulmano della Transoxiana in Persia, tramite due testi. Al-Khwārizmī scrisse nell'825 il trattato *Sul calcolo con i numeri hindi* tradotto in latino nel XII secolo, e *Algoritmi de numero Indorum*, il cui titolo tradotto, dove *Algoritmi* era la translitterazione del nome dell'autore ma fu intesa come il plurale dell'ipotetico termine *algoritmo*, questo portò a coniare questa parola (dal latino *algorithmus*) che venne ad acquisire il significato di *metodo di calcolo*²⁶.

Nella ricostruzione del significato del termine, Picchiarelli ci ricorda che in generale algoritmo indica ogni forma di procedimento che, in forza della sua sistematicità, permetta in un numero finito di passaggi e secondo regole definite e precise, di giungere ad un risultato. Se questo processo si applica al dominio della matematica, allora algoritmo diviene

²⁴ Cfr. Cardon D., *Cosa sognano gli algoritmi. Le nostre vite al tempo dei big data*, Milano, 2016: 7.

²⁵ Cfr. Battaglia S., Ronco G., Sanguineti E., *Algoritmo*, Grande dizionario della lingua italiana, I, Torino, 1970: 303.

²⁶ Cfr. Ifrah G., *The Universal History of Numbers: From Prehistory to the Invention of the Computer*, Hoboken, 1999.

sinonimo di procedimento computazionale – anch'esso di natura processuale – che mediante un numero finito di passaggi giunge ad ottenere l'argomento numerico di una funzione di riferimento. Nell'informatica diviene il modo mediante il quale un computer può ordinare una sequenza finita di operazioni elementari – cioè ogni operazione eseguibile da un processore in un tempo definito non ulteriormente riducibile o scomponibile in altre operazioni – per trasformare un input in un output coerente²⁷.

Bisogna ricordare una differenza fondamentale tra algoritmi informatici e codice di programmazione. I primi sono dei procedimenti logici computazionali per ottenere un risultato mediante una concatenazione di operazioni elementari, mentre il secondo è la sua espressione mediante un linguaggio di programmazione: una lingua formale pensata per realizzare set di istruzioni – input – con cui si producono dati in uscita – output. Fondamentalmente i linguaggi di sviluppo convertono i concetti astratti dell'uomo in codice macchina, ovvero la lingua con cui opera il computer.

I due concetti sono correlati ma a livelli differenti: un algoritmo informatico si può inverare in differenti codici di programmazione. Questi codici possono differire non solo per sintassi differenti – i differenti linguaggi oggi utilizzati – ma anche per il modo con cui il programmatore implementa linguisticamente il processo logico ottimizzandolo – lo stile di programmazione.

Possiamo allora indicare alcuni elementi costitutivi di un algoritmo informatico. Questo è fondamentalmente una sequenza di istruzioni che sono eseguite secondo il verificarsi o meno delle condizioni che vengono generate ad ogni passaggio in modo non ambiguo. La concatenazione dei passaggi impone che non è possibile l'esecuzione dell'istruzione successiva se la precedente non è completa: questa condizione mantiene l'algoritmo informatico nel dominio delle strutture di controllo finite e quindi computabili da ogni macchina di Turing in un tempo finito.

²⁷ Cfr. Istituto della Enciclopedia italiana, *Algoritmo*, Enciclopedia Treccani, I, Roma, 2010: 220-221.

Se l'algoritmo si configura come un artefatto di controllo di natura finita, cioè limitato nel tempo, ossia eseguibile in un tempo di calcolo finito, questo necessariamente assume la forma di una struttura astratta, efficace, effettiva, di natura necessitante ed imperativa. L'algoritmo così compreso non è solo performativo ma anche auto verificante: non solo fornisce determinati risultati a partire da certi parametri iniziali, ma in forza della connessione input output è perfettamente valutabile nella correttezza o nel malfunzionamento della sua implementazione²⁸. Volendo semplificare ulteriormente, possiamo concludere che un algoritmo è una successione finita di regole o istruzioni elementari ordinate secondo uno schema preciso che permette la loro esecuzione passo dopo passo in vista della risoluzione di un certo tipo di problema e che a partire da una serie di dati iniziali produce una serie di dati finali²⁹.

Le istruzioni che compongono un algoritmo possono essere lineari o cicliche. Nei linguaggi di programmazione moderni, inoltre, un algoritmo può richiamare altri algoritmi esterni al codice, mediante le cosiddette librerie, facendo così ricorso ad algoritmi implementati da altri. Il codice moderno è particolarmente complesso non solo per la complessità dei problemi ma perché nessun programmatore di fatto scrive un programma completamente ma lo realizza includendo ed eseguendo algoritmi scritti e distribuiti in pacchetti chiamati librerie. Questa stratificazione del codice genera una certa opacità nel *software*.

Con il passaggio alle IA un ruolo chiave lo rivestono i dati che consentono i processi di apprendimento automatico dell'algoritmo. Per definizione, i modelli di apprendimento automatico vengono formati utilizzando i dati. In effetti, il *set* di dati su cui viene formato un modello di apprendimento automatico influenza fondamentalmente il suo comportamento. Tuttavia, attualmente non esiste un modo standard per documentare come e perché è stato creato un *set* di dati, quali informazioni contiene, per quali attività dovrebbe e non dovrebbe essere utilizzato

²⁸ Cfr. Hill R.K., *What an Algorithm Is*, Philosophy & Technology, 29, 2016: 35-59.

²⁹ Cfr. Ifrah G., *The Universal History of Computing: From the Abacus to the Quantum Computer*, New York, 2001: 277. Si veda anche Brassard G., Bratley P., *Fundamentals of algorithmics*, Englewood Cliff, 1996: 1-6.

e se solleva preoccupazioni etiche o legali. Questa mancanza di documentazione è particolarmente problematica quando i *set* di dati vengono utilizzati per addestrare modelli di apprendimento automatico per applicazioni ad alto rischio.

Oggi l'utilizzo degli algoritmi di IA come strumento di mediazione delle economie digitali e delle *platform* sociali produce, in forza della profilazione della parte umana, inedite forme di *loop* cibernetico. Gli algoritmi di IA progettati per consigliare informazioni e prodotti in linea con le presunte preferenze individuali possono creare *feedback* incontrollati in cui sia le preferenze sulle informazioni dell'utente che la successiva esposizione ai contenuti diventano più estreme nel tempo (le cosiddette bolle). Tali dipendenze dal percorso possono avere effetti trasformativi, modificando le preferenze e i valori degli utenti stessi e portando alla radicalizzazione. Ma il problema è più generale. In breve, abbiamo bisogno di forme di gestione per evitare di scaricare i nostri processi evoluti di ricerca di informazioni su algoritmi che sono in genere progettati per massimizzare la redditività, con incentivi spesso insufficienti a promuovere una società informata, giusta, sana e sostenibile. L'industrializzazione dell'AI sta producendo degli effetti sulle relazioni sociali e sulla relazione uomo macchina che richiede una gestione non solo tecnica ma politico-sociale.

SOSTENIBILITÀ DIGITALE

Se la società dell'informazione può di fatto mediante azioni di *feedback* digitali mettere l'uomo in una condizione di controllo da parte della macchina (sia essa elettronica o algoritmica) e se la relazione cibernetica nella sua forma più radicale di realizzazione di simbiosi uomo-macchina può di fatto negare la necessità di ipotizzare cause finali nell'agire, appare qui un orizzonte distopico in cui la società dell'informazione collassa inevitabilmente in una società del controllo. L'analisi della società digitale ci permette di riflettere sul legame tra cause, necessità e libertà che il digitale realizza nella sua forma di attuazione politica: mette in questione la stessa possibilità che esista un destino dell'uomo che dipenda dal suo volere libero.

Questa forma di digitalizzazione cibernetica, che qui definirei come “forte” per sottolineare come questa sia una possibile forma di società qualora non si mettano in atto forme di sostenibilità digitale³⁰, rischia di elidere la stessa possibilità di una *libertà positiva*. Con questo termine s'intende, ripercorrendo le riflessioni di Bobbio, nel linguaggio politico la situazione in cui un soggetto ha la possibilità di orientare il proprio volere verso uno scopo, di prendere delle decisioni, senza essere determinato dal volere altrui. Questa forma di libertà si chiama anche “autodeterminazione” o, ancor più appropriatamente, “autonomia”. (...) Della libertà positiva la definizione classica fu data da Rousseau, per il quale la libertà nello stato civile consiste nel fatto che quivi l'uomo, in quanto parte del tutto sociale, come membro dell'‘io comune’, non ubbidisce ad altri che a se stesso, cioè è autonomo nel senso preciso della parola, nel senso che dà leggi a se stesso e non ubbidisce ad altre leggi che a quelle che si è dato: «L'obbedienza alla legge che ci siamo prescritti è la libertà»³¹. Tale concetto di libertà fu ripreso, per influsso diretto di Rousseau, da Kant nella *Metafisica dei costumi*, ove la libertà giuridica viene definita come «la facoltà di non obbedire ad altra legge che non sia quella a cui i cittadini hanno dato il loro consenso»³². «(...) Le libertà civili, prototipo delle libertà negative, sono libertà individuali, cioè inerenti all'individuo singolo: storicamente, infatti, sono il prodotto delle lotte per la difesa dell'individuo considerato o come persona morale, e quindi avente un valore di per se stesso, o come soggetto di rapporti economici, contro l'invasione di enti collettivi come la Chiesa e lo Stato (...). La libertà come autodeterminazione invece è generalmente riferita, nella teoria politica, a una volontà collettiva, sia questa volontà quella del popolo o della comunità o della nazione o del gruppo etnico o della patria»³³.

³⁰ Su questo sta prendendo forma una riflessione approfondita in un'opera scritta a quattro mani con Sebastiano Maffettone (cf. https://www.corriere.it/opinioni/21_maggio_17/intelligenza-artificiale-frontiera-principi-697e5326-b71d-11eb-ba17-f6e1f3fff06b.shtml).

³¹ Rousseau J.J., *Contrat social*, I, 8.

³² Kant I., *Metafisica dei costumi*, II, 46.

³³ Bobbio N., *Libertà*, Enciclopedia del Novecento, Treccani, Roma, (https://www.treccani.it/enciclopedia/liberta_%28Enciclopedia-del-Novecento%29/).

Alla luce di queste brevi riflessioni, ci sembra di poter sottolineare che la matrice epistemologica del controllo insito nello sviluppo del digitale come cultura dell'informazione cibernetica abiti ancora in maniera implicita e irriflessa all'interno delle applicazioni tecniche della società dell'informazione. Spetta alla società civile la creazione di un dibattito perché i processi di innovazione tecnologica digitale siano messi in questione. Tuttavia, il mondo della tecnologia è oggi descritto dalla categoria dell'innovazione.

Se continueremo a guardare la tecnologia solamente come innovazione rischiamo di non riuscirne a percepire la portata di trasformazione sociale, e quindi di risultare incapaci di orientarne verso il bene gli effetti.

Per poter parlare di innovazione come di un bene, e per poterla orientare al bene comune, abbiamo bisogno di una qualifica che sia in grado di descrivere come e quali caratteristiche del progresso contribuiscano al bene dei singoli e della società. Per questo, con Sebastiano Maffettone, abbiamo deciso di adottare la categoria della sostenibilità digitale.

L'idea di sostenibilità digitale porta l'attenzione su un concetto di ampia portata, che comprende l'espansione durevole delle possibilità di scelta degli individui e il miglioramento equo delle loro prospettive di benessere. Parlare di sostenibilità digitale significa non mettere la capacità tecnica al centro dell'attenzione, bensì tenere l'uomo al centro della riflessione e come fine che qualifica il progresso.

Utilizzare eticamente la tecnologia digitale oggi, rispettare l'*ecologia umana*, significa cercare di trasformare l'innovazione in un mondo *digitale sostenibile*. Significa indirizzare la tecnologia verso e per lo sviluppo umano, e non semplicemente cercare un progresso fine a sé stesso. Sebbene non sia possibile pensare e realizzare la tecnologia senza delle forme di razionalità specifiche (il pensiero tecnico e scientifico), porre al centro dell'interesse la sostenibilità digitale significa dire che il pensiero tecnico-scientifico non basta.

Perché ci sia libertà, abbiamo bisogno che la coscienza e le coscienze interrogino la tecnica orientando il suo sviluppo verso il bene comune.

LE SCIENZE SOCIALI DI FRONTE ALL'INTELLIGENZA ARTIFICIALE

Carla Collicelli

I dubbi posti dal recente sviluppo tecnologico nel campo della IA, della robotica e dei social media – e che ruotano attorno a concetti quali *human enhancement*, trans-umanesimo e post-umanesimo – hanno quale terreno principale di analisi e discussione le discipline filosofiche, all'interno delle quali da tempo vengono sollevate importanti questioni sia di tipo ontologico che di tipo etico. E dove sono nati filoni di riflessione filosofica nuovi e dedicati, come la filosofia della mente, la bioetica e la neuro-filosofia. Ma anche le discipline raggruppate sotto l'etichetta di scienze sociali avvertono la sfida che lo sviluppo tecnologico e scientifico dell'ultimo periodo, e in special modo quello legato alla IA, pone rispetto a molti aspetti teorici ed applicativi.

La sociologia nasce come disciplina autonoma e come strumento di lettura della realtà governato dalla ragione quando, nel corso del 1700, diventa possibile supportare l'analisi della realtà con procedimenti scientifici e tecniche di misurazione ed inferenza statistica. Da cui lo sviluppo di un'area disciplinare che, pur discendendo dalla filosofia, intende studiare la realtà umana e sociale a partire da una base di ipotesi di lavoro, per fare poi ricorso a strumenti di osservazione e rilevazione empirica, e sfociando quindi nella produzione di indicazioni paradigmatiche di tipo prospettico.

Lo sviluppo delle moderne tecnologie pone a questa area disciplinare una serie di problemi. Innanzitutto le scienze sociali vedono ampliarsi e modificarsi le possibilità di investigazione e ricerca nel proprio specifico campo di osservazione a seguito della introduzione di nuove e più ampie tecnologie di rilevazione, di algoritmi per elaborare i dati di riferimento e di metodi e strumenti inimmaginabili, fino a poco tempo fa, per immagazzinare quantità enormi di informazioni che fanno da base alle successive elaborazioni.

Al tempo stesso, il livello raggiunto, ed in particolare le prospettive di simulazione e sostituzione del pensiero umano, pongono alle scienze sociali il cruciale quesito se sia ipotizzabile una tenuta della società umana, come la conosciamo, nonché dei paradigmi consolidati di analisi della stessa. E se non si vada piuttosto delineando da un lato uno scenario di depotenziamento dell'umano e del sociale rispetto al tecnologico; dall'altro lato un processo di sovvertimento dei principi cardine del patrimonio scientifico e culturale consolidatosi nel tempo nelle discipline sociali.

Infine in molti dei campi applicativi della ricerca sociologica, cui corrispondono le sue diverse sotto discipline – sociologia del lavoro, del *welfare*, della comunicazione, ecc. – sono ormai numerosi gli studi e le analisi che mettono in evidenza forti contraddizioni e segnalano evidenze e dati che mettono in discussione la bontà di alcune delle più recenti applicazioni tecnologiche alla vita umana, ed in particolare di quelle che si pongono in un'ottica di tendenziale sostituzione del pensiero umano.

IL PROBLEMA DEL LIMITE

Dal punto di vista dei principi di riferimento, non è da oggi che la sociologia avanza perplessità rispetto al valore della tecnologia, specie se applicata al campo delle funzioni più evolute della natura umana. E non mancano voci importanti che hanno teorizzato la fine dell'umano, ad esempio nei termini in cui si è espresso Baudrillard¹ con "l'implosione dell'identità e delle differenze" e l'avvento di una era "post-soggettuale".

Volendo tentare di ridurre la complessità della questione ai suoi snodi più importanti, si può dire che sono almeno quattro le aree rispetto alle quali l'evoluzione più recente della tecnica e le sue applicazioni nell'umano pongono i maggiori problemi all'impalcatura della disciplina sociologica. E si tratta di: il principio del limite applicato allo sviluppo umano, sia individuale che collettivo; la natura composita del processo conoscitivo della mente umana; la creatività e le funzioni di progettazione del futuro; le motivazioni ed i valori di riferimento alla base delle scelte di vita e delle attribuzioni di senso all'esistenza.

¹ Farisco M., *Uomo-natura-tecnica, Il modello, il post-umanesimo*, Roma-Teramo, 2008.

Fa parte della storia del pensiero il riconoscimento della pregnanza del concetto di limite, come dimensione incancellabile della vita umana e barriera alle possibilità ed ai desideri. Sappiamo che la modernità nasce, come ci ricorda Remo Bodei², proprio sull'idea del superamento delle "colonne d'Ercole", e cioè dei limiti fisici e mentali dell'umanità. E la cultura occidentale, in particolare, è connotata dal valore attribuito alla evoluzione progressiva verso obiettivi sempre più ambiziosi, quasi all'infinito. L'idea delle potenzialità senza limiti si è poi fortemente consolidata con l'avvento della tecnologia moderna, che ha avanzato proprio l'ipotesi del superamento di ogni limite.

Ma sappiamo anche che la vita di ciascuno, e quella della società umana e dei suoi costrutti, sono condizionate da limiti insormontabili – primo fra tutti quello della morte – che rendono necessario un continuo *trade off* tra aspirazioni e possibilità reali. E la scienza sociale ci segnala che, ogni qual volta l'umanità ha preteso e tentato di superare questi limiti, è andata incontro a situazioni di evidente criticità, se non addirittura a drammi. Esempi ne sono le varie forme di eccesso, di supremazia e di illusione: eccesso nei consumi e nella comunicazione – fino all'infodemia; delirio di onnipotenza rispetto al superamento delle regole collettive della convivenza; forme di supremazia e di oppressione nei confronti dei più deboli dettate da sentimenti di prevaricazione; assunzione di rischi insostenibili a seguito della spinta a superare i confini etici delle sperimentazioni; illusione di poter porre fine in via definitiva alla malattia, alla miseria, alla vecchiaia; abuso di prodotti chimici che permettono l'alterazione degli stati di coscienza e del funzionamento della mente. Ed altri esempi si potrebbero citare.

È stato spesso grazie alle situazioni di crisi che l'umanità ha capito, a proprie spese, che la mancata considerazione dei limiti che condizionano la nostra vita e le nostre realizzazioni sfocia spesso in pesanti negatività e rischi. In particolare, la recente pandemia ci ha costretto ad ammettere che proprio il nostro modello di vita e di sviluppo economico e sociale, basato sull'Antropocene, e cioè sul superamento dei limiti posti dagli equilibri ecologici, con predominio della specie umana sul resto delle forme

² Bodei R., *Limite*, Bologna, 2016.

viventi e fiducia illimitata nei prodotti ed artefatti che derivano dalla sua attività speculativa, è alla base di molti dei mali della nostra epoca.

Le scienze sociali, a tale proposito, mettono in guardia rispetto alla perdita del senso del limite insita nel tentativo di superare i confini del pensiero umano, segnalando che solo una consapevolezza del contesto generale di vita nel quale l'uomo è inserito e dei suoi limiti può salvaguardarci dalla "deriva catastrofica del futuro", che è di fronte a noi – per riprendere i termini utilizzati da Mauro Magatti³.

L'evoluzione tecnologica degli ultimi decenni, con la sua spinta sempre più accelerata al superamento dei confini dell'umano – di cui l'IA è l'emblema – pone radicalmente in discussione il concetto di limite, su cui buona parte delle scienze sociali basano i propri presupposti, e prefigura scenari decisamente problematici e in contraddizione con la sostenibilità complessiva del sistema vivente.

CREATIVITÀ E PROGETTAZIONE DEL FUTURO

Un secondo aspetto attiene al tema della creatività di cui la mente umana è capace, che si collega alle funzioni, tipiche delle forme più evolute del pensiero umano, della progettazione e programmazione del futuro, fino a ricomprendere la sfera della cosiddetta immaginazione creativa ed utopica. Come si diceva, è con l'avvento dell'era moderna che si consolida un approccio al tempo ed all'esistenza umana di tipo progressivo, con valorizzazione della dimensione progettuale e propositiva. In sociologia è soprattutto il contributo della sociologia francese degli ultimi settant'anni che ha alimentato questo approccio, laddove partendo dall'osservazione della realtà si è dedicato un particolare impegno alla riflessione sulle scelte da compiere e sugli scenari anche innovativi da perseguire, dando seguito con ciò alla teoria secondo la quale, come sostiene *in primis* lo storico della scienza Thomas Kuhn (nel suo *La struttura delle rivoluzioni scientifiche*⁴), le scienze sociali, come quelle esatte,

³ Magatti M., *La tecnica tra enhancement e precarietà: apprendere dall'umano alla prova del suo limite*, Seminario UniCattolica 13 settembre 2021.

⁴ Kuhn T., *La struttura delle rivoluzioni scientifiche*, Torino, 1962.

non possono fare a meno di coltivare, oltre alla descrizione della realtà, anche la sua interpretazione, e soprattutto lo studio dei processi di cambiamento e di trasformazione dei paradigmi di riferimento.

Da questo filone della riflessione sociologica deriva il valore delle speculazioni sul futuro, dell'utopia come "oggetto di una aspirazione ideale", come "funzione stimolatrice nei riguardi dell'azione politica" e come frutto di processi complessi di riflessione creativa e propositiva. Una funzione che rischia di essere mortificata, se non annullata, dall'uso della tecnologia come sostituto del pensiero umano. Una tecnologia che per sua natura intrinseca fonda le proprie elaborazioni sugli schemi del passato rispetto ai quali è stata programmata. E ciò si sposa con le tendenze cosiddette "retrotopiche" della società moderna, che si sostanziano nel connubio tra nostalgia del passato e uso di strumenti tecnologicamente avanzati ma privi di capacità immaginifica e creativa.

Come spiegato molto chiaramente da Zygmund Bauman nel suo volume intitolato proprio *Retrotopia*⁵ «un tempo senza memoria (passato) e senza slanci (futuro) diventa liquido, e poi evapora, incastrato nell'affanno dell'attimo breve, brevissimo (presente). Sfumano le radici, solide ancora durante la navigazione della vita, e si appannano le aspettative, i sogni, le energie che fanno davvero crescere, non solo in senso anagrafico. Nella prigione del tempo snaturato, a un'unica dimensione, la civiltà occidentale deve misurarsi con un nuovo assioma: tutto è presente, esclusivamente presente». *Retrotopia* è per Bauman dunque «una visione situata nel passato perduto/rubato/abbandonato ma non ancora morto», che ci ricorda che già Marx «svelava la logica psicosociale che spinge a cercare nel passato un aiuto nell'aprire le porte del presente». E questa tendenza si sposa con la caratteristica delle tecnologie moderne di costruire i propri assiomi sui dati del passato.

A fronte di questa deriva retrotopica, il movimento di pensiero legato agli obiettivi dello sviluppo sostenibile ed ai principi dell'Enciclica *Laudato Si'* rappresenta un'alternativa importante, che pone in guardia rispetto al venir meno della capacità di progettazione innovativa del fu-

⁵ Bauman Z., *Retrotopia*, Roma-Bari, 2017.

turo. Come si legge nel testo di Enrico Giovannini *L'utopia sostenibile*⁶, è a partire dalla responsabilità nei confronti del futuro e delle future generazioni che occorre rivalutare la creatività ed il pensiero utopico – nel senso originario del termine di riflessione sul modello ideale di società. E sta all'umanità ed alle istituzioni politiche e sociali trovare le soluzioni più corrette per contrastare la deriva presentista, distopica e retrotopica che caratterizza lo sviluppo degli ultimi decenni. Monito che tocca direttamente molti degli aspetti che connotano la produzione scientifica nel campo della ricerca sull'IA tutta centrata sulla riproduzione, e nel migliore dei casi sulla efficientizzazione del presente e del passato, con scarse possibilità di agire sul miglioramento e sulla costruzione del futuro.

LA DIMENSIONE EMOTIVA DELLA CONOSCENZA

Una terza questione di non poca importanza dal punto di vista sociologico si pone rispetto alla natura intrinseca del pensiero umano nelle sue funzioni conoscitive, di decisione e di giudizio. In altre parole, rispetto alla natura di quel pensiero che l'IA intende imitare e per molti versi superare. Sulla base di una letteratura ormai sufficientemente ampia, che affonda le radici oltre che nelle discipline sociologiche e psicologiche, anche in quelle biomediche ed in particolare nelle cosiddette neuro-scienze è acquisito che il pensiero, e di conseguenza l'intelligenza, sono realtà complesse e articolate, all'interno delle quali convivono funzioni ed elementi sia di tipo razionale, sia di tipo irrazionale ed afferenti alla sfera della emotività e dei sentimenti.

Le emozioni sono una componente importante, spesso determinante, nella formazione del giudizio umano e nel processo di soluzione dei problemi, e non si tratta solo di un vago coinvolgimento emotivo, ma di una modalità profondamente umana di anticipazione del pensiero razionale da parte della sfera emotiva. Una forma essenziale di legame tra corpo e mente, come è stato detto, definito anche come “colore dell'intelligenza”⁷. In sostanza, una prima forma di conoscenza che prelude –

⁶ Giovannini E., *L'utopia sostenibile*, Roma-Bari, 2018.

⁷ Nussbaum M. C., *L'intelligenza delle emozioni*, Bologna, 2009.

come dice Marinoff⁸ – all'analisi razionale, al discernimento e alle scelte. Ed anche i sentimenti – la parte relazionale delle emozioni – entrano a pieno titolo nel processo conoscitivo al pari delle emozioni, ed anzi vi si sovrappongono e si combinano tra loro, e spesso proprio da questa combinazione e da questo intreccio tra emozionale, relazionale e razionale sgorgano le idee e gli stimoli più interessanti.

Una complessità questa che è evidentemente non riproducibile attraverso strumentazioni tecnologiche e procedure altamente standardizzate. Da cui le preoccupazioni di fronte al tentativo di proporre processi conoscitivi e di pensiero (scelta, decisione, ecc.) che escludono questa complessità ed in particolare non sono in grado di comprenderne la dimensione emotiva.

VALORI E IDEALI ALLA BASE DEL PENSIERO E DELLE SCELTE

Strettamente collegato ai punti precedenti, rilevante è il ruolo dei valori di riferimento che stanno alla base delle scelte che la persona e l'umanità, nelle loro varie espressioni, compiono. Tema questo ampiamente trattato in altri contributi di questo volume (cfr. Zamagni, Becchetti, Palazzani e altri). Aggiungiamo qui che la responsabilità etica e la considerazione dei valori di riferimento per le scelte da compiere costituiscono elementi fondamentali del nostro assetto societario, che non possono essere riprodotti in processi decisionali meccanici, che difettano di una dimensione etica di riferimento, come ben spiegato da Hans Jonas⁹.

«L'uomo è misura di tutte le cose è un'affermazione che può avere ancora senso così intesa: l'uomo, in virtù delle strutture cognitive di cui si è dotato nel corso della propria evoluzione biologica, si pone come il solo osservatore in grado di conferire senso a quell'insieme di fenomeni che le suddette strutture cognitive rendono visibili». Queste le parole utilizzate da Giovanni Leghissa¹⁰.

⁸ Marinoff L., *Platone è meglio del Prozac*, Casale Monferrato, 1999.

⁹ Jonas H., *Il principio responsabilità*, Torino, 2002.

¹⁰ Leghissa G., *Postumani per scelta. Verso un'ecosofia dei collettivi*, Milano-Udine, 2015.

Di rilievo a questo proposito anche le riflessioni, che ben si sposano con i cenni di cui sopra, sulle strumentalizzazioni che la tecnica e le sue applicazioni rischiano di produrre nel sociale, trasformando ciò che dovrebbe costituire lo scopo (l'umano, lo sviluppo, il bene) in un mezzo¹¹.

La libertà di scelta e di comportamento degli individui e delle collettività ed i processi di definizione del bene e del male sfuggono in sostanza in maniera evidente al controllo da parte delle tecnologie, e suggeriscono particolare cautela nell'immaginare una delega ampia delle funzioni decisionali, a tutti i livelli, dalla persona alla macchina.

COMUNITÀ E FIDUCIA

Ma la sociologia e le scienze sociali non si occupano solo di concetti astratti e di principi di riferimento della convivenza umana. Il loro ambito disciplinare comporta, ed in una misura sicuramente preponderante, lo studio delle forme concrete di convivenza, di relazione e di comunicazione. Anche rispetto a questi settori disciplinari, di tipo applicativo, si pongono innumerevoli interrogativi relativamente allo sviluppo della cosiddetta intelligenza artificiale.

Va innanzitutto rimarcato che molte delle tecnologie innovative, che popolano la nostra realtà sociale, sono di fatto derivati del pensiero umano e si pongono come strumenti aggiuntivi importanti ed utili per la realizzazione di opere e progetti pensati e gestiti dall'uomo, dunque come protesi dell'umano piuttosto che come negazione o superamento dell'umano. I casi più evidenti da questo punto di vista sono quelli delle tecnologie dell'informazione e della comunicazione e quelli della medicina robotica. Dall'introduzione delle tecnologie più avanzate in questi settori sono derivate e derivano peraltro potenzialità altamente positive, ma con un posizionamento di chiara subalternità della tecnologia rispetto all'umano, che mal si coniuga con l'aspirazione a superare, e quali a scalzare, l'umano con l'IA.

Va però detto che in molti degli ambiti applicativi della ricerca sociale emergono sempre più frequentemente evidenze che segnalano

¹¹ Marassi M., *Tecnica e valori dell'umano*, Seminario UniCattolica 13 settembre 2021.

contraccolpi negativi e rischi sociali che derivano dall'uso massiccio, e in molti casi strategicamente poco controllato, delle tecnologie.

Un primo aspetto riguarda la dimensione relazionale e comunitaria della convivenza umana, la cui importanza viene continuamente ribadita dalle scienze sociali come modalità essenziale alla sopravvivenza ed al benessere delle persone e dei gruppi. La necessaria condivisione di spazi e tempi di vita, lo scambio delle idee e delle esperienze, la regolazione comunitaria del rapporto con le istituzioni, le forme di relazionalità positiva, la solidarietà e la cooperazione sono tutti aspetti essenziali per l'esistenza stessa di una società, la cui tutela e sviluppo rientrano a pieno titolo tra gli obiettivi più importanti delle politiche sociali, economiche ed ambientali, in contrapposizione ai rischi della solitudine, dell'isolamento, dello spaesamento, dell'individualismo e dell'egoismo sociale, alla base di tante patologie sociali e psicologiche e di tante situazioni di criticità.

Tutte le ricerche sociologiche più recenti confermano che il benessere e la vera felicità degli individui dipendono, assai più che dalla ricchezza materiale e dal raggiungimento di obiettivi quantitativi, dalla qualità delle relazioni sociali primarie, dalla fiducia e dal senso di utilità reciproca, tipici delle comunità coese e dei sistemi sociali connotati da una intensa relazionalità. È quanto ribadito dai sostenitori di una necessaria rigenerazione sociale attraverso il coinvolgimento dei soggetti vitali. Ed è quanto sfugge ai processi tecnologici che si dipanano in gran parte in maniera asettica rispetto al contesto comunitario e relazionale.

IL LAVORO

Un ulteriore ambito applicativo della ricerca sociale che vale la pena citare è quello del lavoro, dove le tecnologie hanno creato grandi opportunità ma anche grandi sconquassi. I vantaggi della digitalizzazione di molti processi produttivi sono evidenti, specie per ciò che riguarda l'efficienza, la qualità ed il risparmio di fatica umana. Ma altrettanto evidenti sono i relativi rischi e le criticità che si sono mano a mano determinate e continuano a determinarsi a seguito dell'intervento tecnologico: dalla perdita di posti di lavoro, alla accelerazione dei ritmi di produzio-

ne, alla mortificazione delle funzioni di interscambio e confronto, alla burocratizzazione di molte funzioni del terziario, alla delocalizzazione industriale, alla crescita del lavoro precario, alla perdita del potere di acquisto ed al cosiddetto *digital divide*. Fino ad arrivare alle problematiche più recenti legate all'introduzione dello *smart working* (meglio sarebbe chiamarlo lavoro a distanza), ed alla sostituzione degli incontri di lavoro in presenza con incontri a distanza *online*.

La ricerca sociale in questo campo ci dice che, laddove la tecnologia interviene a sostituzione parziale o totale della fatica fisica ed a supporto delle funzioni umane, il bilancio è sicuramente positivo. Ma, laddove comporta uno schiacciamento della dimensione umana e relazionale del lavoro, la crescita dei livelli di sfruttamento dei lavoratori e la strumentalizzazione dei fattori produttivi a fini esclusivamente quantitativi e non qualitativi, si verificano crisi e contraccolpi negativi. Soprattutto la ricerca sociale mette in guardia dalle situazioni nelle quali si pretende di sostituire la decisione, il ragionamento, la progettazione ed il controllo, svolti da soggetti umani, con tecnologie standardizzate che non tengono conto della dignità umana e della funzione positiva dell'autorealizzazione lavorativa.

EDUCAZIONE E FORMAZIONE SCOLASTICA

Analoghi risultati si rilevano nel campo dell'educazione e della scuola. Anche in questo caso non vanno disconosciuti gli effetti positivi legati alla introduzione di strumentazioni tecnologiche nella didattica scolastica, né tanto meno quella di supporti tecnologici per il sostegno educativo, familiare e scolastico, per le persone con disabilità e per i bambini ed i giovani in difficoltà.

Certo è, però, che nulla di tutto ciò può sostituire la relazione educativa che si instaura tra genitori e figli o tra insegnanti e allievi, tanto meno una IA o un robot. Come molti studi di ambito pedagogico e psicologico hanno dimostrato¹², di fronte all'introduzione di tecnologie roboti-

¹² Si veda Riva G., *Human Robot confluence. Towards a multimodel analysis of human robot interaction*, Seminario UniCattolica 13 settembre 2021.

che nell'educazione e nella scuola, i bambini riconoscono la differenza tra persona e robot, e individuano un robot come qualcosa di "non umano". E spesso, se sottoposti a sperimentazioni che prevedono l'uso massiccio di "macchine per insegnare", reclamano la presenza di persone vive, che appaiono quindi essenziali per la formazione della propria identità.

MEDICINA E SANITÀ

I vantaggi ed i successi ottenuti nel campo della medicina grazie all'uso delle moderne tecnologie sono fuori di dubbio ed anche su questo il libro contiene contributi di approfondimento specifico. Si può anzi affermare senza tema di essere contraddetti che l'ambito medico è probabilmente quello di maggiore impatto positivo della tecnica ed anche delle forme di IA. Ma la ricerca sociale applicata a questo campo ci ricorda che questi aspetti positivi non hanno annullato fino ad oggi, e non possono sostituire in futuro, il ruolo della relazione terapeutica che si instaura tra paziente e terapeuta, medico o altra figura sanitaria che sia.

La tempesta sociale, sanitaria ed economica che ha investito l'Italia e il mondo con il Covid-19 ha cambiato la prospettiva epidemiologica di riferimento, ma ha al tempo stesso rafforzato, invece che sminuito, l'importanza del rapporto medico-paziente, della medicina del territorio e dell'integrazione socio-sanitaria, di quelle branche della medicina cioè che guardano alla comunità di vita ed al rapporto diretto con le persone, sia in termini preventivi che in termini di cura, come fondamentali per il buon esito delle cure. Come hanno scritto i medici dell'Ospedale Papa Giovanni di Bergamo il 21 marzo del 2020¹³, abbiamo capito fino in fondo quanto sia importante la dimensione della comunità in sanità, e non solo per le patologie croniche e le disabilità, ma anche di fronte ad una crisi pandemica di tipo infettivo, che è anche e soprattutto una crisi umanitaria, che tocca tutta la popolazione e richiede un approccio comunitario.

¹³ Nacoti M. et al., *At the Epicenter of the Covid-19 Pandemic and Humanitarian Crises in Italy. Changing Perspectives on Preparation and Mitigation*, NEJM Catalyst, Innovation in Care Delivery, March 21, 2020:1-5.

Numerosi studi e ricerche evidenziano, inoltre, il valore di una medicina e di processi di cura che attraverso un rapporto umano profondo, non simulabile con nessuna tecnologia, costruisca percorsi personalizzati, anche con il supporto di elementi tratti dal mondo delle emozioni e dei sentimenti e da quello della produzione artistica, e con l'eventuale supporto di tecnologie, ma sotto la guida consapevole del terapeuta che mette a frutto le sue risorse intellettuali ricche di aspetti emotivi e relazionali, oltre che razionali¹⁴.

Molti altri esempi si potrebbero portare, tratti da altri ambiti della ricerca sociale applicata, dalla comunicazione, al governo della cosa pubblica e delle politiche. I risultati che potremmo trarne non sarebbero diversi, tutti esplicitamente orientati verso la convinzione della non replicabilità della complessità del pensiero e della azione umana attraverso le tecnologie.

SPUNTI DI PROSPETTIVA

È evidente, sulla base degli esempi e delle considerazioni generali di cui sopra, che la società ha trovato e può trovare nella strumentazione tecnologica un supporto, a volte molto utile, alla propria sussistenza e sostenibilità. Ma è altrettanto evidente che nulla ancora nel panorama dei prodotti dello sviluppo tecnologico sembra preludere alla individuazione di forme soddisfacenti di riproduzione dei processi umani e sociali più complessi, di sostituzione delle forme di aggregazione e relazione tra individui, dei rapporti terapeutici ed educativi, delle attività lavorative. Il processo particolarmente avanzato e veloce cui assistiamo di ampliamento della realtà naturale, di quella culturale e dello stesso pensiero umano nel mondo attraverso forme di IA, non intacca pertanto, almeno fino ad oggi e secondo le evidenze prodotte dalla ricerca sociologica, le prerogative specifiche dell'uomo e del suo ruolo nella storia e nell'universo, pur ponendo sfide nuove di sostenibilità, controllo etico e umanizzazione.

¹⁴ Si veda Valentini V., *Rhapsody: realizzazione di servizi automatizzati ed avatar per lo sviluppo di una medicina personalizzata basata sul real world*, Seminario UniCattolica 13 settembre 2021.

Piuttosto che sui rischi di un annullamento dell'umano, o di una sua radicale trasformazione in altro da sé, la sociologia sembra richiamare quindi l'attenzione sull'importanza degli strumenti di governo, controllo e orientamento dello sviluppo tecnologico e delle sue produzioni. Ed il riferimento va in particolare alla rappresentanza dei bisogni e degli interessi di individui e collettività di fronte alle nuove tecnologie, alla riflessione etica rispetto ai confini da porre, agli orientamenti e indirizzi in termini di finalità, all'armonizzazione dei percorsi di ricerca tecnologica con i principi della dignità umana e della giustizia allocativa.

La "*Rome Call for AI Ethics*", che si è svolta il 28 febbraio 2020 per iniziativa della Pontificia Accademia della Vita e di Monsignor Vincenzo Paglia, ha promosso un dibattito approfondito su questi temi, con il titolo "*The Good Algorithm? Artificial Intelligence: Ethics, Law, Health*", ed ha prodotto un documento in cui si afferma che «occorre identificare e presidiare alcuni principi-cardine che devono governare per il futuro le ricerche e gli ambienti o prodotti dell'intelligenza artificiale»¹⁵. In particolare, il documento richiama la necessità di difendere «trasparenza, inclusione, responsabilità, imparzialità, tracciabilità, sicurezza e privacy», sette principi guida per un accordo tra sostenitori dell'etica della responsabilità e attori del mondo della IA¹⁶. E ribadisce che «i sistemi di intelligenza artificiale devono essere concepiti, progettati e migliorati per servire e proteggere gli esseri umani e l'ambiente in cui viviamo». Riprendendo il concetto di "algor-etica", come etica dei problemi e dei risvolti etici connessi all'applicazione degli algoritmi, da promuovere nei diversi campi applicativi, dalla politica industriale all'educazione.

D'altra parte, la stessa Pontificia Accademia per la Vita aveva affrontato questioni analoghe nell'incontro del 24 febbraio 2019, con il titolo "*Robo-ethics*", arrivando alle medesime conclusioni. Per riprendere le parole di Monsignor Paglia: «Ci troviamo di fronte a un panorama nuovo: una sorta di invasione della tecnica che aiuta certamente l'umanità a

¹⁵ Alessandrini G., *La Rome Call for AI Ethics: un approccio di svolta verso la fondazione dell'Algor-etica*, 2021.

¹⁶ Alla giornata partecipavano tra gli altri Fao, IBM, Microsoft e Ministero dell'Innovazione.

risolvere problemi oggettivi, come l'autosufficienza di alcuni anziani o di chi deve fare i conti con gravi handicap: ma il rischio è che possa essere la macchina a guidare l'uomo, non viceversa. (...) Senza una risposta etica il futuro può diventare denso di insidie legate alla possibilità per l'uomo di avere un'occupazione».

L'esigenza di natura etica e scientifica di controllare lo sviluppo tecnologico specie nel campo dell'IA è dunque una delle grandi sfide del nostro tempo. Ma altrettanto rilevante è l'esigenza di indirizzare le politiche pubbliche dello sviluppo tecnologico e sociale in una direzione eticamente e socialmente sostenibile a lungo termine, e sia dal punto di vista ambientale, che economico e sociale. Una attenzione particolare va di conseguenza dedicata alle nuove linee di programmazione nazionale in merito all'innovazione ed allo sviluppo tecnologico.

Ad esempio, il Piano Nazionale innovazione 2025, varato a novembre 2020 dal Ministro per l'innovazione tecnologica prende le mosse da una dichiarazione di principio che dice: «L'innovazione e la digitalizzazione devono far parte di una riforma strutturale dello Stato che promuova più democrazia, uguaglianza, etica, giustizia e inclusione e generi una crescita sostenibile nel rispetto dell'essere umano e del nostro pianeta»¹⁷. Un approccio quindi, quello proposto, rispetto alle sfide del digitale e della IA, che pone al centro dei criteri da adottare lo sviluppo inclusivo, etico e sostenibile, e propone venti azioni molto concrete (sull'identità digitale, l'innovazione nella PA, le ICT nelle imprese, l'assistenza agli anziani e molto altro ancora), tutte basate sul criterio della intersezione tra innovazione, sostenibilità ed etica.

Il PNRR (Piano Nazionale di Ripresa e Resilienza), a sua volta, prevede un massiccio investimento per i prossimi tre anni sull'innovazione digitale (49,26 miliardi), volto a promuovere la transizione digitale nella PA, nelle infrastrutture di comunicazione e nel sistema produttivo, migliorare la competitività delle filiere industriali e rilanciare due settori che distinguono l'Italia: il turismo e la cultura. I principali campi applicativi nei quali si attuerà il Piano per la parte di innovazione digitale sono:

¹⁷ Ministero per l'innovazione tecnologica, Piano Nazionale innovazione 2025, 9 novembre 2020.

la scuola (programmi didattici, competenze di docenti e studenti, funzioni amministrative); la sanità (infrastrutture ospedaliere, dispositivi medici, competenze e aggiornamento del personale); l'agricoltura (aggiornamento tecnologico); i processi industriali e nel settore terziario; la pubblica amministrazione (dotazioni tecnologiche, capitale umano e infrastrutturale, organizzazione e modalità di erogazione dei servizi ai cittadini).

Vista l'ampiezza del campo applicativo e l'ammontare delle risorse disponibili, è assolutamente necessario che le scelte ed i concreti progetti di attuazione del PNRR tengano conto dei principi contenuti nel Piano Nazionale innovazione 2025 e delle valutazioni critiche espresse da buona parte del mondo scientifico e culturale sulla necessaria armonizzazione tra tecnologie e principi etici di riferimento della società.

Di conseguenza, anche le altre sei Missioni del Piano in linea con i sei Pilastri del NGEU (*Next generation EU*), dedicate a rivoluzione verde e transizione ecologica; infrastrutture per una mobilità sostenibile; istruzione e ricerca; inclusione, coesione e salute, debbono essere trattate ed attuate secondo una strategia comune che tenga conto dei risvolti etici richiamati.

CONTRO L'INFOSFERA. SALVARE LA PRESENZA

Eugenio Mazzarella

1. IL PECCATO DI FACEBOOK

Se “il peccato di *Facebook*” fosse solo quello documentato dai *Facebook Papers*¹ di aver messo in pericolo la democrazia americana il 6 gennaio 2021 con l’assalto al Congresso, sobillato e organizzato sul social web, di alcune migliaia di *followers* di Donald Trump, durante la proclamazione del risultato elettorale che avrebbe sancito la vittoria di Joe Biden alle elezioni presidenziali, la situazione sarebbe grave, gravissima, ma non ai limiti del sostenibile nel complessivo “shock antropologico”, sospinto da tecnica e globalizzazione, della “modernità del rischio” in cui viviamo (per usare espressioni di Ulrich Beck) ai tempi, già molto avanzati, del digitale e della IA.

Se nell’operato di *Facebook*, connivente con la tossicità sociale degli algoritmi di propagazione delle interazioni sulla sua piattaforma, ci fosse solo un peccato di lesa democrazia, si potrebbe rispondere, ancorché difficile visto il “potere” di Zuckerberg e dei padroni del web, con “algoritmi di polizia”, da implementare obbligatoriamente sui social pena di chiuderne le piattaforme, in funzione di custodi dell’ordine sociale richiesto anche sul *social web*, che non ha diritto di franchigia di “parola”, di interazione sociale quale che sia, rispetto alla società *offline*, alla società reale.

Purtroppo, le cose stanno peggio. E lo si può vedere con l’annuncio che da *Facebook* nascerà *Metaverso*. Il rilancio, non solo del *marketing*, dopo il disastro del sei gennaio per l’immagine dell’azienda, ma del “ver-

¹ *The Facebook Papers ‘History will not judge us kindly’*, The Atlantic Monthly, 25 ottobre 2021; lo si può leggere in italiano su Il Foglio del 1° novembre 2021.

bo” con cui il giovane Zuckerberg aveva definito la “missione” di *Facebook*: la creazione di «un’infrastruttura sociale per dare alle persone il potere di costruire una comunità globale che funzioni per tutti noi». Una *mission* che diventa ben presto, per la comunicazione pubblica dell’azienda, il “verbo” di *Facebook*: con una prostituzione delle parole – “comunità”, “per noi” – che da sola meriterebbe uno studio sulla “retorica” del digitale. La *Community*, la “comunità per tutti noi”: l’*aperti Sesamo* della caverna in cui inghiottire, millantando il ritrovamento della comunità perduta, la solitudine di massa, lo spaesamento dei miliardi di sradicati – dalle comunità locali, dalle culture tradizionali – della globalizzazione. Questa *community* in cui tutti hanno potuto entrare con un *click*, l’annuncio della comunità a venire sul web a chi l’aveva persa, a fronte della nuova creatura di Zuckerberg, è poco più di un proto-evangelo rispetto al nuovo vangelo che si annuncia con *Metaverso*.

E qual è questo vangelo? Il trascendere oggi a disposizione, grazie al digitale e all’IA, del mondo reale nel mondo virtuale per il tramite della *transitività* tra i due mondi. Il nostro essere entrati nell’epoca dell’*onlife*², dove la dimensione vitale, relazionale, sociale e comunicativa, lavorativa ed economica, è vista, agita e proposta come frutto di una continua interazione tra la realtà materiale e analogica e la realtà virtuale e interattiva. Dove l’effetto gorgo, il buco nero dell’*online* si mangia sempre più la realtà *offline*, la vita come tale. Almeno fin qui come tale. Un peccato di “lesa vita” che è il vero *shock antropologico* in cui siamo immersi, e di cui abbiamo segni consistenti, ma forse non piena consapevolezza.

Uno *shock* la cui sostanza è una *ri-ontologizzazione digitale*, agita dalle ICT (*Information and Communication Technologies*) e dalla AI (*Artificial Intelligence*), della realtà³, trasformata in «*infosfera*, neologismo

² Il neologismo si deve a Luciano Floridi. Per la sua contestualizzazione teorica: Floridi L., *La quarta rivoluzione. Come l’infosfera sta trasformando il mondo*, Milano, 2017 [ed. inglese 2014]. Per il suo orizzonte programmatico e di orientamento delle *policies* pubbliche, L. Floridi (a cura di), *The Onlife Manifesto. Being Human in a Hyperconnected Era*, Cham, 2015.

³ Floridi L., *Agere sine intelligere. L’intelligenza artificiale come nuova forma di agire e i suoi problemi etici*, in Floridi L., Cabitza F., *Intelligenza artificiale. L’uso delle nuove macchine*, Martini Lecture, Milano, 2021: 138, 149, 176.

coniato negli anni '70 e basato sul termine biosfera, che fa riferimento a quella limitata porzione del nostro pianeta caratterizzata dalla vita». Concetto, *infosfera*, «in rapida evoluzione, a un livello minimo [indicante] l'intero ambiente informazionale costruito da tutti gli enti informazionali, le loro proprietà, interazioni, processi, reciproche relazioni, un ambiente paragonabile ma al tempo stesso differente dal cyberspazio, che è soltanto una sua regione dal momento che l'infosfera include anche gli spazi di informazione *offline* e analogici, [per proporsi al suo livello massimo come] un concetto che può essere utilizzato anche come sinonimo di realtà, laddove interpretiamo quest'ultima in termini informazionali: in tal caso l'idea è che ciò che è reale è informazionale e ciò che è informazionale è reale»⁴.

Come è evidente, una parossistica e vieta riedizione dell'hegeliana razionalità del reale non come (*rectius*) comprensione delle sue ragioni in essere, divenute e divenienti, anche per guadagnarvi un margine critico di correzione dal magistero della sua storia, della storia del "reale", ma piuttosto nel senso della razionalizzazione ideologica dell'*agere* che nella realtà si dà e si è dato, anche come promozione della sua ineludibilità; in un'idea non lontana dall'orientamento direttivo dell'*antiquarizzazione* dell'umano in essere nell'epoca della tecnica denunciata da Anders, in una ripresa della "legge di Gabor", che "ciò che si può fare, si deve fare, e comunque si farà"⁵. Un approccio alle cui spalle c'è, come vedremo, una ben debole "etica della responsabilità", accorciata e acconciata all'efficienza presunta di proceduralità di controllo, incapace di dirsi di no, di dire "no" allo stesso *agere* possibile; in definitiva *incapace di dire no al presente*, "*anche*" no al presente, per scegliere *quale* presente far diventare futuro.

⁴ Floridi L., *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, cit., p. 44-45.

⁵ È al premio Nobel Dennis Gabor, che si deve l'individuazione di questa criticità dello sviluppo tecnico, passata poi ad essere designata come "legge di Gabor": «È la sua stessa velocità che fa avanzare la tecnica e questo per due motivi: il primo è che è necessario mantenere le industrie tradizionali. La seconda non è altro che la legge fondamentale della società tecnica: "ciò che si può fare tecnicamente sarà necessariamente". È così che il progresso applica nuove tecniche e crea nuove industrie *senza considerare se siano desiderabili o meno*». "Legge" che è anche al centro delle riflessioni [1977] di Jacques Ellul, *Il sistema tecnico. La gabbia delle società contemporanee*, tr. it. di Carbonelli G., Milano, 2004.

Ma così, in nome delle “magnifiche sorti e progressive” della realtà virtuale, della realtà “aumentata” (e chi non vorrebbe aumentare quello che ha o quello che può! Forse solo i saggi...) gestita dagli algoritmi dell’intelligenza artificiale, noi stiamo sradicando la nostra vita, il nostro *esserci*, dall’*essere-nel-mondo di presenza* fin qui abitato, promettendo un ampliamento degli spazi “vitali” accessibili all’esperienza individuale.

Ancora una volta “l’individuo e i suoi diritti”, senza nessun dovere neanche verso sé stesso, che è il mantra sempre più nichilistico della razionalizzazione strumentale della modernità occidentale. Della modernità che noi, l’Occidente, abbiamo dato al mondo globale, che per altro, per una legge del contrappasso, su questo terreno ci sta dando ampie lezioni di più “efficiente” interpretazione. Ed è in questa direzione che, dopo l’infortunio di percorso del sei gennaio, si sta muovendo Zuckerberg con *Metaverso*.

Ma di che si tratta? Lasciamolo dire a un ingegnere informatico:

Facebook diventa *Meta*. Che significa veramente? E cosa dobbiamo aspettarci? L’idea, annunciata da Mark Zuckerberg alla conferenza annuale Connect (28 ottobre 2021), è quella di un ecosistema tecnologico che da semplice social network, il *Facebook* di oggi, si trasforma in un ambiente dove le persone possono vivere e lavorare in un complesso mondo di realtà virtuale connessa: non più e non solo foto e video da guardare, bensì mondi virtuali nei quali immergersi. Mondi non più solo osservati attraverso un monitor, piuttosto confini da valicare grazie a visori per realtà virtuale. Confini che una volta valicati porteranno l’utente, attraverso il suo *avatar*, a vivere la propria vita virtuale nella stessa maniera della vita reale: la fusione di *offline* e *online*, la realizzazione dell’*onlife*. Attraverso il nostro *avatar* faremo acquisti, andremo al teatro o a cena da amici, andremo al lavoro e in vacanza. Il tutto comodamente seduti a casa nostra e semplicemente indossando dei visori per realtà virtuale. (...) Zuckerberg a Connect ha parlato della possibilità di “teletrasportarsi da un’esperienza all’altra” e del fatto che “siamo all’inizio del prossimo capitolo di Internet e del prossimo capitolo della nostra società”. I benefici di un *Metaverso* sono sicuramente tanti. Basti pensare a tutti gli ambiti professionali, sociali e di intrattenimento nei quali un’esperienza immersiva rappresenta un enorme fattore abilitante: immaginiamo solamente i vantaggi per le persone con disabilità e alla possibilità da parte loro di poter avere una vita piena seppur “onlife”. Enormi benefici e potenzialità, accompagnati però da altrettanti

rischi. Non stiamo parlando infatti più di un prodotto commerciale bensì di quello che potrebbe diventare il “sistema operativo” delle nostre vite e della nostra società⁶.

Segnatevi la carota del sostegno esistenziale che da questa vita *onlife* verrebbe alla disabilità [che sta diventando per la crescita dell’aspettativa «la questione sanitaria più rilevante per l’umanità»⁷] di una vita reale, per ragionare poi sui pericoli. Un po’ come donne e bambini legati alle torrette dei carri armati per avanzare in un territorio da conquistare o da sottomettere: la vita fin qui (riduttivamente dal punto di vista dei conquistatori⁸) reale. Segnatevelo, perché è tipico del *marketing* dell’e-vangelo del *social web* che ci promette di farci vivere in modo aumentato

⁶ Antonio Pescapè, *Il Mattino* 31 ottobre 2021.

⁷ Jacques Ellul, *Il sistema tecnico*, cit.: 1.

⁸ Nell’ingenua convinzione (basti pensare al rapporto dell’esperienza comune con il bugiardino di un farmaco o con le clausole in calce di un contratto-tipo di fornitura di servizi) che il *bombing* informativo di un’infosfera priva di frizione dove assistiamo «a una sostanziale erosione del diritto di ignorare» e però «a una crescita esponenziale della conoscenza comune», possa abilitarci a una maggiore responsabilità: «Le ICT stanno rendendo l’umanità sempre più responsabile dal punto di vista morale (...) poiché le interfacce divengono progressivamente meno visibili diviene più sfumato il limite tra il qui (analogico, basato sul carbonio e *offline*) e il là (digitale, basato sul silicio e *online*), per quanto ciò accada tanto a vantaggio del là quanto del qui. Per parafrasare Orazio “l’infosfera conquistata sta conquistando il suo vincitore”. Il mondo digitale *online* trabocca nel mondo analogico *offline*, con il quale si sta mescolando. Si è soliti fare riferimento a tale recente fenomeno con una varietà di espressioni: dalla “computazione ubiquitaria” all’“ambiente intelligente”, dall’“Internet delle cose” alla “realtà delle cose aumentare dal web”. Per quanto mi riguarda, preferisco parlare di *esperienza onlife*. È, o presto sarà, il prossimo stadio nello sviluppo dell’età dell’informazione. In misura crescente conduciamo le nostre vite *onlife*. La graduale informatizzazione degli artefatti e dell’intero ambiente (sociale) fa sì che sia diventato difficile rappresentarsi come fosse la vita in epoca predigitale. Nel prossimo futuro la distinzione però *online/offline* è destinata a diventare ancora più sfumata e quindi a scomparire» (Floridi L., *La quarta rivoluzione*, cit.: 47-48). Questa volta però, la Grecia conquistata che conquistò il vincitore (Orazio, *Epistole*) non è un passaggio nella storia (l’arrivo “umanistico” delle arti nel Lazio agreste del *fiero* vincitore, salutato da Orazio), ma un passaggio dalla storia all’*iperstoria*, per restare al lessico della “quarta rivoluzione”, quella informatica, annunciata da Floridi nella sua cronologia della civilizzazione umana: preistoria, storia, *iperstoria*. Un passaggio antropologico nella storia, non una sua scansione culturale.

e accessibile a tutti (e quindi “democratico”) l’esperienza. Immaginatevi questo nostro “evolvere” in avatar. Questo traslare la propria esperienza nel digitale tramite avatar, ritenendola ancora la “propria” esperienza. Un’immaginazione di futuro che dimentica una piccola domanda che il buon Nietzsche nel secolo storico, l’Ottocento, aveva posto (all’inverso del *Metaverso* futurologico di Zuckerberg) al mondo di ieri che veniva a coscienza storica, al peso dell’enfasi del passato sulla vita.

E cioè quanto futuro ignoto, quanto digitale, quanta *expeditio in novum*, può tollerare la nostra vita come l’abbiamo conosciuta senza doversi “aumentare” nella sua “essenza” (più banalmente: nella sua struttura percettiva, psico-fisica) per inseguire le indefinite possibilità di aumento dell’esperienza nella realtà digitale? Per giocare in un mondo digitale che ha sempre più i tratti di un *videogame* dove i morti si rialzano dopo essere stati abbattuti? A quale crisi di “prestazione”, se vuole essere all’altezza di questo “mondo” digitale, sarà sospinto l’uomo che conosciamo (si pensi solo per un dettaglio empirico alla già evidente crisi di prestazione della sessualità giovanile per il consumo del porno in rete)? Per tacere della già classica domanda nietzscheana strutturante il nostro rapporto con il passato, su quanta *memoria*, nei termini dell’*onlife*, della realtà *ri-ontologizzata* dal digitale, dalle ICT (cioè su quanti *Data*, ovvero informazioni già date, quante tracce mnestiche-cognitive magari affluenti in tempo reale, quello di una digitazione informativa) sia in sé capace di reggere l’*hardware* psico-biologico umano conosciuto; quello almeno che l’evoluzione fin qui ci ha consegnato nelle mani. Dietro una tale, inedita *promesse de bonheur* sembra celarsi una pulsione neo-gnostica (tecno-gnostica) che è vero e proprio disprezzo per il corpo, odio per la carne, “svilita a *wetware*, a sostanza putrida e vischiosa dalla quale bisogna liberarsi”⁹.

Al di là dell’ipotesi futurologica, su cui però *si ricerca* e *si opera*, di un potenziamento di massa del nostro attuale *hardware* psico-biologico per reggere come interfaccia – o più realisticamente per servirla meglio

⁹ Su questo tema rimando a Mazzearella E., *L’androide Philip Dick*, in Id., *L’uomo che deve rimanere. La smoralizzazione del mondo*, Quodlibet, Macerata, 2017: 171-198 (qui p. 175).

nella servitù volontaria di cui si vedono ormai imbevuti i milioni, a voler solo essere un po' *apoti*, di quelli che non se la bevono, – la “Megamacchina” dell'ambiente *onlife*. Megamacchina, della cui “intelligenza”, a noi superiore in ogni senso pensabile della *computazione*, al più possiamo ambire ad essere un “motore semantico” secondo il paradigma dell'*human-based computatio* (HBC, calcolo basato sull'essere umano). Acronimo di recente formulazione per definire “un paradigma per utilizzare la capacità umana nel processare i dati per risolvere problemi *non ancora risolvibili dai computer*”¹⁰. Nelle more ovviamente di questo “non anco-

¹⁰ In concreto una delle nuove figure professionali create dalle ICT e dalla AI: «Gli esseri umani stanno diventando “nuovi mezzi di produzione digitale”. Il punto è semplice: a volte l'AI ha bisogno di capire e di interpretare ciò che sta accadendo; quindi, ha bisogno di motori semantici come noi per svolgere il lavoro. Questa tendenza abbastanza recente è nota come *human-based computatio* (HBC, calcolo basato sull'essere umano). La definizione classica della HBC è recente: “Un paradigma per utilizzare la capacità umana nel processare i dati per risolvere problemi non ancora risolvibili dai computer”. (...) Si tratta di un servizio web di *crowdsourcing* che consente ai cosiddetti *requesters* di sfruttare l'intelligenza dei lavoratori umani detti *providers* o, più informalmente *turkers*, per svolgere compiti di HIT (*Human Intelligence Tasks*), che i computer non sono attualmente in grado di eseguire. Un richiedente pubblica un HIT, ad esempio per trascrivere registrazioni audio o taggare contenuti negativi in un film (due esempi reali). I *turkers* possono scegliere tra gli HIT esistenti, e completarli per un pagamento stabilito dal richiedente. A tutt'oggi, i richiedenti possono verificare se i *turkers* soddisfano alcune qualifiche specifiche prima di assegnare un HIT. Possono anche accettare o rifiutare il risultato inviato da un *turker*, cosa che si riflette sulla reputazione di quest'ultimo. *Human inside* sta diventando il prossimo slogan. La formula vincente è semplice: *agere artificiale + intelligere umano = sistema intelligente*». È singolare che il contesto di questa descrizione [«gli esseri umani possono diventare inavvertitamente parte del meccanismo, comportando non soltanto una separazione tra *agere* e *intelligere*, ma una inversione della relazione tra i due, mettendo l'intelligere al servizio dell'*agere*», dando ragione all'antivedente messa in guardia Norbert Wiener sulle pericolose potenzialità delle ICT di un loro egoistico sfruttamento per rendere più efficiente “l'uso umano di esseri umani” nella sua *Introduzione alla cibernetica*, 1954], dove si richiama oltretutto l'evidenza del ben più massivo entrare a far parte del *meccanismo onlife* degli esseri umani come «interfacce influenzabili», e dove «il mondo diventa sempre più complesso da gestire e quindi sempre più bisognoso di AI per funzionare bene», si muova poi in un contesto discorsivo di sostanziale sottovalutazione della servitù volontaria di massa ai calcoli di mercato e al mercato del calcolo in tutti gli ambiti della vita, solo perché questo nuovo mondo non avrebbe nulla a che fare con l'immaginario postumano di «una singolarità futura, in cui si oltrepasserà un punto (detto *Singularity*) oltre il quale l'AI avrà uno sviluppo incon-

ra”, e cioè nell’attesa che l’IA, servoassistita dalle ICT, realizzi il transito definitivo dall’attuale struttura sintattica dell’IA, alle sue ambizioni di farsi autonoma capacità *semantica* (significante, significativa, e produttore senso); cioè non più solo “intelligenza” computazionale ingegneristicamente *riproduttiva* dell’*agere*, del *fare* che le è stato programmato, ma *intelligenza para-umana* in senso proprio, *produttiva* cioè del *sensu* che, per quanto efficiente computazione, ancora non è capace di filare da sé al pari di quella particolare macchina biologica che siamo (la «cosa che pensa» di Cartesio, che cioè «dubita, concepisce, afferma, nega, vuole, immagina, sente») di cui vuol essere mimetica. Fosse pure il mero senso di *voler* vivere del biologico in quanto tale, che ci accomuna a ciò che nella biosfera ha “anima”: il vento (*anemos*) dello “spirito” insufflato nelle “cose”, o sortovisi da sé, *che vuole respirare*, che dice il greco *physis* in opposizione descrittiva all’*artificio* di cui si riconosce l’intenzionalità agente che lo mette in essere – “anima” sia pure minima, quella muta dell’istinto, non espressa a sé di un castoro che costruisce la sua diga o di un’ape la sua cella.

Un transito, un *salto evolutivo*, magari agito da un ultimo (e *primo* del futuro) intervento umano, un *fiat lux* o meglio *fiat neon* del nuovo “futuro”, della nuova creazione *onlife*, il cui ideale regolativo – l’infinito da conquistare del simbolo grafico ∞ di *Metaverso* – sembra essere la “smaterializzazione” sempre più spinta nel virtuale della vita reale.

trollato, si gestirà da sola e finirà per schiavizzarci». Senza rendersi conto che ciò di cui bisognerebbe preoccuparsi non sono questa ed altre distopie narrative, ma l’entificazione già in campo nel mondo del *social web*, che si avvia a transitare nell’*onlife*, di concretissimi processi di alienazione sociale, esistenziali e finanche percettivi in obbedienza ad un *esse est percipi* ormai declinato sempre più grazie al web in senso mediale-passivo come un essere percepiti che rimbalza e costruisce non solo il nostro percepire ma il nostro stesso percepirci. Il web essendo per comune ammissione la più potente tecnologia di manipolazione del sé sociale – individuale e collettivo – che si sia mai conosciuta. Non ci si rende conto, che il web è *la nuova gleba a cui siamo asserviti*, paradossalmente ancora più *stanziale* della vecchia gleba, perché è racchiusa nel fazzoletto di terra di uno schermo che ci viene fornito a “casa”, senza neppure necessità che si esca “in campagna”. Il processo di liberazione sarà lungo, l’implementazione degli esseri umani nella gleba digitale ancorché ben cominciata è ancora agli inizi, nella sua fase di sviluppo. Per parlare come i virologi di moda, non ha ancora raggiunto il *plateau*, da cui possa poi scendere.

Quasi una promessa dell'*onlife* di portare a termine una volta per tutte l'antica guerra dello spirito – la sua ambizione all'eterno, alla sua sempiternità modellata sulla presenza a sé della coscienza, patrimonio di religioni e filosofie – contro la sua materia biologica corruttibile di subíto accompagnamento esistenziale.

Ma in verità anche nell'*onlife*, passeremo solo da materia a materia, trascendendo non nello spirito eterno (promessa in cui la metafisica dell'ontologia creaturale cristiana, per completezza del desiderio di noi che esperiamo, ha incluso anche i “corpi” e non solo l'anima-mente), ma nella materia senza patema dei *chips*, dei *byte* come pura vittoria non dello spirito, ma della *materia fredda* della nostra origine *minerale*, prima che in essa, da essa stessa, venisse, emergesse, la febbre dello spirito che siamo *come spirito incarnato*, e cioè come carne che prende addosso sé stessa, la sua “vita”. E ne vive la febbre che “sa” del suo bruciarsi, la “sente” – come una candela che mentre illumina e si illumina si spegne. Portassimo questa nostra vita negli avatar, avremmo certo – fino a quando non degradi il supporto digitale in cui si proietta – la replicabilità virtuale, all'infinito, della nostra esperienza nei nostri avatar, come in un classico racconto di Adolfo Bioy Casares, *L'invenzione di Morel*. Ma a costo di restituire la “vita”, lo spirito che siamo, alla sua inerte figura senza fiamma – senza l'*interiorità* che si consuma, che brucia della sua *carne*, la *cera* del *biologico* che l'alimenta.

La *grande dismissione* nel virtuale del reale, questa è la posta in gioco dell'*infosfera*, la nuova parola-mondo con cui una filosofia troppo integrata al suo tempo, troppo poco inattuale, descrive e promuove oggi questa deriva dell'antropologia della tecnica. La forma-mondo della nostra realtà oggi, di un tempo-spazio in cui si trascendono l'uno nell'altro *online* ed *offline* della vita, l'*onlife*, appunto, dove “ciò che è reale è informazionale e ciò che è informazionale è reale”, come si è visto pretendere, con l'estensione dell'*esserci* umano in un nuovo ambiente: l'*infosfera*; un nuovo mondo a cui non ci si può sottrarre, come al reale hegeliano che in una sua equivocata lezione è sempre razionale, ha cioè le sue ragioni a cui non ci si può sottrarre e davanti a cui possiamo solo alzare le mani.

Come anche da *Wikipedia* possiamo sapere (per un utilizzo ironico di un segmento, ormai archeologico per i suoi profeti, dell'*infosfera*), nel

suo libro *Digital Dharma: A User's Guide to Expanding Consciousness in the Infosphere* Steven Vedro scriveva:

Emergendo da ciò che il prete-filosofo francese Pierre Teilhard de Chardin chiamava *noosfera* – come insieme condiviso del pensiero della collettività umana, dell'invenzione e della ricerca spirituale – l'*infosfera* è talvolta termine utilizzato per concepire appropriatamente uno spazio che comprenda i nostri corpi fisici, mentali ed eterici; coinvolge e incide sul nostro sognare, sulla nostra vita culturale. Il nostro sistema nervoso, sempre in evoluzione, può essere ora esteso, come Marshall McLuhan prediceva nei primi anni Sessanta, fino a diventare un abbraccio planetario¹¹. “Noosfera” di Teilhard de Chardin tanto quanto la nozione di “semiosfera” di Jurij Lotman possono essere quindi considerati quali precursori del concetto di *infosfera*, dove quest'ultimo pone maggiormente l'accento sull'organizzazione e la distribuzione delle informazioni nei sistemi elettronici, informatici e telematici¹².

Termine, *infosfera*, con cui già Floridi aveva inteso, in analogia con biosfera, «lo spazio semantico costituito dalla totalità dei documenti, degli agenti e delle loro operazioni»¹³, cioè delle interazioni sociali comunque agite *onlife* – *online* ed *offline* – su dati cognitivi in qualsiasi modo codificati e processati.

Ora, lasciando da parte la *noosfera* di Teilhard de Chardin, il punto da non dimenticare è che noi certo abitiamo da sempre il *noetico* (lo sappiamo almeno dai primi filosofi greci), ma sempre anche, e a sua base, il *bio-etico*; come l'ontologia creaturale cristiana ha per tempo ricordato come anche da “salvare” per sempre all'*integrità* del nostro esserci, se vogliamo fare le cose perbene con il nostro desiderio di vita “reale” eterna, da eternizzare, che è poi la nostra; vita che è spirito e carne. E l'abitiamo, il *bio-etico*, il nostro radicamento organico-vitale, come dimensio-

¹¹ Un abbraccio planetario in cui gli addetti ai lavori riconosceranno facilmente la costrittiva *dispositività* operativa e di controllo annunciata dal *Gestell* heideggeriano come cifra del mondo della tecnica, e prima ancora la weberiana *gabbia d'acciaio* di economia e società della seconda modernità.

¹² Steven Vedro, *Digital Dharma: A User's Guide to Expanding Consciousness in the Infosphere*, Wheaton, 2007.

¹³ Floridi L., [Infosphère, une définition](#), in *Boson2x*, 20 dicembre 2002.

ne di emergenza del noetico; radicamento da cui questo non può mai spogliarsi, anche se vuole “eternarsi” – sempre che non voglia diventare altra cosa che quello che sa da sempre di essere, cioè “mortale”. Condizione, la mortalità, da cui il nostro esserci reale non può mai redimersi da solo anche nell'*onlife*, cercando di esorcizzare l'altra faccia dell'*onlive*, che conosce da sempre, e cioè la sua esposizione all'*offlive*. Insomma, per dirla in altri termini, noi non abitiamo il noetico se non nel *bio-etico*, il biologico che abitiamo e da cui siamo abitati, cioè *com-portati* mentre ci *com-portiamo* in esso; *bio-etico* come sfera dei modi in cui la vita si porta a sé stessa e si comporta con sé stessa (il regno del biologico) di cui il noetico, il noetico che siamo, è la deperibile emergenza, che da sempre sogna in vario modo di non deperire assieme alla sua condizione di possibilità, l'*anima vitale*, l'*animazione vitale* che ci fa lo spirito che siamo: e cioè *l'incarnazione come presenza a sé di un'entità, un esserci – che è anche sempre un (eco)sistema di relazioni – che si prende addosso la sua carne*¹⁴.

2. L'INFOSFERA

Con l'*infosfera* come ambito dell'*onlife* saremmo giunti, per i nuovi profeti, alla “quarta rivoluzione” – a far data dal XVI secolo – dell'essere sociale dell'uomo. Saremmo cioè giunti all'*essere sociale* come *social web*, dove andrebbe in congedo l'obsoleta dicotomia tra vita reale e vita digitale: «La pervasività sempre crescente delle tecnologie di informazione e comunicazione (ICT) scuote le strutture di riferimento consolidate attraverso le seguenti trasformazioni: lo sfocamento (*blurring*) della distinzione tra reale e virtuale; della distinzione tra umano, macchina e natura; l'inversione dalla scarsità dell'informazione all'abbondanza dell'informazione e il passaggio dal primato delle *entità* al primato delle *interazioni*»¹⁵.

Uno scenario in cui la nostra vita sarebbe una *fluida transitività* tra *esperienze online* ed *esperienze offline*, senza maggiore o minore “autentici-

¹⁴ Mazzarella E., *La carne addosso. Annotazioni di antropologia filosofica*, in Buti I., Calcione C., Di Salvo S., Masi C., Doria, Reduzzi Merla F., Salerno F. (a cura di) *Fides Humanitas Ius. Studi in onore di Luigi Labruna*, Napoli, 2007: 3453-3463.

¹⁵ Floridi L. (editor), *The Onlife Manifesto. Being Human in a Hyperconnected Era*, cit.:7.

tà” delle une rispetto alle altre, e non ci sarebbe motivo di ritenere che ciò che avviene *online* sia meno “vero” di ciò che avviene *offline*. Il che è un truismo, un’ovvietà se per autenticità o verità degli eventi intendiamo che essi *in quanto accadono*, incidono, *hanno effetti*, siano entrambi – *eventi online* ed *eventi offline*, per altro nella loro fluida transività – “reali”. Ma proprio per questa ovvietà sono eventi che incidono potentemente sull’esserci dell’*entità* – cioè *noi*, i soggetti (attivi o passivi) dell’esperienza – che con essi si *co-eventua*, cioè accade e si fa insieme ad essi. Per cui è poco più che una *boutade* pensare che noi «siamo esseri umani che si immergono temporaneamente nel mondo digitale per poi riemergere, scollarci tutto di dosso e riprendere la nostra vita regolare: le due esperienze sono costantemente e profondamente intrecciate»¹⁶. Perché proprio *il passaggio dal primato delle entità al primato delle interazioni* come vuole *The Onlife Manifesto*, si traduce nella permeabilità e nella plasmabilità sempre più determinante dell’*entità* che, reali o virtuali, vive quelle esperienze; quell’*entità* che siamo *noi*, *la soggettività esperiente*; in un potenziamento dell’*essere* “ciò che fa” (ciò che esperisce e come lo esperisce) del nostro esserci (già assioma – nel mondo reale, non-virtuale – della costruzione percettiva, sociale e di esperienza del nostro io, che è sempre un io di relazione, un noi ambientato), che toglie proprio sostanza di io “autentico”, cioè *consapevolezza autogovernata*, al nostro “io” – e di conseguenza tanto più al nostro esserci collettivo, rinforzandone le già potenti dinamiche persuasive “anonime”; un “io” sempre più intriso di *avatara*. *Avatara*, cioè letteralmente “discesa”, “incarnazione” del virtuale nel reale, di fusione intima delle due nature, reale e virtuale, che verrebbe a farsi *sinolo* nell’entità umana nell’*onlife* (Treccani: «*Avatara*. - Letteralmente “discesa”; a un dipresso “incarnazione”. L’intimo significato di questo concetto religioso dell’India è, non di una manifestazione transitoria del divino, ma della presenza reale della divinità in un essere umano, di una fusione intima delle due nature in un unico essere, che è veramente dio e veramente uomo» – con il che è bello e deciso quale doveva essere lo “spirito”, quello *virtuale*, della tecnica, in definitiva una nostra protesi che ci torna addosso, che doveva scendere sull’uomo e nell’uomo).

¹⁶ Così il commento alle prospettive aperte dall’*onlife* di Andrea Daniele Signorelli, *cheFare*, 6 febbraio 2020.

Questo *avatarsi*, questo farsi avatar del nostro Sé, del nostro esserci, sarebbe richiesto dopo che con la “quarta rivoluzione” della modernità, quella del digitale e dell’IA, abbiamo perso l’ultima “centralità” che ci era rimasta dopo la perdita già patita della nostra centralità nell’universo (con la rivoluzione copernicana), nell’evoluzione (con la crisi dell’antropocentrismo nel darwinismo), nell’io cosciente (con la scoperta freudiana dell’inconscio). E cioè, la centralità che ci residuava, *quella della nostra capacità di calcolo*: «Ci era rimasta una sola centralità, che ci portava ancora a dire cose tipo “a scacchi non potrà mai batterci nessuno” oppure “vorrei proprio vedere se un robot è in grado di parcheggiare in uno spazio così stretto”»¹⁷.

Oggi che l’IA sconfigge da tempo i maestri di scacchi e le macchine si guidano e si parcheggiano da sé, anche la nostra centralità nell’universo del calcolo, delle prestazioni intellettuali è fortemente dubbia, se non obsoleta. Sempre ammesso che quelle del calcolo, della computazione siano prestazioni intellettuali *stricto sensu* se sganciate dall’orizzonte motivazionale di senso in cui siano “operate”. Per riguadagnarla, la nostra centralità, per essere all’altezza dell’infosfera, del suo potere di calcolo, per abitarla e non per esserne semplicemente abitati, cioè sostanzialmente “programmati” dal *sistema totale delle interazioni sociali real-virtuale*¹⁸, c’è bisogno – ci sarebbe bisogno per i profeti del transito nell’*iperstoria* dell’infosfera – appunto di *avatarsi*, di disincarnarsi (o di potenziare, nel e tramite il digitale, la carne sempre più disabile del “reale” alle funzioni che deve assolvere) nel nuovo mondo del virtuale. Di stare sempre *up-to-date* alla *documentalità* in incessante divenire dello *script*, della *scriptura digitale* che sta *ri-ontologizzando*, scrivendolo in proprio, in senso attivo, in algoritmo, il già galileiano libro della natura scritto in caratteri mate-

¹⁷ Floridi L., *Infosfera: Idee per capire il digitale*, Lectio al Teatro Parenti del 3 luglio 2012 (disponibile su: <https://www.youtube.com/watch?v=vlqLxj935Nc&t=3s>).

¹⁸ Uno spaventoso incremento dell’alienazione, del *Man* heideggeriano, del “si dice”, “si fa”, “si pensa” dell’anonimo sociale; a paragone dell’oggi, il *Man* heideggeriano di *Sein und Zeit* può ben apparire una tenera romanticheria da esistenzialisti, mentre ben più corposa anticipazione di realtà la sua evoluzione nel *Gestell* come critica della tecnica che radicalizza nella totalità del nesso struttura-sovrastuttura l’alienazione di Marx.

matici; ora però non più scritto *per noi ma “non” da noi*, ma scritto *da noi* fino a immaginare di poterlo sovrascrivere e persino riscriverlo *ex novo* cancellando nelle *entità* le scritture precedenti (in definitiva, come orizzonte di possibilità, scritto *da noi ma “non” per noi*, bensì *per un “altro” noi*, diverso da quello che *si* sovrascrive, lui stesso, nella sua stessa psico-bio-sociologia).

Ora, se si ha chiaro – come è chiaro – che la *memoria digitale*, l'*iperdocumentalità*, prodotta e depositata in rete, dell'*iperstoria* dell'*infosfera*, ha anch'essa, per problemi di sostenibilità della rete, necessità di svuotare le sue *cache* e aggiornare i suoi *cookies*, e pertanto «salva questo documento» di *default* per lo più «significa sostituisci le versioni precedenti, per cui ogni documento di qualsiasi genere è destinato a questa natura astorica e il rischio è che le differenze siano cancellate, le alternative amalgamate, il passato costantemente riscritto e la storia ridotta a un perenne qui e ora», per cui «il lavoro di custodia dell'informazione è destinato a diventare sempre più importante»¹⁹, è sorprendente immaginare di cavarsela nella situazione così descritta con una nuova figura professionale di “custodi della memoria digitale” (e chi custodirà poi i custodi, chi gli dirà cosa custodire?), per non vedere partecipare quest'uomo *iperstorico* in prima linea allo *shock antropologico* dell'età della tecnica e della globalizzazione, potenziata da un *digitale non di servizio all'analogico*, come bene o male è stato fin qui, ma dell'analogico *colonizzatore*. Lo *shock* di abitare un mondo, l'*onlife*, la realtà involupata (*envelopped*) nel digitale e dal digitale, «in cui siamo più ospiti (*guest*) analogici che ospitanti (*hosts*) digitali, [dove] non c'è da stupirsi che i nostri agenti artificiali [per potenza di calcolo e di connettività] funzionino sempre meglio, e meglio di noi; sono nel loro ambiente (...) una profonda trasformazione ontologica (vivamente! *n.d.r.*) che implica importanti sfide etiche»²⁰. Scenario la cui analisi di dettaglio, in questo approccio *positivo*, rigorosamente *non* apocalittico (“tanto non c'è niente da fare”), vede sempre e solo il *dito locale* di ogni

¹⁹ Floridi L., *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, cit.:19.

²⁰ Floridi L., *Agere sine intelligere. L'intelligenza artificiale come nuova forma di agire e i suoi problemi etici*, in Floridi L., Cabitza F., *Intelligenza artificiale. L'uso delle nuove macchine*, cit.:161-162.

problema emergente per cui si troverà la *soluzione*, o lo *specialista*, ma non la *luna (piena)* di una partecipazione più o meno irriflessiva all'ideologia progressiva dell'*infosfera* – al più moderatamente sorvegliata per consigliare le *policies* sociali più opportune al liberismo socialmente controllato vigente, e da far vigere, di una *biopolitica* che mentre crede di inglobare nella vita il virtuale, sta piuttosto inglobando la vita nel virtuale, tramite un *bioliberismo* che non ha più santuari intoccabili (*ineffabili*, non dicibili e non “scriptibili”) dell'individuo e della comunità nella sua continuità storica e biologica, epperò gestito illiberalmente in un controllo capillare, digitale, molecolare, a cui si unisce l'orizzonte coatto del *politically correct*, “sfogati” nel privato, ma “conformati” nel pubblico: «liberi di farsi e disfarsi, di coltivare l'erba voglio, ma poi coatti, vigilati e conformati nel nuovo sistema globalitario: bioliberismo in un regime illibertario»²¹.

Ma davvero l'*umanità* dell'uomo nell'*infosfera*, la si difende su questa frontiera del *calcolo*? Tanto da affidarvi – così pare debba essere in un inseguimento ad *handicap* dell'AI – l'*autocentratura* oggi possibile dell'uomo, il tratto specifico della sua “soggettività”? Che poi in effetti non è altro che un affannato tentativo di integrarsi alla centralità del calcolo come autonomo operatore di realtà di un mondo ristrettosi a generale sistema di calcolo. Il punto è che l'evangelo dell'*infosfera* l'*umanità* dell'uomo non la vede proprio più, e piuttosto ci mette a repentaglio di farcela perdere, *sfuocare*, “*sconfinata*”, nel virtuale. Questa centralità del *calcolo*, come capacità di interazione (di successo) con la realtà, il mondo, la nicchia ambientale – perché di questo si tratta quando diciamo calcolo – non è affatto, se non ha la *consapevolezza* dei *fini* del calcolo operato – lo specifico dell'umano. Questa capacità di calcolo come interazione (più o meno efficiente) con il reale – ammesso e non concesso che l'interazione di successo con il reale possa ridursi a calcolo – la condividiamo non solo con tutti i viventi (le *entità* della biosfera), ma persino con tutto l'«essere», la natura, la materia (inclusiva di quella che si cerebralizza dall'*octopus* all'*anthropos*). Quello che non condividiamo – e che *ci dividiamo solo tra noi*, cioè *con-dividiamo* tra noi, e ci fa *le entità che siamo*, cioè *noi*, gli uomini

²¹ Così efficacemente, Marcello Veneziani, *Se vince il bioliberismo, l'unico diritto che resta è quello di farsi e disfarsi*, “La Verità”, 14 febbraio 2022.

– è che noi siamo le uniche entità che sanno che cosa significa perdere una partita a scacchi, al di là della performatività del calcolo che metta sotto scacco il Re dell'avversario. E che sanno *che cosa significa perdere in generale*: la partita a scacchi della vita, e della propria vita, lo scacco dell'esistenza. Gli unici che sanno la fallibilità del calcolo. E come entità finite la sua fallibilità di principio. Che sanno le regole della scacchiera, *come* ci si muove su, ma non *perché* le regole sono quelle e non altre, e *perché* in genere *ci si muove*, *perché* un atomo si lega e poi decade e la vita ne sorge e ne declina, restituita al suo inanimato, alle sue radici minerali. L'*autocentratura* dell'umano, il *centro* che dell'uomo fa il suo *autos*, è nient'altro che questo: la nostra capacità di *soffrire*, *sufferre*, di portare su di noi il calcolo che siamo come speranza dello spirito di rimanere presso sé stesso nella sua forma conosciuta – *vita di presenza e vita di presenze*. Non c'è nessuna metafisica, neanche di avatar che ci portino nell'eterno digitale, dopo il reale, dopo *la fisica* che siamo. *Dopo la fisica*, anche quella dell'*onlife*, c'è solo *la fisica che muore*, l'unica *metafisica* conosciuta, cioè noi. Almeno che un Dio non ci salvi, nonostante i progetti di "ricerca" in merito dell'AI e delle *humanities* in ambiente *onlife*²². Ma questo a tutta evidenza è un'altra storia. Se così non fosse, non si spiegherebbe, perché oggi, ai tempi del Covid, stiamo disperatamente cercando di salvare, tra DAD e digitale, la vita di presenza, il calore umorale di un abbraccio. Di resistere nonostante tutto alla spinta che viene da ogni parte – mi si consenta un neologismo brutto quanto la realtà che descrive e non promuove – alla sostituzione *webinale* delle nostre vite. Di salvare la *presenza*²³.

²² Si veda in *The Online Manifesto*, cit., a pp. 125-159, il contributo di Yiannis Laouris, *Reengineering and Reinventing both Democracy and concept of Life in the Digital Era*, che si applica alla necessità di ridefinire in era digitale i concetti (in presenza delle nuove entità ibride umano-digitali) di mortalità/immortalità, di vita, essere vivo, umano, stante la pressione che l'immortalità emergente di artefatti e informazioni esercita sul raggiungimento dell'immortalità della mente e/o dell'umano; e la conseguente necessità di pensare la sostenibilità sociale, e l'esigenza di gestirla democraticamente, di questa "immortalità" degli artefatti intelligenti, fino a dover prevedere "il diritto all'eutanasia digitale" (sic!).

²³ Per una trattazione più complessiva di questo tema, rimando a Mazzearella E., *Contro Metaverso. Salvare la presenza*, Milano, 2022.

COSA RESTA DELL'UMANO NELL'EPOCA DELLA IA

Laura Palazzani

1. INTELLIGENZA ARTIFICIALE E/O INTELLIGENZA NATURALE?

L'IA "forte" si pone un obiettivo ambizioso: quello di progettare macchine *simili* all'uomo, che ne *imitino* le *funzioni*. Gli studi stanno sempre più andando nella direzione della imitazione tenendo come riferimento le neuroscienze, per costruire una "macchina pensante".

Emerge l'interrogativo: potrà una macchina pensare come l'uomo? Si tratta di un interrogativo tecnico se per "possibile" si intende "tecnicamente realizzabile". Ma non è di questo che si occupa la filosofia. La domanda della filosofia è antropologica: l'uomo è riducibile a macchina? La domanda morale si può articolare nel modo seguente: ammesso che la macchina possa diventare tecnologicamente simile all'uomo, è bene che la "macchina diventi umana" o "l'uomo diventi macchina"?¹.

2. L'ARTIFICIALE "OLTRE" IL NATURALE: LA PROSPETTIVA TECNO-CENTRICA

I tecnofili nell'orizzonte di una visione tecno-scientista, favorevoli ad ogni intervento di tecnologicizzazione del corpo e della mente, delineano orizzonti in cui l'artificiale diviene sempre più simile al naturale e tende a "con-fondersi", per annullare intenzionalmente la differenza tra uomo e macchina, in una simbiosi tra uomo e tecnologia, tra vita organica ed inorganica. I tecnofili spingono nella direzione di uno sviluppo delle tecnologie convergenti e della robotica/IA (ove il robot è l'*embodiment* o incorporazione della IA) che sostituisca e superi l'umano, unica strada

¹ Coeckelbergh M., *AI Ethics*, MIT, Cambridge (Mass.) 2020; Boden M.A., *AI: Its Nature and Future*, Oxford, 2016.

per superare i limiti biologici del corpo e i limiti neurologico-cognitivi della mente verso una perfezione trans-umana, post-umana, o anche “super-umana”.

L’elaborazione del concetto di “persona tecnologica/elettronica/artificiale”, nella duplice direzione dell’uomo che si artificializza e della macchina che si umanizza antropomorficamente², presuppone filosoficamente una concezione riduzionistica, funzionalista e dualista.

Una concezione riduzionistica parte dalla visione materialistica e meccanicista dell’uomo, che si riduce a oggetto o somma di parti anche producibili e assemblabili dalla tecnologia: il corpo si riduce a parti modificate e sostituibili meccanicamente e la mente a componenti neurali modificate e sostituibili informaticamente. La tecnologicizzazione del corpo (robotica) e la digitalizzazione della mente (IA) sostituiscono le funzioni umane (movimento e intelligenza) con funzioni artificiali.

La visione riduzionistica è strettamente connessa alla concezione del funzionalismo che ritiene che l’esercizio delle funzioni sia ciò che conta per identificare la persona, a prescindere dalla natura, considerata irrilevante. La natura dell’organismo vivente umano si svuota ontologicamente. Ciò che conta sono le funzioni espresse o esibite: è indifferente se di un uomo o di una macchina.

Materialismo, riduzionismo, funzionalismo sono propri di una prospettiva non cognitivista, che nega la possibilità per la ragione di conoscere “oltre” i fatti, gli unici che possono dirsi veri o falsi in quanto verificabili o falsificabili empiricamente. Si torna al dualismo antropologico corpo/mente di derivazione platonico-cartesiana: secondo questa visione la persona è scissa in corpo e mente, ove l’essere “incarnato” in un corpo biologico è considerato un “peso”, che può e deve essere superato dalla macchina, in grado di esercitare la funzione razionale e cognitiva senza corporeità biologica (la “mente disincarnata”), oppure con una corporeità meccanicizzata (robot). L’alterazione tecnologica diviene un “dovere evolutivo” (la *enhancing technological evolution*³) che consente la “scorcia-

² Benanti P., *The cyborg. Corpo e corporeità nell’epoca del postumano*, Assisi, 2012.

³ Harris J., *Enhancing evolution, The Ethical Case of Making Better People, Enhancing Evolution. The Ethical Case for Making Better People*, Princeton, 2007.

toia tecnologica” (*technological shortcut*) per il miglioramento dell'umano e della specie e un “imperativo tecnologico”, la cui omissione può essere considerata colpa e negligenza.

Si aprirebbe tecno-profeticamente un orizzonte futuro che dovrebbe portare ad un'alterazione radicale della natura dell'uomo, sino alla totale artificializzazione dell'umano sostituendo corpo e mente con sussidi meccanici ed informatici. La convergenza delle tecnologie diverrebbe convergenza umano/artificiale. Il condizionale è d'obbligo: trattasi di scenari futuribili e meramente speculativi.

Il trans-umanesimo, teoria presente nel dibattito contemporaneo, promuove eticamente l'abbandono del biologico e la transizione verso il virtuale/artificiale/digitale al fine di espandere le capacità umane, per avere vite migliori e menti migliori. Il movente del trans-umanesimo⁴ è la desiderabilità sovra-umana e iper-umana del perfezionamento, che si esprime in modo moderato⁵ nella ricerca di aumentare bellezza, resistenza fisica e aspettativa di vita (il potenziamento estetico, fisico e biologico); in modo radicale con la cancellazione della condizione umana stessa, percepita e vissuta come limite. L'obiettivo è espresso nella Massima Centrale del Transumanesimo (MCT): “è etico e desiderabile utilizzare mezzi tecnoscientifici per superare la condizione umana (data)”. In questa prospettiva il potenziamento dell'uomo si realizza nella “condizione tecno-umana” che si compie con l'evoluzione tecnologica⁶. Dal potenziamento delle funzioni umane si passa al potenziamento dell'uomo inte-

⁴ World Transhumanist Association, *Transhumanism Declaration*, <http://humanity-plus.org/learn/philosophy/transhumanist-declaration>; Bostrom N., *Superintelligence*, Oxford, 2014.

⁵ McNamee M.J., Edwards S.D., *Transhumanism, Medical Technology and Slippery Slope*, *Journal of Medical Ethics*, 32, 2006: 513-514.

⁶ Cfr. FM-2030, *Are you a Transhuman?*, London, 1989; Bostrom N., *Welcome to a World of Exponential Change*, in Miller P., Wilsdon J. (a cura di), *Better Humans? The Politics of Human Enhancement and Life Extension*, London, 2006: 40-50; Savulescu J., Maslen H., *Moral Enhancement and Artificial Intelligence: Moral AI?*, in Romporti J., Zackova E., Kelemen J. (a cura di), *Beyond Artificial Intelligence: The Disappearing Human-Machine Divide*, Switzerland, 2015: 79-96.

so come uomo potenziato (*enhanced human*) post-umano⁷ o uomo come macchina. È l'orizzonte teorico che tenderà a svuotare i corpi umani in carne ed ossa, riducendoli a componenti biotecnologiche meccaniche ed elettroniche, a flussi di informazioni mutanti in grado di assistere, fino a sostituire, i processi vitali dell'organismo promettendo una perfezione illimitata. Ciò che conta è che la funzione "funzioni" il più perfettamente possibile, il corpo non è necessario, se non come involucro della funzione. Nella convergenza delle tecnologie è l'uomo stesso che diventa tecnologia in una sorta di "pan-tecnologismo"⁸.

Si tratta, forse, di «fantasie sciocche o prognosi verosimili, tardivi bisogni escatologici oppure inediti modelli fantascientifici»⁹, come sostiene J. Habermas. Sono comunque scenari che si stanno prefigurando almeno speculativamente (si pensi alle teorizzazioni della superintelligenza e della singolarità), rispetto ai quali non ci si può limitare a prenderne atto, ma è indispensabile una riflessione critica che ne colga le problematiche. Non si tratta, a fronte di paure rispetto all'avanzare di queste prospettive, di aderire al tecno-catastrofismo tecno-fobico, ma è indispensabile elaborare una riflessione filosofica critica sull'umano, sul suo significato e valore, che individui i limiti eventuali delle tecnologie. Non si tratta di esaltare la tecnologia disprezzando l'uomo o di esaltare l'uomo disprezzando la tecnologia: l'obiettivo è quello di consentire interventi sull'umano senza snaturarne l'identità, evitando trasformazioni irrimediabili.

Partendo da una considerazione: la tecnologia non è un "destino", ma siamo noi che costruiamo la tecnologia. In questo senso non dobbiamo limitarsi a prendere atto di cosa "resta umano" delle tecnologie,

⁷ Marchesini R., *Post-human. Verso nuovi modelli di esistenza*, Torino, 2004; Pepperell R., *The Posthuman Condition Consciousness. Beyond the Brain*, Bristol-Portland (OR) 2003; Pirni A. (a cura di), *Il postumano realizzato. Orizzonti di possibilità e sfide per il nostro tempo*, Novara, 2016.

⁸ Viola F., *Umano e post-umano: la questione dell'identità*, in Russo F. (a cura di), *Natura, cultura, libertà*, Roma 2010: 89-98; Grion L. (a cura di), *La sfida postumanista. Colloqui sul significato della tecnica*, Bologna, 2012.

⁹ Habermas J., *Il futuro della natura umana. I rischi di una genetica liberale*, Torino, 2002: 43-44.

ma spingerci oltre e considerare anche e soprattutto cosa “deve restare umano” con le tecnologie.

3. LA RILEVANZA DELL'UMANO: LA PROSPETTIVA UMANO-CENTRICA

Per quanti avanzamenti abbia compiuto la ricerca di una macchina che imiti e riproduca la complessa organizzazione cerebrale e che possa imitare tutte le caratteristiche dell'intelligenza umana, sembra lontana dalla portata degli sviluppi tecnici realisticamente prevedibili. Certamente le macchine (i software e i programmi informatici) sono oggi in grado di compiere funzioni e operazioni considerevolmente complesse comparabili – o in alcuni ambiti, anche superiori – a quelle umane: si pensi in particolare alla raccolta e alla selezione e catalogazione/archiviazione di informazioni in tempi più veloci che siano umanamente possibili, la capacità di calcolo¹⁰. Più complessa da realizzare e dispendiosa sia finanziariamente che energeticamente la mobilità del robot¹¹.

I tecno-ottimisti pensano che sia solo un problema di tempo, ma il progetto è chiaro: quello a breve di sostituire ogni funzione per superare i limiti fisici, di un corpo fragile e mortale, che si possa meccanizzare antropomorficamente con costruzioni robotiche; i limiti di una mente che si potrebbe estendere quantitativamente e qualitativamente in modo indefinito e indefinibile a operare come e oltre l'umano in ogni contesto, ambito e situazione.

Ma, esistono differenze qualitative o essenziali tra intelligenza artificiale e intelligenza umana? Potranno le macchine sostituire interamente l'uomo? E, ammesso che sia tecnologicamente possibile, sarebbe auspicabile? La domanda riguarda la possibilità e auspicabilità di riprodurre macchine che in *qualsiasi ambito* di pensiero e azione tipicamente umani possano sostituire l'umano, la possibilità o auspicabilità di riprodurre *interamente* l'intelligenza umana e anche oltrepassarla. Ossia, ri-

¹⁰ Comitato Nazionale per la Bioetica, Comitato Nazionale per la Biosicurezza le Biotecnologie e le Scienze della Vita, *Intelligenza artificiale e medicina: aspetti etici*, 2020.

¹¹ Comitato Nazionale per la Bioetica, Comitato Nazionale per la Biosicurezza le Biotecnologie e le Scienze della Vita, *Sviluppi della robotica e della roboetica*, 2017.

produrre le funzioni/operazioni e la stessa struttura o la costituzione dell'intelligenza umana come substrato.

Ci sono ragioni per ritenere che sia “meglio” sul piano antropologico, ontologico, morale difendere la *non riproducibilità*, la *non sostituibilità* e l'*unicità* dell'intelligenza umana? Vi sono alcuni aspetti che rendono difficile, di fatto, pensare che un giorno le macchine potranno sostituire interamente l'intelligenza umana.

Vi è una dimensione umana che non appartiene alle macchine: la autoconsapevolezza. Le macchine pensanti non sono consapevoli di sé, non si autoidentificano. Non sono in grado di riferire la somma delle proprietà/ funzioni e la serie degli atti/operazioni che compiono ad una unità che ne costituisce la sintesi. La IA non ha un'identità nel tempo che permane alla modificazione dei caratteri.

Nell'ambito del pensiero occidentale la filosofia ha riconosciuto anche la rilevanza di affetti, emozioni, sentimenti, e “appetiti” nell'ambito della cognizione umana. Anche oggi dal punto di vista sperimentale emerge che le emozioni influiscono sui processi decisionali e razionali delle scelte. La scelta umana non deriva solo da un processo interamente logico che colleziona informazione, lavora e calcola le informazioni immagazzinate in vista della decisione che massimizzi i benefici e minimizzi i rischi. Da quanto sappiamo oggi, l'intelligenza artificiale non è in grado di riprodurre gli aspetti emotivi e affettivi: e questo pone già un limite al sogno tecnologico dell'intelligenza artificiale forte. Chi sta progettando “macchine pensanti” predilige la dimensione “razionale” e “calcolante” nella direzione della convenienza in linea con il pensiero utilitarista. Ma una macchina intelligente che volesse imitare e riprodurre integralmente l'intelligenza umana dovrebbe includere anche la generale influenza degli stati emotivi-emozionali sulla cognizione e sulla decisione. È questo un percorso che necessiterebbe di un confronto tra informatici, filosofi, psicologi, teorici della mente. Il riconoscimento della stretta connessione tra dimensione cognitiva ed emotiva, nell'ambito individuale ed interindividuale, ha evidenziato il c.d. “errore di Cartesio”¹², mettendo in luce la complessità della soggettività, irriducibile ad

¹² Damasio A.R., *L'errore di Cartesio. Emozione, ragione e cervello umano*, Milano, 1995.

oggettività. In questo senso si coglie un aspetto dell'inappropriatezza delle profezie dell'avvento di "menti disincarnate". Il presupposto da cui partono le ricerche convergenti è la possibilità di simulare e sperimentare su calcolatore il funzionamento della mente, in una prospettiva funzionalista, misconoscendo la complessità della mente umana e del rapporto cervello/mente con le emozioni, nella esperienza vissuta.

Un aspetto emergente problematico dell'intelligenza artificiale (forte) riguarda la distinzione – chiara nella filosofia del linguaggio contemporanea – tra sintassi e semantica. I calcolatori e software, che supportano l'intelligenza artificiale, possono operare sui legami sintattici tra simboli, mentre non potrebbero considerare la dimensione semantica, ossia interpretare i significati. Questo porrebbe un ulteriore ostacolo alla possibilità di realizzare esaustivamente un'intelligenza artificiale che possa imitare *in toto* l'intelligenza umana. La semantica non è riducibile alla sintassi e la potenza di calcolo per quanto potrà essere sempre più avanzata in futuro non sarà in grado di trattare gli aspetti semantici complessi sul piano della interpretazione possibile all'intelligenza umana. Il ragionamento umano si basa su concetti e significati. Se la sintassi è costituita da una serie di regole di composizione generali applicabili ad ampie classi di elementi linguistici senza riguardo al significato, la semantica non è solo la rete di relazioni tra termini, ma è anche *l'esperienza reale*. Su questa base, dunque, per realizzare macchine pensanti non sarebbe sufficiente la potenza di calcolo e la computazione sintattica di simboli, ma ci sarebbe bisogno di interpretare e forse anche di esperire, fare esperienza/sentire, come un essere umano (soffrire e gioire, desiderare e temere, vedere, udire, toccare, odorare e gustare). A questo punto, dunque, l'intelligenza artificiale dovrebbe anche essere progettata con una "vita artificiale", dunque non solo "macchine pensanti" ma anche "macchine viventi".

Vi è anche una dimensione "motivazionale" che, al momento, resta solo umana. La motivazione a cercare la verità, a conoscere la realtà, a inventare, a immaginare, a creare. Una macchina pensante che volesse riprodurre l'intelligenza umana non dovrebbe solamente essere in grado di compiere operazioni intelligenti per risolvere compiti assegnati dall'uomo, ma dovrebbe essere anche in grado di auto-assegnarsi compi-

ti, di avere aspirazioni nella direzione della crescita e della conoscenza. Oggi, reti-neurali sofisticate sono in grado di compiere operazioni intelligenti (nel senso di cognitive e calcolanti) in maniera anche superiore rispetto all'essere umano, con programmi in grado di auto-modificarsi per migliorare le proprie prestazioni. Ma senza una motivazione, che deriva sempre, ad oggi, dall'input umano.

Ciò che manca alla macchina e resta proprio solo dell'umano è la *possibilità di percepire se stesso come un io e di essere riconosciuto dagli altri come un tu*¹³. L'uomo è in grado di percepire la relazionalità come un "essere" (ossia come una condizione ontologica e antropologica) e come un "dover essere" (ossia come una condizione etica e normativa). In questo senso si giustifica l'eccezionalità dell'uomo rispetto agli altri esseri viventi e alle macchine: solo all'uomo va riconosciuta una dignità intrinseca in quanto è l'unico essere in grado di riconoscere il suo dovere relazionale, dunque, è in grado di agire moralmente (rispetto agli altri esseri, simili e dissimili da lui). In questo senso l'uomo, e non la macchina, è propriamente autonomo, in quanto in grado di darsi una norma morale, mentre la macchina può al massimo essere imprevedibile, mediante gli algoritmi.

Le dimensioni emotiva-emozionale, esperienziale, interpretativa, intuitiva e intenzionale-motivazionale, insieme alla autoconsapevolezza, alla libertà e alla relazionalità interpersonale, sono elementi e segni che mostrano la unicità umana, che evidenziano la differenza uomo-macchina. Una differenza che esiste ora di fatto, e che deve rimanere, di principio, per difendere la unicità dell'umano. Nella misura in cui identifichiamo la specificità umana e le *dimensioni insostituibili* dell'umano, porremo dei *limiti* alla possibilità tecnologica di costruire robot androidi o macchine pensanti: un limite al dominio tecnologico e all'aspirazione di "rifare" (*remake*) la natura, in quanto la tecnologia può anche disumanizzare e deumanizzare¹⁴. L'umano va preservato, perché "conta" anche nell'epoca delle macchine.

¹³ D'Agostino F., *La dignità umana, tema bioetico*, in Cattorini P., D'Orazio E., Pocar V. (a cura di), *Bioetiche in dialogo. La dignità della vita umana, l'autonomia degli individui*, Milano, 1999: 103 e ss.

¹⁴ Kaplan D.M. (a cura di), *Readings in the Philosophy of Technology*, Lanham (MD), 2004.

L'IA è progettato e prodotto da un essere umano: siamo noi che dobbiamo decidere come produrre l'IA. Non sappiamo cosa sarà fattualmente e tecnicamente possibile: ma sappiamo oggi cosa è umano e cosa vogliamo conservare di ciò che è umano in una prospettiva incentrata sull'uomo (c.d. *human-centric* o *human-centered*).

4. IL CONTROLLO UMANO SIGNIFICATIVO DELLA MACCHINA

Anche se gli esseri umani costruiscono AI, selezionano dati, elaborano algoritmi, addestrano macchine, devono mantenere il controllo e la supervisione su ciò che progettano, programmano, applicano; le macchine dovrebbero rimanere un "supporto" alle decisioni umane, che cognitivamente le "assistono" ma non le sostituiscono.

Le macchine non dovrebbero "competere", ma "completare" le azioni umane. In questo senso anche il linguaggio va mantenuto antropocentrico, considerando le macchine di IA come "automatiche" piuttosto che "autonome" nell'apprendimento. Il termine "autonomia" non può essere applicato ad artefatti, anche se sistemi complessi o cognitivi molto avanzati. La terminologia dei sistemi "autonomi" è tuttavia ampiamente utilizzata nella letteratura scientifica e nel dibattito pubblico per riferirsi al più alto grado di automazione e al più alto grado di indipendenza dagli esseri umani in termini di operazioni e decisioni¹⁵.

La necessità di mantenere il controllo umano resta essenziale anche per evitare il possibile problema della "delega tecnologica". Un siste-

¹⁵ The High-Level Expert Group on Artificial Intelligence, in *Ethics Guidelines for Trustworthy AI* (April 2019) sottolinea «"human-centric" dimension of the new technologies». The European Group on Ethics in Science and New Technologies afferma «the importance that humans - and not computers and their algorithms - should ultimately remain in control, and thus be morally responsible»; «The principle of human dignity, understood as the recognition of the inherent human state of being worthy of respect, must not be violated by 'autonomous' technologies », (p. 16). The European Commission, *White Paper On Artificial Intelligence - A European approach to excellence and trust* (19 February 2020: 12) ammette che «the specific characteristics of many AI technologies, including (...) 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 right»; cfr. Recommendation of the Council on OECD, *Legal Instruments Artificial Intelligence*, 2020.

ma esperto che diventa ottimale nel suggerire “decisioni” all’uomo pone il rischio di diminuire l’attenzione umana con la possibile conseguenza di ridurre le abilità umane (il cosiddetto fenomeno di de-professionalizzazione), riducendo la responsabilità andando verso l’artificializzazione delle scelte che possono impoverire e perfino cancellare il rapporto interpersonale. In questo senso, è importante disciplinare la sinergia produttiva come complementarità tra uomo e macchina, ricercando modalità di “sostegno” intelligente che consentano all’uomo di avere “un controllo umano significativo” (*meaningful human control*) in termini di attenzione, contributo, controllo e responsabilità. Il dibattito sulla affidabilità, spiegabilità, tracciabilità, trasparenza e algor-etica ne sono un importante esempio¹⁶.

5. PER UN UMANESIMO TECNOLOGICO

Rifondare un umanesimo tecnologico significa evitare da un lato di “antropomorfizzare” la macchina (umanizzare la tecnologia) o artificializzare l’uomo a “macchina biologica” (tecnologizzare l’uomo). La “macchina” può essere una alleata dell’uomo e l’integrazione tra naturale ed artificiale può divenire complementare, sempre nella priorità del rispetto dell’umano, salvaguardando ciò che “conta” dell’umano, nel quadro dei diritti umani fondamentali.

¹⁶ Palazzani L., *AI and health: ethical aspects for regulation*, Teoria e Critica della Regolazione Sociale, 1, 2021: 16.

LE DOMANDE DELLA TEOLOGIA CRISTIANA SULLA IA

S.E. Mons. Antonino Raspanti

1. IL PROBLEMA DELL'ORIGINE E DEL LIMITE

Le frontiere di ricerca scientifica e tecnologica pongono radicali domande alla teologia, forse non del tutto nuove; proviamo ad elencarne alcune, cercando di indicare spiragli sui quali si può incamminare una riflessione compiuta. Sono domande radicali perché interpellano *in radice* capitoli della teologia come quello delle origini, che include, se non in vero parte dal compimento cristologico, nel quale non solo si svela con pienezza il disegno divino, ma si comprende anche la profonda lacerazione del vissuto umano e cosmico che riassumiamo nell'inclusiva questione del male. La teologia cattolica nella prima metà del Novecento ha posto le basi per una revisione della concezione sull'evento della creazione e sulla concezione della "natura", acquisendo nuovi approcci ermeneutici ai testi scritturistici. Si è accantonata l'ipotesi di una "natura" esistente e autonoma sulla quale sopraggiunge in modo estrinseco un "soprannaturale" che la destina alla comunione con Dio. Peraltro, dopo gli scontri frontali con le teorie evoluzioniste, che si sono contrapposte a quelle fissiste che la teologia sposava, assumendole da posizioni filosofiche, una nuova convergenza tra la teologia, la scienza e la filosofia si è fatta strada. Il mondo scientifico, per suo conto, ha chiarito meglio i vari livelli di discorso (quello dei dati documentari, quello delle teorie esplicative specifiche e quello delle visioni totalizzanti)¹.

¹ Cfr. Muratore S., *Creazione*, in Barbagallo G., Bof G., Dianich S. (a cura di), *Teologia*, Cinisello Balsamo, 2002: 313. Sono ormai classici alcuni trattati: Westermann C., *Creazione*, Brescia, 1974; Moltmann J., *Dio nella creazione. Dottrina ecologica della creazione*, Brescia, 1986; Gisel P., *La creazione*, Genova, 1987; Ganoczy A., *Teologia della natura*, Brescia, 1997; Arnould J., *La teologia dopo Darwin. Elementi per una teologia*

Pur consapevoli che le varie tribù ebraiche conoscessero e condividersero la credenza dei popoli circostanti circa l'esistenza di un Dio che avesse originato il mondo, la teologia biblica di area protestante e, in seguito, anche cattolica hanno evidenziato come gli ebrei siano pervenuti a una precisa visione di quel Dio dell'alleanza che è il creatore di tutto, quale fonte inesauribile e costante del suo essere e del suo operare. Questo non toglie nulla all'esistenza di una vita intelligente, che esige una compiuta intelligibilità di sé e del creato intero, e un suo ineliminabile riferimento a un orizzonte di trascendenza. Inoltre, la vita intelligente comporta la responsabilità umana nel creato, anzi la rende possibile e la sollecita nella ricerca di un senso universale che la provoca e la invita.

Da questa prospettiva ci si avvicina all'origine della vita umana, alla possibilità di riprodurla e, pertanto, all'eventuale limite con un trans-umano di qualsiasi genere. Questi tentativi sembrano soltanto una narrazione, puro frutto di aspirazioni umane, mai nascoste nel cammino storico, che si costruisce partendo dalle proprie scoperte o queste ultime sono guidate, almeno in parte, secondo quelle aspirazioni ed espresse in una narrazione avveniristica e, come sempre, con una visione del futuro idilliaca? Come individuare quel limite?

Secondo la rivelazione possiamo immaginare questo limite come una linea che, in quanto tale, separa nettamente e nello stesso tempo congiunge indissolubilmente le due parti: Dio e uomo/donna. Come attestano gli studi esegetici della Bibbia, gli autori ispirati pur assumendo pezzi delle narrazioni umane circolanti nei contesti culturali locali dell'epoca, smontarono il racconto totalizzante, ricreando una nuova narrazione che fotografava il tessuto storico similmente alle radiazioni che attraversano insensibilmente la materia, facendone conoscere le profondità meno manifeste. Dai racconti genesiaci all'ampia letteratura veterotestamentaria la sovranità *ab-soluta* del Dio di Abramo e di Mosè rispetto a qualsiasi creatura è continuamente affermata, con un cammi-

della creazione in una prospettiva evolucionista, Brescia, 2000. Qualche pubblicazione più recente: Kehl M., «E Dio vide che era cosa buona» - *Una teologia della creazione*, Brescia, 2009; Williams R., *Cristo, cuore della creazione*, Brescia, 2020; Bonino S. T., Mazzotta G. (a cura di), *Dio creatore e la creazione come casa comune - Prospettive Tomiste*, Roma, 2017.

no che va dall'enteismo al monoteismo e con la corrispondente battaglia anti-idolatrice combattuta per secoli dai profeti.

Nell'incontro con la filosofia ellenistica questa sovranità sarà tradotta con la trascendenza e la piena immaterialità del Dio dei padri, il quale con la sua sapienza tutto crea, ordina e pervade, mantenendo in vita ma anche sotto il suo giudizio. Tale concezione rigorosamente monoteista si è "scontrata" fino a ricomprendersi con l'evento Cristo e l'impensabile unione tra divino e umano, che quest'ultimo effettua, introducendo un'articolazione vitale trinitaria all'interno della divinità. La comprensione della persona di Gesù Cristo avvenuta nei quattro secoli successivi ha sancito la perfetta unione tra Dio e uomo, implicante la piena solidarietà tra i due, insieme alla loro netta distinzione, e con la non-identità Padre-Figlio spiega la possibilità di una creazione libera, autonoma eppure destinata alla partecipazione alla relazione interna alla divinità.

«Dio ha creato l'uomo: come va pensato questo evento? – nella misura in cui lo possiamo pensare?»², si chiede Guardini. Egli risponde non solo ricordando che se Dio come un artista allestisse un'opera e poi la lasciasse a se stessa, l'uomo affonderebbe nel nulla; al contrario, lo mantiene costantemente fuori dal nulla. Inoltre, l'uomo è rapporto a Dio, incomprendibile se si partisse dal basso. La chiave di sé non è in mano all'uomo, perché egli è frutto di un atto sovrano, non necessitato, libero e autonomo, di Dio, un atto che non è "nel" tempo, ma istituisce il tempo e lo mantiene in essere. Per altro, affidando all'uomo e alla donna la cura dell'intero creato, è legittimata, con una chiara visione desacralizzata di esso, anche ogni possibilità di plasmarlo totalmente, di cooperare all'atto creativo di Dio, rimanendo il limite/trascendenza dentro l'umano medesimo e tra il tutto, umano compreso, e Dio. Non c'è pertanto un limite all'azione dell'umano sul creato; da questo punto di vista la ricerca scientifica e le applicazioni tecnologiche sono aperte e opportune. Se si apre la strada alla violenza e alla distruzione, ciò avviene perché il vero limite non è stato riconosciuto e rispettato. Tanzella-Nitti aggiunge il guadagno della nozione di partecipazione, che il pensiero cristiano ha rielaborato dal platonismo, per avere una migliore comprensione dell'autono-

² Guardini R., *Antropologia cristiana*, Brescia, 2013: 37-38.

mia del creato e del rapporto d'indipendenza dal Creatore. La suprema causalità di Dio permette ad ogni effetto creato di partecipare dell'essere e delle sue perfezioni trascendentali "prendendone parte, senza esserne una parte". Solo Dio può partecipare l'essere in questo modo incomposto perché, essendo una causa staccata dal mondo, è capace di crearlo *ex nihilo*, dal nulla. La nozione di partecipazione risolve in modo definitivo la visione panteista, chiarendo che Dio è l'essere di ogni cosa non come costitutivo essenziale, bensì come causa³.

Da qui la domanda converge sulla natura di quel limite, ma anche sulla sua riconoscibilità, se sono due (tra Dio e il creato, tra l'uomo e le altre creature) o uno solo. La rivelazione lega strettamente una signoria/trascendenza umana sulle creature alla signoria/trascendenza divina su tutto il creato. Quella umana è ovviamente costituita e donata dalla divina. L'israelita/umanità impara la propria signoria/trascendenza rispetto alle altre creature scendendo verso le proprie origini tramite l'esperienza del Dio della liberazione, del roseto che arde e non si consuma, che guida le sorti dei popoli, che tutto ha posto, conserva e domina, resistendo ai superbi e facendo grazia agli umili. Sperimentando un Dio signore/trascendente, dovremmo dire propriamente "tre volte santo", egli è condotto a sapere la propria signoria e impara di essere ad immagine e somiglianza di lui. Sperimenta persino che quegli si fa difensore e custode dell'uomo/donna, anche quando questi fosse assassino e gli si ribellasse, ma guarda dall'alto la pretesa di costruire dal basso una torre che giunga fino a lui, disperdendone il tentativo con un piccolo soffio, che suona ancora come protezione dell'umano da se stesso.

Questa signoria/trascendenza dell'uomo su tutte le creature trova soprattutto il fondamento nella chiamata alla comunione con Dio, destinazione già contenuta nell'atto creatore. Tale realizzazione suprema e definitiva è conosciuta soltanto alla luce dell'incarnazione del Figlio eterno in forma umana. «La relazione del Figlio al Padre non è superabile da alcun'altra forma di rapporto con Dio ... L'uomo già come tale è ormai elevato al di sopra del mondo della natura e in certo senso al di sopra dei rapporti di potere vigenti nel mondo sociale. Il fatto che egli sia destinato alla

³ Cfr. Tanzella-Nitti, G., *Creazione*, (online) <https://disf.org/creazione>.

comunione con Dio rende la sua vita, nella persona di ciascun individuo, inviolabile. È essa quella che giustifica una dignità che noi riconosciamo propria e imperdibile in ogni persona umana»⁴.

Se nella risurrezione di Cristo questo compimento del destino umano è definitivo, allora l'intera causa umana, e del creato che in lui si riassume, è messa al sicuro da ogni minaccia o, quanto meno, dalla possibilità che la minaccia raggiunga il suo scopo. Non solo, pertanto, la chiave del proprio esistere non è nelle mani dell'umano, e ancor meno lo è la sua conoscenza, a lui soltanto rivelata, ma non è neanche nelle sue mani la distruzione di sé.

Di contro, però, assistiamo a una catena inarrestabile di violenza e morte in ogni angolo della storia, alla presenza spesso dominante del dolore e del disordine, con la trafiggente percezione di non riuscire ad aver ragione degli effetti devastanti di ciò che possiamo riassumere con il termine "male". «La tematizzazione biblica, e soprattutto cristiana, del male come peccato, cerca le radici del male nell'uomo stesso, anzi in linea di principio in ciascun uomo, e non semplicemente negli altri»⁵, o all'esterno in strutture sociali. Ciò mostra che la questione è legata a quella della signoria/trascendenza, che è dentro l'uomo stesso, al tentativo di ignorare il fatto che essa è donata, appropriandosene come "*ab-soluta*". La conseguenza è il totale disorientamento, compreso il fatto di non individuare più l'esistenza e la collocazione di un limite.

Il cristianesimo per male/peccato non intende soltanto fenomeni manifesti, anche perché raramente esso si manifesta in tutta la sua virulenza, né intende solo le colpe attuali dei singoli, che altrimenti potrebbero essere facilmente estirpabili bloccando e condannando costoro. Il cosiddetto peccato originale intende l'universalità e, quindi, la solidarietà di tutti con il male che precede ogni agire individuale e che la legge fa emergere. Con Agostino d'Ipbona il cristianesimo ha individuato la radice di ogni male nell'arbitrarietà umana con la quale egli si pone al centro il proprio "io", sfruttando ogni creatura come mezzo in vista della propria affermazione, un "*amor sui propter se*", direbbe Bernardo di Chiaravalle. Ecco per-

⁴ Pannenberg W., *Teologia sistematica 2*, Brescia, 1994: 204.

⁵ *Ibidem*: 273.

ché la radice in ultimo è svelata come un tentativo di farsi eguali a Dio, un vedere Dio come qualcuno che sbarra la strada all'umano verso la propria felicità, senza riconoscere che in questa spinta alla felicità si nasconde una brama di affermazione di sé, percorribile grazie al libero arbitrio ricevuto, ma letale negli esiti. Cadendo la signoria/limite/trascendenza divina cade contemporaneamente quella umana, e cade insieme la capacità di riconoscere tutto questo, oltre a non poter porvi rimedio. L'uomo non conosce più il dono, la gratuità e la correlativa libertà. Questa dinamica si rende pienamente manifesta nel processo che conduce alla condanna e all'uccisione dell'uomo-Dio, Gesù di Nazareth. Lì emerge contemporaneamente quel che dalle prime pagine delle Scritture si narra, che cioè Dio interviene continuamente e dalle origini a "salvare", a porre rimedio a questo processo che sembrerebbe avviato alla catastrofe.

Nella capacità di autonomia creaturale che l'uomo ha ricevuto è insita la possibilità di non tener conto e trasgredire il limite; si innesca così la dinamica del male/peccato il cui aspetto tragico risiede nella differente gradualità del suo manifestarsi e, soprattutto, nell'incapacità umana di individuare nettamente i contorni e la natura. Questo nuovo modo di vedere, cieco ed empio negli esiti, secondo le Scritture, crede di sapere quali vie conducono alla felicità, all'affermazione di sé, pervenendo allo scontro frontale con la legge di Dio, giudicandola sprezzante della vita. La sua cecità non sa che «il salario del peccato è la morte», come afferma san Paolo (*Rm* 6, 23), cui fa eco Guardini: «una nebbia infernale rende indistinta ogni cosa»⁶. Da un sapere che la propria signoria e libertà è costituita nella sovrana volontà di Dio, da cui gli vennero diritto e potere d'andare avanti nel proprio esistere; da un vivere la finitezza quale felicità e potenzialità capace d'ogni adempimento con coraggio e fiducia e gioia (ne è espressione l'Eden), l'uomo passa all'angoscia, in cui non sopporta il suo esser finito; pretende di non esser più immagine, bensì archetipo, ossia d'essere infinito-assoluto. Così facendo egli rimane bensì finito, perde però la connessione con la propria origine. Ora la fiducia si capovolge in *hybris*, e il coraggio in paura⁷.

⁶ Guardini R., *Antropologia cristiana*, cit.: 53.

⁷ Cfr. Guardini R., *Accettare se stessi*, Brescia 1992: 21.

La rivelazione cristiana non fa che illuminare gli abissi di un'iniquità, percepibile indipendentemente dalla conoscenza di fede; se così non fosse, ricorda Pannenberg, «i discorsi cristiani che si fanno sul peccato dovrebbero accettare l'accusa che viene da Nietzsche e dai suoi seguaci, cioè di essere finalizzati a screditare la vita stessa»⁸. Barsotti, parlando di una rivelazione cosmica di Dio, insiste sull'ascoltare il linguaggio divino insito nel creato, rifiutando il quale si rifiuta Dio e alla fine se stesso: «tolta la creazione Dio non ha rapporto con l'uomo, né l'uomo può stabilire un rapporto reale con Dio. Per questo nella *Leggenda del Grande Inquisitore* la parola diviene puro silenzio»⁹.

Come per altri percorsi della ricerca scientifica e tecnologica umana, non meraviglia se assistiamo ad atteggiamenti compresi tra il sospetto, la paura o il rifiuto e l'apertura, la curiosità o l'entusiasmo nei riguardi dell'ampio e complesso mondo della IA. Il cammino di quest'ultima è tanto complesso e intrigato quanto lo è quello della vicenda umana, perché il nodo drammatico non sta nelle cose ma dentro l'umano, che nondimeno perviene a sé mediante l'interazione con l'intero creato. Per l'uomo la ricerca di sé non è im-mediata, ma sempre intessuta di un travagliato e continuo discernimento rispetto a scelte e azioni storiche che vanno soppesate e non operate secondo la loro prima apparenza, laddove possono essere accolte senza comprendere le letali conseguenze che recano o essere bandite e persino demonizzate, precludendo o ritardando grandi benefici per l'umanità.

Non è minimamente ipotizzabile che l'atto della produzione di macchine di IA e il suo risultato siano comparabili con l'atto creativo per il quale esiste l'umano, con una dignità che in ultimo si fonda nella sua destinazione alla comunione con Dio. Ma è senza dubbio utile e doveroso che si scoprano aspetti e dinamiche delle creature, compresa quella umana, che ci sono ignoti, che ci potrebbero proiettare in modi di vivere anche molto diversi rispetto a quelli cui siamo abituati fino ad ora, con i rischi talvolta dolorosi che il cammino umano conosce e talvolta rimuove e dimentica. Rimane sempre per tutti quale supremo compito etico il

⁸ Pannenberg W., *Teologia sistematica 2*, cit.: 272.

⁹ Barsotti D., *Dostoevskij. La passione per Cristo*, Padova 2003:197.

vigile discernimento di quel che si annida e nasconde nel cuore umano e che si oggettiva nelle opere delle sue mani.

2. L'UMANESIMO TECNOLOGICO: IL CONTRIBUTO DELLA TEOLOGIA CRISTIANA E DELLA FILOSOFIA DELLA PERSONA

La prospettiva cristiana ci insegna a prendere le distanze da uno sviluppo incontrollato del progresso scientifico, a causa della finitezza umana. Il peccato "originale" è un tentativo "di diventare come Dio". Quello che è male, secondo la prospettiva cristiana è l'"arroganza umana" che riguarda il tema del potenziamento quando diventa un modo di "giocare a fare Dio", di diventare Creatore (non co-creatore) e sostituire Dio (negando la Sua esistenza). La prospettiva cristiana ci ricorda che non siamo Dio, né infiniti come Dio. Non possiamo sfuggire alla finitezza della creazione. Il tentativo finisce in una tragedia.

In questo senso, il transumanesimo attraverso il potenziamento funzionale umano cerca una sorta di "salvezza tecnologica" nella realizzazione materialistica dell'eterna immortalità sulla terra senza Dio, come nel progetto della Torre di Babele, una forma di *arroganza*, una ribellione contro la condizione umana, una ribellione contro il Creatore, allo scopo di dimostrare di essere migliore di Dio, tecnologicamente. È una prospettiva post-umanista, anti-cristiana o post-cristiana. Nella prospettiva cristiana lo scopo è la "salvezza teologica" spirituale, nell'accettazione della finitezza e della vulnerabilità come condizioni umane e come "dono" e nell'identificazione della "felicità" con la dimensione spirituale "al di là" del miglioramento materialistico e funzionalistico.

L'IA è progettata e prodotta da un essere umano: siamo noi che dobbiamo decidere come produrre l'IA. Non sappiamo cosa sarà fattualmente e tecnicamente possibile: ma sappiamo oggi cosa è umano e cosa vogliamo conservare di ciò che è umano in una prospettiva incentrata sull'uomo, ispirata da una visione teologica cristiana e ad una visione della filosofia della persona umana che ne possono segnare l'orizzonte valoriale di riferimento.

Stefano Zamagni

1. FIDUCIA E RESPONSABILITÀ NELLA ECONOMIA DI MERCATO DI FRONTE ALLE TECNOLOGIE CONVERGENTI

È sul fronte dell'etica pubblica che le conseguenze della diffusione nelle nostre società delle tecnologie convergenti vanno ponendo i problemi più delicati, primo fra tutti quello di capire come la digitalizzazione della nostra vita stia riuscendo a modificare anche il modo di percepirla. Eppure, è proprio su tale fronte che si registra una sorta di "fin de non recevoir" da parte dell'alta cultura, scientifica e filosofica. Di due (soli) aspetti particolari desidero qui dire in breve. Il primo concerne la questione della fiducia: può l'intelligenza artificiale creare la fiducia che è necessaria per il corretto funzionamento delle nostre economie di mercato? Il secondo aspetto chiama in causa il problema della responsabilità, di cosa significhi essere responsabili nell'era della digitalizzazione. Sono le "smart machines" agenti morali e dunque responsabili? Saranno gli algoritmi a governarci, in tutti i casi in cui le persone non sono in grado di comprendere appieno le questioni sulle quali debbono esprimere valutazioni? Comincio dal primo aspetto. Prima, però, una considerazione di natura generale. Non è saggio continuare ad ignorare di fare i conti con quel progetto, politico e filosofico insieme, che è il transumanesimo, termine, per primo, coniato da Julien Huxley per descrivere una realtà in cui si arriverà ad una completa ibridazione tra umano e macchina. L'ambizione dei transumanisti è sia fondere l'uomo con la macchina per ampliarne le potenzialità in modo indefinito sia (e soprattutto) arrivare a dimostrare che la coscienza non è un ente esclusivamente umano. Ecco perché si va parlando di coscienza artificiale con insistenza crescente, posto che la IA è data per qualcosa di acquisito. L'obiettivo non è tanto commerciale o finanzia-

rio: è politico, e in un certo senso religioso e ciò nel senso che il progetto ambisce a trasformare – non tanto a migliorare – il nostro modo di vivere, oltre che i nostri valori di riferimento. Il transumanesimo è l’apologia di un corpo e di un cervello umani “aumentati”, arricchiti cioè dall’intelligenza artificiale, il cui utilizzo consentirebbe di separare la mente dal corpo e quindi di affermare che il nostro cervello per funzionare non avrebbe necessità di avere un corpo.

Generale è il consenso sul fatto che è la fiducia uno dei fattori decisivi per assicurare i vantaggi dell’agire collettivo e, per questa via, sostenere il processo di sviluppo, ma si noti il paradosso tipico dell’attuale fase storica. Mentre la fiducia nei confronti delle istituzioni, sia politiche sia economiche, va declinando per una pluralità di ragioni, tra cui l’aumento endemico della corruzione, il mercato globale è conquistato sempre più da imprese e organizzazioni che chiedono ai loro clienti e utenti prove di fiducia, mai viste in passato. È come se gli individui stessero imparando la lezione della ben nota vicenda della Tosca di Puccini: la mutua sfiducia genera sempre risultati subottimali. Per Tim Wu, affermato giurista della Columbia University, quello cui stiamo assistendo è un massiccio trasferimento di fiducia sociale: abbandonata la fiducia nelle istituzioni, ci si rivolge alla tecnologia. “La fiducia – scrive R. Botsman – è la nuova valuta dell’economia mondiale. È un vero moltiplicatore di opportunità di guadagno perché consente di far fruttare beni sottoutilizzati”¹. Si pensi al fenomeno delle criptomonete – la più nota delle quali, ma non certo l’unica, è il bitcoin – che sono valute digitali che si scambiano tra pari. Le transazioni non sono garantite da alcuna autorità centrale, ma convalidate dagli stessi partecipanti alla rete mediante un algoritmo. Al tempo stesso, la forza di queste criptomonete è che esse consentono di effettuare transazioni anonime non soggette a tassazione e al riparo da confisca da parte dello Stato. L’infrastruttura che ne è alla base è la *blockchain*, che è un registro di proprietà distribuita su cui sono annotati tutti gli scambi, senza possibilità di modifica. La tecnologia *blockchain* – finora utilizzata praticamente solo in ambito finanziario – consente già oggi una vasta gamma di applicazioni, da quelle in ambito sociale a quelle di tipo politico-amministrativo.

¹ Botsman R., *Di chi possiamo fidarci*, Milano, 2017.

Si pensi alla gestione dei processi amministrativi, dove la *blockchain* può certificare in modo certo e per sempre un determinato atto senza bisogno di un soggetto Terzo certificatore. Si consideri anche che le Nazioni Unite stanno progettando di avvalersi della medesima tecnologia per la gestione degli aiuti, di varia natura, ai profughi e migranti. E così via.

Il cuore del paradosso odierno è in ciò che l'economia di mercato contemporanea ha ancora più bisogno di quella del passato di fiducia reciproca per poter funzionare al meglio. Al tempo stesso, però, gli straordinari livelli di efficienza finora raggiunti dai nostri sistemi economici sembrano far dimenticare che è necessario rinforzare le reti fiduciarie tra persone perché il mercato mentre "consuma" sempre più fiducia non riesce, stante l'attuale assetto istituzionale, a produrne a sufficienza. Di qui l'inquietante dilemma sociale: chiediamo sempre più efficienza per accrescere il benessere materiale, la ricchezza, la sicurezza, ma per conseguire un tale obiettivo decumuliamo irresponsabilmente il patrimonio di fiducia che abbiamo ereditato dalle generazioni passate.

2. POTENZIARE I CERVELLI UMANI

Che fare per sciogliere questo dilemma? È nota la proposta di David Hume. Per il fondatore dell'empirismo filosofico (e iniziatore del non cognitivismo etico) la disposizione ad accordare fiducia, e a ripagare la fiducia concessa, trova il proprio fondamento nei vantaggi personali che scaturiscono da una buona reputazione. «Possiamo soddisfare i nostri appetiti meglio in un modo indiretto e artificiale (...). È così che imparo a prestare un servizio ad un altro senza provare per lui una vera benevolenza. Infatti io prevedo che egli mi renderà il servizio attendendosene un altro dello stesso tipo, per conservare la medesima *reciprocità* di buoni uffici con me o con altri»². È quasi incredibile che un grande filosofo come Hume sia potuto cadere in una così patente svista concettuale, quella di confondere la reciprocità con una sequenza di scambi auto-interessati. La reciprocità, a differenza dello scambio di equivalenti, è un insieme di relazioni di dono (si badi, non di donazioni) tra loro interrelate.

² Hume D., *Trattato sulla natura umana*, Roma-Bari, [1740] 1971: 552-3.

Anche la soluzione dell'imperativo categorico kantiano non ci è di grande aiuto agli scopi presenti. "Segui la regola che, se ognuno la seguisse, tu potresti volerne il risultato". È questo un principio di eguaglianza del dovere. Tuttavia, la teoria di Kant soffre di una evidente aporia quando si cercasse di porla in pratica. Infatti, l'individuo kantiano sceglie la regola (la massima) che va ad applicare assumendo che anche tutti gli altri la applichino. Ma poiché persone diverse, in generale, hanno preferenze diverse circa il risultato finale, anche le regole kantiane da esse preferite saranno a priori diverse. Ne consegue che ciascuno seguirà la sua regola preferita (da cui la sua azione) assumendo che gli altri agiscano nel modo in cui in realtà essi non agiranno affatto. Ciò significa che il principio kantiano non può applicarsi a se stesso; non può validare se stesso: davvero una seria incongruenza logica per una dottrina morale che ambisce ad essere universale. Solamente se tutti gli individui fossero tra loro identici quanto al loro sistema preferenziale l'aporia in questione scomparirebbe. Ma è evidente che se così fosse il principio kantiano perderebbe tutta la sua rilevanza pratica.

La recente ricerca nell'ambito delle neuroscienze va oggi suggerendo la seguente via d'uscita dal dilemma sopra indicato. Lavori recenti dimostrano che se si disattiva, mediante stimolazione magnetica transcranica, una particolare zona della corteccia cerebrale, i soggetti aumentano notevolmente il loro comportamento prosociale, il che conduce ad un sostanziale incremento del loro grado di fiducia. In particolare, somministrando per via nasale una certa quantità di ossitocina (un ormone naturalmente prodotto dall'organismo di molti mammiferi) si è scoperto che esso deattiva l'attività cerebrale di una specifica regione del cervello (l'amigdala) deputata a controllare il comportamento degli individui nei rapporti fiduciari³. Si pensi anche agli interventi volti al potenziamento cognitivo che agiscono su capacità come l'attenzione, la memoria, la tendenza all'affaticamento intellettuale. Già vengono praticate tecniche come la stimolazione celebrale profonda (*deep brain stimulation*) che prevede l'impianto di un microchip nel cervello; come la stimolazione transcranica a corrente diretta (*transcranical direct cur-*

³ Narvaez D., *Neurobiology and the Development of Human Morality*, New York, 2014.

rent stimulation) che prevede la stimolazione dell'encefalo con dosi di corrente elettrica.

Pochi anni fa, ricercatori dell'Università di Berkeley hanno testato su un campione di trentacinque soggetti un farmaco "in grado di produrre artificialmente sentimenti di bontà e di benevolenza verso gli altri"⁴. I risultati ottenuti confermerebbero che il tolcapone, altro ormone umano, contribuisce ad accrescere il tasso di equanimità nei confronti anche di sconosciuti e ad accrescere, per questa via, la propensione alla fiducia. (Si tratta di tentativi che mirano al *mood enhancement* delle persone, per modificarne il carattere e aumentarne il benessere psicologico, contrastando la disposizione alla tristezza e all'introversione). Non è questa la sede per discutere della plausibilità di risultati simili e per giudicare l'efficacia, nella pratica, di proposte come quella di somministrare per via chimica molecole atte a potenziare la nostra moralità. Mi limito ad osservare che il tentativo di attribuire l'origine del senso morale alla biologia, tentativo che riduce tale senso a mera chimica cerebrale, se da un lato può sortire effetti desiderati rispetto a ciò che è funzionale al buon andamento degli affari, dall'altro riduce lo spazio della libertà (positiva) e quindi della responsabilità individuale. Vedere il pensiero morale come intrinseco al cervello umano, piuttosto che come prodotto di volontà e di virtù, comporterebbe un pericoloso arretramento rispetto al nostro modello di civilizzazione.

3. POTENZIAMENTO OLTRE L'UOMO

Non v'è chi non veda come approcci del genere si collochino, al di là delle apparenze o delle dichiarazioni ufficiali, nell'ambito di quel grande progetto, politico e filosofico insieme, che è il transumanesimo, la cui ambizione è sia fondere l'uomo con la macchina per ampliarne le potenzialità in modo indefinito sia (e soprattutto) arrivare a dimostrare che la coscienza non è un ente esclusivamente umano.

La strategia perseguita da Ray Kurzweil, responsabile del progetto che Google va da qualche tempo implementando, mira alla produzione

⁴ Sáez I., Zhu L., Set E., Kayser A., Hsu M., *Dopamine Modulates Egalitarian Behavior in Humans*, *Current Biology*, 25, 2015:912-919.

di cyborg dotati di sembianze fisiche e capacità cognitive simili a quelle dell'*homo sapiens*. È l'obiettivo del *playing God* (recitare la parte di Dio) che nasconde il desiderio di prendere in mano le redini dell'evoluzione. L'approccio fiscalista (secondo cui esisterebbe soltanto una realtà – quella fisica – che le scienze cognitive cercano di comprendere per spiegare come si genera la conoscenza), accolto dalle neuroscienze pone in discussione il nesso tra responsabilità e libertà. Veniamo da una lunga stagione durante la quale era assodato ritenere che alla libertà come espressione della responsabilità corrispondesse la responsabilità come consenso all'applicazione della stessa libertà. Cosa significa per un operatore lavorare tutto il giorno con un robot collaborativo? Sappiamo già come l'avvento dei social network e l'uso degli smartphone stiano cambiando le nostre abitudini e i nostri stili di vita. Ma possiamo pensare un futuro in cui l'uomo trascorre tutta la sua giornata lavorativa "dialogando" – si fa per dire – con un robot, senza cadere in forme nuove e più gravi di alienazione? Si tenga presente la grossa differenza tra automazione e IA: mentre la prima interviene sull'oggetto – come nel caso di *Internet of Things* – la seconda agisce sul soggetto.

4. LA RESPONSABILITÀ NELL'ERA DELLE MACCHINE

Passo ora al tema della responsabilità. Come sappiamo, la responsabilità possiede significati diversi. Si può dire responsabilità per significare una libertà che possiede il senso della responsabilità. Ma si può dire responsabilità in senso molto diverso quando si è incaricati di un compito di cui si deve rispondere. (È il concetto americano di "*accountability*"). Infine, si può dire responsabilità per indicare che si è colpevoli di un'attività portata a compimento. In tal senso, "io sono responsabile" significa che sono colpevole di qualcosa. Responsabilità e libertà risultano pertanto fortemente correlate, anche se, in tempi recenti, sull'onda degli avanzamenti registrati sul fronte delle neuroscienze, si tende ad allentare il nesso tra libertà e responsabilità. Si considerino gli interventi di potenziamento sull'uomo. Il soggetto potenziato prenderebbe le sue decisioni non sulle ragioni pro e contro, ma in seguito all'influsso causale esercitato sul suo cervello dai mezzi di manipolazione biotecnologica. Quanto a dire che per

migliorare la performance degli esseri umani li si priva della loro autonomia morale, che è il bene più prezioso.

Mentre sembra relativamente facile identificare la responsabilità diretta degli agenti – come quando il proprietario di uno *sweatshop* sfrutta il lavoro minorile per trarne un vantaggio – che dire dell'azione economica che è intrapresa con l'intenzione di non svantaggiare nessuno e tuttavia provoca effetti negativi in capo ad altri? Ad esempio, di chi è la responsabilità della disoccupazione, della povertà, delle disuguaglianze, etc.? Le risposte tradizionali, in economia, consistono nel sostenere che si tratta di conseguenze non volute delle azioni intenzionali (le “*unintended consequences of intentional actions*” di cui ha parlato la Scuola dei moralisti scozzesi del XVIII secolo). E dunque l'unica cosa da fare è di attribuire alla società il compito di porre rimedio (o di alleviare) le conseguenze negative. E infatti il *welfare state* è sorto e si è sviluppato precisamente per rendere collettiva e impersonale la responsabilità dei singoli. Ma è veramente così? Siamo sicuri che i meccanismi del libero mercato siano inevitabili e che gli effetti che ne derivano siano inattesi, come si tende a far credere?⁵

Valga un solo esempio. Albert Carr nel saggio *Is business bluffing ethical?* – il saggio più citato di sempre in teoria della finanza – scrive: «La finanza, dove si cerca di fare agli altri ciò che non si desidera gli altri facciano a noi (*sic!*) dovrebbe essere guidata da un insieme di standard etici diversi da quelli della morale comune o della religione: gli standard etici del gioco. Se un'azione non è strettamente illegale, e può dare un profitto, allora compierla è un *obbligo* dell'uomo d'affari»⁶. È questo modo di pensare – fondato sulla tesi della doppia moralità – che è all'origine di tutti i grandi scandali finanziari, tra cui quelli dell'ultimo ventennio. Come, tra i primi, aveva notato Zygmunt Bauman, l'organizzazione sociale della seconda modernità è stata pensata e disegnata per neutralizzare la responsabilità diretta e indiretta degli agenti. La strategia adottata – di grande raffinatezza intellettuale – è stata quella, per un verso, di allungare la distanza (spaziale e temporale) tra l'azione e le sue conseguenze e, per l'altro verso, di realizzare una grossa concentrazione di attività economica senza

⁵ Si veda la *Caritas in Veritate* di Benedetto XVI, cap. 3 e 4.

⁶ Carr A., *Is business bluffing ethical?*, Harvard Business Review, 1968.

una centralizzazione di potere. È in ciò il carattere specifico dell'impresa adiaforica, una figura di impresa ignota alle epoche precedenti la Seconda guerra mondiale e il cui fine è quello di annullare la questione della responsabilità morale dell'azione organizzata. Adiaforica è la responsabilità "tecnica" che non può essere giudicata in termini morali di bene/male. L'azione adiaforica va valutata in termini solamente funzionali, sulla base del principio che tutto ciò che è possibile per gli agenti sia anche eticamente lecito, senza che si possa giudicare eticamente il sistema, come Luhmann ha insegnato.

Ebbene, la responsabilità adiaforica ha ricevuto, in tempi recenti, nuovo impulso proprio dalla Intelligenza Artificiale, la quale va producendo "mezzi" che sono alla ricerca di "domande" o di problemi da risolvere. Esattamente il contrario di quanto era accaduto con le precedenti rivoluzioni industriali. Invero, cosa ne è del principio di responsabilità nella società degli algoritmi? Dalle nuove tecnologie industriali alla diagnostica medica, dai social networks ai voli degli aerei, dai big data ai motori di ricerca: ci affidiamo a complesse procedure cui deleghiamo la buona riuscita di operazioni che gli esseri umani, da soli, non saprebbero eseguire. Eppure, gli algoritmi sono irresponsabili, pur non essendo neutrali, né oggettivi, come invece erroneamente si crede. Quando un programma commette un errore non ne paga le conseguenze, perché si ritiene che la matematica resti al di fuori della moralità. Ma non è così, perché gli algoritmi non sono pura matematica; sono opinioni umane incastonate in linguaggio matematico. E dunque discriminano, al pari dei decisori umani. Ad esempio, il processo delle assunzioni di lavoro si va sempre più automatizzando, perché si pensa di rendere obiettivo il reclutamento del personale, eliminando pregiudizi. Ma le dinamiche discriminatorie, anziché diminuire, stanno aumentando nelle nostre società.

5. L'ETICA DELLE MACCHINE

Generalizzando un istante, il vero problema delle *smart machines* comincia nel momento in cui queste compiono azioni che coinvolgono la necessità di scegliere oppure di decidere. Il soldato-robot, l'automobile-robot, la scopa-robot potrebbero compiere scelte esiziali per vite non robotiche.

Di chi è la responsabilità in questi casi? Quale il messaggio subliminale della recente provocazione di Bill Gates di tassare i robot, che andrebbero dotati di personalità elettronica, al pari delle *corporations* che sono dotate di personalità giuridica? Come ha lucidamente spiegato Gunther Anders, il XXI secolo ha inaugurato l'era dell'irresponsabilità umana, immunizzando i soggetti dalle loro relazioni. Le *smart machines* (quelle dotate di intelligenza artificiale) sono in grado di prendere decisioni autonome, che hanno implicazioni sia sociali sia morali. (Si veda il caso dell'auto senza pilota Tesla, creata da Elon Musk, che nel maggio 2016 uccise un passeggero). Come assicurare, allora, che le decisioni prese da tali oggetti siano eticamente accettabili? Dato che queste macchine possono causare danni di ogni sorta, come fare in modo che esse siano poste in grado di differenziare tra decisioni corrette" e "sbagliate"? E nel caso in cui un qualche danno non possa essere evitato - si pensi al caso dell'auto senza conducente che deve scegliere se gettarsi contro un altro veicolo uccidendone i passeggeri oppure investire dei bambini che attraversano la strada -, come istruire (nel senso di programmare) tali macchine a scegliere il danno minore? Gli esempi in letteratura sono ormai schiera. E tutti concordano sulla necessità di dotare l'IA di un qualche canone etico, per sciogliere dilemmi morali del tipo "guida autonoma".

Le divergenze nascono nel momento in cui si deve scegliere il modo (cioè l'approccio) secondo cui procedere: *top-down* (i principi etici sono programmati nella macchina intelligente: l'uomo trasferisce all'intelligenza artificiale la sua visione etica del mondo) oppure *bottom-up* (la macchina impara a prendere decisioni eticamente sensibili dall'osservazione del comportamento umano in situazioni reali). Entrambi gli approcci pongono problemi seri, che non sono tanto di natura tecnica quanto piuttosto concernono la grossa questione se le macchine intelligenti debbano o meno essere considerate agenti morali (cioè *moral machines*). Siamo appena agli inizi di un dibattito culturale e scientifico che già si preannuncia affascinante e preoccupante ad un tempo. Si veda, ad esempio, la recente presa di posizione di Amitai Etzioni e Oren Etzioni⁷ che negano la possi-

⁷ Etzioni A., Etzioni O., *Incorporating Ethics into Artificial Intelligence*, *Journal of Ethics*, March, 21, 2017: 403-418.

bilità di attribuire lo status di agente morale all'IA e dunque negano ogni fondamento al programma di ricerca della *Internet Ethics* che studia gli aspetti etici della *Internet Communication* nelle sue varie articolazioni.

Da ciò essi traggono la conclusione che non vi sarebbe alcun bisogno di insegnare etica alle macchine, anche se ciò potesse essere fattibile. Non è del medesimo parere, ad esempio, il gruppo di ricerca che opera per la *Neurolink Corporation*, in California, che da qualche tempo sta sviluppando tecnologie digitali per realizzare connessioni tra computer e mente umana e che sta progettando un uomo-cyber con microchip nel cervello. È la intricata e delicata questione concernente la possibilità di attribuire "personalità elettronica" ai robot intelligenti e, più in generale, la opportunità di favorire il passaggio dalla selezione naturale darwiniana alla scelta deliberata del processo di selezione mediante la scorciatoia biotecnologica⁸.

Per concludere, il problema serio che ci sta di fronte è che l'IA costituisce un divario tra agire e intelligenza. Invero, l'IA è una trasformazione che ha a che fare non tanto con l'intelligenza, quanto piuttosto con l'agire. Il *machine learning*, il *natural language processing*, la robotica sono straordinarie soluzioni per far svolgere alla macchina compiti propri della persona umana, in maniera più efficace, ma senza intelligenza. Si sta separando la capacità di risolvere problemi dalla necessità di essere intelligenti nel farlo. Questa è la grande e rischiosa novità dell'oggi. Mancando l'intelligenza, non c'è intenzionalità, senso, eticità dell'agire. La grande minaccia è che l'IA si trascini dietro un'etica artificiale.

È giunto il tempo di mettersi a pensare seriamente, e in modo sistematico la *governance* del digitale. Ma per questo, ci vuole un'idea di progetto umano, occorre aver chiaro dove si vuole andare.

⁸ Cfr. Kampowski S.M., Moltisanti D. (a cura di), *Migliorare l'uomo? La sfida dell'enhancement*, Siena, 2011.

ALLA RICERCA DI UN *HABEAS CORPUS* PER L'INTELLIGENZA ARTIFICIALE

Luca Antonini e Antonella Sciarrone Alibrandi

1. LE RIFLESSIONI DEL CORTILE DEI GENTILI E IL LORO RILIEVO NEL DIBATTITO GIURIDICO

Il sempre più diffuso affermarsi dei sistemi di intelligenza artificiale (di seguito: IA), la prospettiva di ulteriori, enormi potenzialità di sviluppo e, al contempo, la rilevanza dei rischi ad essi correlati sollevano serie questioni di natura giuridica. Non ci troviamo, infatti, semplicemente davanti, come tante volte è accaduto nella storia umana, a una nuova tecnologia volta a migliorare o ampliare le nostre capacità, ma a sistemi complessi che possono arrivare via via a sostituirci in decisioni che sempre ci eravamo riservati, sino al punto da porre in discussione lo stesso nucleo intangibile dell'umano.

Di fronte a una sfida così radicale, cresce l'esigenza di disegnare un quadro regolatorio che, senza porre eccessivi (e ingiustificati) freni all'innovazione e alla ricerca, persegua però in modo adeguato la tutela dei diritti fondamentali e sappia garantire uno sviluppo realmente finalizzato al bene dell'umanità. È questo, peraltro, l'approccio fatto proprio, circa un anno fa, dalla Commissione europea, che, per prima a livello mondiale, ha messo in cantiere una giustamente ambiziosa proposta di Regolamento sull'*Artificial intelligence* (AI Act), il cui *iter* per l'approvazione è ancora piuttosto lungo ma finalizzato all'introduzione di regole armonizzate vincolanti.

Gli interessi in gioco, nel loro rimandare a questioni ultime di natura antropologica, sono tali da andare ben oltre la tecnica giuridica, esigendo scelte normative di pregnante valore politico. Scelte che, a loro volta, come forse mai prima d'ora, impongono ai giuristi chiamati ad assumerle di abbeverarsi alla fonte di numerosi, e fra loro eterogenei,

saperi: dalla filosofia alle neuroscienze, dalla medicina alla teologia e molti altri ancora.

Al fine di contribuire allo sviluppo di un ampio dibattito in chiave transdisciplinare sul tema, fra i componenti della Consulta scientifica del Cortile dei Gentili ha iniziato a svolgersi un intenso e prolungato confronto, guidato da Giuliano Amato. Di tale confronto si intende dare conto qui di seguito (in via di sintesi pur senza rinunciare a un riferimento puntuale alle varie opinioni emerse), per poi proporre alcune considerazioni focalizzate su un punto di caduta più strettamente giuridico.

2. IL CUORE DEL PROBLEMA

Sebbene la narrazione intorno all'IA possa rischiare di assumere i tratti di una costruzione mitologica, le questioni interrogano, innanzitutto, su cosa si possa ancora definire a tutti gli effetti "esperienza umana". Infatti, trasladosi sempre di più le percezioni esperienziali nel mondo digitale, si assiste a una sostanziale dismissione nel virtuale del reale, che rischia di condurre alla dissoluzione di una delle categorie fondamentali dell'essere umano.

Del resto, il transumanesimo è l'apologia di un corpo e di un cervello umani "aumentati", potenziati cioè dall'intelligenza artificiale, il cui utilizzo consentirebbe di separare la mente dal corpo e quindi di affermare che il nostro cervello, per funzionare, non avrebbe necessità di avere un corpo. Per questa via, si sviluppano argomenti in grado di incidere sul significato della persona e della sua unità, svelando così la reale ambizione del transumanesimo: l'essere un progetto diretto a trasformare, non tanto a migliorare, il nostro modo di vivere e i nostri valori di riferimento.

Appare quindi indispensabile rivalutare la libertà umana, perché altrimenti, perdendola di vista, si potrebbe scendere fino a sostenere che le decisioni degli esseri umani siano esse stesse il prodotto di un algoritmo dal funzionamento biochimico su una base neuronale, in quanto tali assimilabili al prodotto di un algoritmo elettrico su una piattaforma formata da semiconduttori.

Si tratta di un rischio ben attuale, che identifica come priorità quella di rifondare un umanesimo tecnologico, che eviti da un lato di “antropomorfizzare” la macchina, umanizzando quindi la tecnologia, o, dall’altro, di artificializzare l’uomo a “macchina biologica”, tecnologizzando cioè l’uomo.

Pertanto, nel ribadire quel legame tra corpo e mente definibile anche come il “colore dell’intelligenza”, non resta che riconoscere come l’esigenza, etica e scientifica, di controllare lo sviluppo tecnologico nel campo dell’IA costituisca una delle grandi sfide del nostro tempo.

3. LA NECESSITÀ DI UNA SOSTENIBILITÀ DIGITALE

Si profila, quindi, la necessità di dare adeguato rilievo alla “sostenibilità digitale”, in modo da non mettere la capacità tecnica al centro dell’attenzione, bensì mantenervi l’uomo, inteso come fine che qualifica il progresso.

Pertanto, va considerato che lo stesso progresso tecnologico non è neutrale ma è guidato da imprese che hanno come obiettivo la massimizzazione del loro profitto, talvolta (spesso) senza prestare particolare attenzione ai risvolti di sostenibilità – ambientale, sociale ed umana – della loro azione.

Pertanto, anche al fine di contenere gli effetti perversi del c.d. “capitalismo di sorveglianza”, sarebbe opportuno non lasciare lo sviluppo incontrollato dei sistemi di IA alle piattaforme digitali dei privati, bensì introdurre adeguati strumenti di governo delle decisioni private nonché coinvolgere in misura maggiore nei processi attori di natura pubblica.

E, in questa prospettiva, apprezzabilmente si pone la *Raccomandazione sull’etica dell’IA* approvata dalla Conferenza generale dell’Unesco nel novembre 2021, che introduce il concetto di IA *actors* (soggetti operanti nel campo dell’IA), comprensivo di enti pubblici e privati. Certo, la *Raccomandazione* Unesco ha un peso normativamente leggero, se misurato sulla scala della vincolatività, ma molto ampio, considerata la sua vocazione multidisciplinare e globale ed avendo come sfondo i diritti umani delle dichiarazioni universali.

4. L'EVOLUZIONE DELLA IA E I DILEMMI IMPLICATI

Nell'economia del discorso non può essere sottovalutata l'evoluzione che ha caratterizzato lo sviluppo dei sistemi di IA e, in particolare, il cambio di paradigma verificatosi rispetto ai sistemi esperti, quando, all'inizio del nuovo millennio, si è cominciato ad avere facile accesso a enormi volumi di dati. L'approccio, infatti, è diventato di tipo induttivo: si è passati dal codificare regole per i sistemi esperti, al farle scoprire ai computer in via autonoma per correlazione e classificazione, sulla base di un'enorme mole di dati.

In altri termini, gli attuali sistemi non si basano più, come era agli albori dell'IA, sul tentativo di riprodurre la capacità teorica e di ragionamento dell'intelligenza umana, bensì, in gran parte, su un approccio di tipo "statistico".

Questo è un punto chiave anche per comprendere i dilemmi implicati nell'utilizzazione giuridica di sistemi di IA e che di seguito si proverà ad esplicitare.

A titolo di esempio si pensi al sistema *Compas*, utilizzato dai giudici in Wisconsin per valutare il rischio di recidiva, che si è osservato finire per discriminare i neri, non in quanto nella costruzione dell'algoritmo sia stato volutamente immesso un fattore discriminante, ma perché, nella realtà, i casi di recidiva riguardano i neri in misura maggiore. Il problema allora non è la dissimulata intenzione discriminatoria di chi ha costruito l'algoritmo, ma il modo con cui lo *statistical machine learning* ha reagito su quei dati determinando un risultato iniquo.

Bisogna anche dire però, in positivo, che un sistema di IA, così come strutturato su un approccio statistico, ha potuto apportare un contributo importante al miglioramento delle condizioni di vita.

Sul piano del quotidiano modo di vivere, si deve constatare che esso è determinante per la possibilità di usare la posta elettronica (ormai divenuta di gran lunga prevalente su quella cartacea), giacché, con tutta probabilità, essa non sarebbe usufruibile senza il filtro operato dalla IA su migliaia di mail di posta indesiderata.

Su piani più specifici, come ad esempio in campo medico, è innegabile che, attraverso la IA, si possano analizzare enormi volumi di casi, con

una capacità quantitativa di analisi irraggiungibile dalla pura e semplice attività umana. Per cui, banalizzando, una persona potrebbe preferire farsi curare secondo le terapie suggerite dall'IA rinvenibili nei migliori ospedali del mondo, rispetto a quelle suggerite dai medici di un ospedale di scadente livello, con alta percentuale di mortalità.

Ciò considerato, resta tuttavia sempre aperto il problema di come mantenere un equilibrio fra le *soft skills* umane dei sanitari (astrazione, intuizione, flessibilità, empatia) e le *hard skills* dell'IA (enormi capacità di calcolo e gestione dei database). Così come altrettanto aperta rimane un'ulteriore questione sull'uso che può essere fatto dalle industrie farmaceutiche e in generale dai soggetti privati della enorme mole di dati sensibili di cui si devono alimentare i sistemi di IA

Si deve quindi concludere che quello delle opportunità fornite dalla IA è solo uno dei lati della medaglia, perché vi sono, ormai, costanti algoritmiche che stanno generando in via automatica regole, addirittura decisioni, alle quali coloro che altrimenti dovrebbero decidere sono abilitati a rimettersi. E ciò sia che si tratti di stabilire se una persona deve essere assoggettata a carcerazione preventiva, o se un incidente debba essere risarcito, o se una terapia vada preferita a un'altra, o ancora se in consiglio d'amministrazione una data delibera debba essere approvata oppure no.

Tutto ciò, peraltro, con un'ulteriore complicazione: mentre il progettatore di un programma *software* può dirsi pienamente responsabile di quanto ha prodotto, un'analogia responsabilità non è infatti configurabile in capo a chi progetta l'algoritmo di una IA, perché questa ha una autonoma capacità di apprendimento, che, proprio in virtù della sua autonomia, pone in dubbio il sorgere di una imputabilità, perlomeno piena, in capo a chi l'ha progettata.

5. ALCUNE DOMANDE DA PORRE ALL'INTERNO DEL DIBATTITO, NELL'OT- TICA DI UN'ETICA PROCEDURALE

I problemi aperti dalla IA in ambito giuridico sono quasi infiniti, ma è forse possibile operare delle semplificazioni, se ci si pone nei termini di un'etica procedurale, più che altro votata a svelare con specifiche domande i punti di maggiore criticità.

Un primo snodo essenziale è il seguente: si può convenire sul fatto che la dignità della persona implica sempre, in sede giurisdizionale, il diritto a essere giudicato da un essere umano (è in fondo la garanzia dell'*habeas corpus*, sulla quale si tornerà tra breve) e non da sistemi di IA, ritenendo che questi possano essere utilizzati solo nell'ambito amministrativo della giustizia? E ciò anche perché lo *statistical machine learning* basato sui precedenti esclude la possibilità dell'*overruling*, che invece è una possibilità fondamentale nella funzione giurisdizionale.

In modo analogo, particolare attenzione va prestata a tutte le situazioni in cui dignità umana e principio di non discriminazione vengono fortemente in rilievo: in ognuno di tali casi, il dilemma circa l'uso della IA difficilmente sembrerebbe poter essere sciolto se non ritenendone preclusa l'applicazione.

È questo, ad esempio, il caso del sistema *iBorderCtrl* utilizzato dalla polizia di frontiera di alcuni Stati per individuare i migranti che non hanno diritto di asilo e che funziona sulla base dell'analisi delle espressioni del volto, per capire la veridicità delle dichiarazioni. Va precisato che si tratta di un sistema *imposto* ai migranti, il che rende questa applicazione della IA ben diversa dalla facoltà che, invece, viene *concessa* da alcune giurisdizioni americane di sottoporsi volontariamente alla cd. "macchina della verità".

In altri settori, dove invece è implicata la libertà di impresa o il diritto al lavoro, si potrebbe valutare come meno pressante l'esigenza di tutela di diritti fondamentali e ritenere che l'IA possa essere utilizzata nell'ottica di una decisione che dovrebbe però, in ogni caso, rimanere assoggettabile a una verifica umana: ad esempio, tecniche decisorie fondate su elementi tratti da sistemi di IA potrebbero essere usate in fase di valutazione del merito creditizio qualora non sia eliminato in radice lo spazio per un intervento, anche *ex post*, di natura umana. Lo stesso potrebbe valere per la selezione dei lavoratori.

In questa seconda tipologia di casi, il punto cruciale è quello relativo alla responsabilità umana, che dovrebbe essere salvaguardata come ultima istanza di imputabilità. Ma come si può mantenere una responsabilità umana di ultima istanza in una decisione, presa attraverso l'ausilio della IA, che comunque inerisce a diritti della persona come quelli al lavoro e all'iniziativa economica?

Una volta assegnata all'IA valenza di *Non-Binding Opinion*, sarà, ad esempio, importante prestare attenzione alla disciplina delle responsabilità legali in capo all'uomo che da tale *Non Binding Opinion* si discosti; ad esempio, se venisse imposta l'inversione dell'onere della prova, il *Non Binding* tenderebbe a diventare nei fatti *Binding*.

L'appena suggerita distinzione fra tipologie di casi d'uso dell'IA in funzione del differente rilievo che in essi assume la dignità umana e il principio di non discriminazione riporta, tuttavia, all'ulteriore e assai delicato problema di verificare la solidità del criterio distintivo fra ciò che è vietato e ciò che è consentito: una verifica da compiersi necessariamente alla stregua dell'adeguatezza del criterio rispetto alla tutela degli interessi di volta in volta in gioco.

Si aprono allora, in quest'ultima prospettiva, ulteriori domande: il rischio di violazioni del principio di non discriminazione può portare a ritenere sempre non lecita la sottoposizione a sistemi di IA per decidere il riconoscimento o meno di diritti fondamentali concernenti soggetti deboli e vulnerabili? E, ancora, l'utilizzo di sistemi di IA con riferimento a diritti attinenti alla libertà di iniziativa economica e, più in generale, aventi a che fare con l'utilizzazione di servizi finanziari in senso lato (bancari, d'investimento, assicurativi) può sempre dirsi effettivamente non pericoloso dal punto di vista della tutela di diritti fondamentali della persona?

6. L'INSUFFICIENZA DELLA *SOFT LAW* E L'OPPORTUNITÀ DI UN INTERVENTO DI *HARD LAW*

Come più sopra già accennato, una prima risposta ad alcune di queste domande si sta costruendo a livello europeo, dove, dopo alcune iniziative preparatorie dirette a comprendere e inquadrare l'IA, ci si è distaccati dall'approccio americano, più votato alla *soft law*, per imboccare la strada dell'*hard law*.

Più in dettaglio, la Commissione ha approvato un *Libro Bianco* a febbraio 2020, ove si evidenzia la necessità di una regolamentazione europea dell'IA, cui hanno fatto seguito, nell'ottobre del medesimo anno, tre risoluzioni del Parlamento europeo in tema di IA, nello

specifico riguardanti: a) la questione etica, b) la responsabilità e c) la proprietà intellettuale. Tali risoluzioni hanno messo in luce che le regolamentazioni dell'IA devono essere indirizzate ad assicurare sicurezza, trasparenza, assunzione di responsabilità, nonché a evitare la creazione di pregiudizi e di discriminazioni, a stimolare la responsabilità sociale e ambientale, ad assicurare il rispetto dei diritti fondamentali.

La prima risoluzione ha sottolineato, soprattutto, l'opportunità di mettere al centro della regolamentazione la persona, secondo un approccio qualificabile come "antropocentrico"; la seconda ha proposto un regime differenziato basato sulla pericolosità dei sistemi di IA, suddividendo quindi questi ultimi in sistemi "ad alto rischio", sistemi "non ad alto rischio" e in "pratiche di IA vietate", mettendo comunque al centro la responsabilità e la sorveglianza umana.

Sulla scia di tali lavori preparatori, si è giunti quindi a presentare, il 21 aprile 2021, una proposta di regolamento volta a fornire un quadro armonizzato di regole vincolanti per lo sviluppo, l'immissione sul mercato e l'utilizzo di sistemi IA nell'Unione conforme «ai valori, ai diritti fondamentali e ai principi dell'Unione».

7. LUCI E OMBRE DELLA PROPOSTA EUROPEA

Al fine di elaborare requisiti minimi necessari per affrontare i problemi connessi all'intelligenza artificiale, la proposta, condivisibilmente, adotta un approccio basato sul rischio e di tipo orizzontale, ovvero riferibile a ogni settore del mercato unico.

È interessante notare che la proposta classifica i prodotti che utilizzano completamente o parzialmente l'IA in base al rischio di impatto negativo su diritti fondamentali quali la dignità umana, la libertà, l'uguaglianza, la democrazia, il diritto alla non discriminazione, la protezione dei dati e, in particolare, la salute e la sicurezza. Più il prodotto è suscettibile di mettere in pericolo questi diritti, più severe sono le misure adottate per eliminare o mitigare l'impatto negativo sui diritti fondamentali, fino a vietare, come si evince dal *Considerando* n. 41, quei prodotti che si ritengono eccessivamente rischiosi rispetto a tali diritti.

Ad una prima analisi, il *Regolamento* presenta aspetti senz'altro positivi, anche se forse gli interessi del mercato sembrano in alcuni passaggi sbilanciati a danno di una più rigorosa tutela dei diritti fondamentali e della dignità umana. Nella relazione alla proposta emerge, infatti, con insistenza l'accento inerente la finalità di garantire «il buon funzionamento del mercato interno per i sistemi di intelligenza artificiale (“sistemi di IA”), nel contesto del quale tanto i benefici quanto i rischi legati all'intelligenza artificiale siano adeguatamente affrontati a livello dell'Unione. Essa contribuisce all'obiettivo dell'Unione di essere un leader mondiale nello sviluppo di un'intelligenza artificiale sicura, affidabile ed etica, come dichiarato dal Consiglio europeo, e garantisce la tutela dei principi etici, come richiesto specificamente dal Parlamento europeo».

Alla luce di tale impostazione, emerge quindi un dubbio di fondo: pur apprezzando l'approccio antropocentrico della Commissione, può dirsi adeguato il bilanciamento tra esigenze di sviluppo del mercato della IA e la tutela dei diritti fondamentali? Può, ad esempio, ritenersi sufficiente che la proposta si limiti a classificare ad alto rischio l'uso di alcuni sistemi di IA, come quelli in materia di giustizia o immigrazione, o tali sistemi dovrebbero piuttosto essere vietati in radice sulla base delle esigenze di tutela dei diritti fondamentali sanciti a livello europeo? E, ancora, è stata prestata sufficiente attenzione ai rischi per la persona (esclusione sociale derivante da quella finanziaria, giusto per fare un esempio) generati dall'utilizzo di IA nel settore finanziario?

8. UN'HABEAS CORPUS SULLA IA

In conclusione di questo percorso, drasticamente sintetico rispetto alla vastità dei temi toccati, non pare inutile evocare un atto che ha assunto, nel mondo, il valore di un simbolo delle conquiste della civiltà giuridica.

Si tratta dell'*Habeas Corpus*, principio coevo alla Magna Carta del 1215, e allora diretto a scongiurare l'abuso dei poteri di polizia e gli arresti illegali.

Nel suo significato letterale “che tu abbia il corpo”, la solennità di quel simbolo, che lo pone all'inizio della storia del costituzionalismo, po-

trebbe essere oggi rievocata per affermare, di fronte alle sfide della IA e delle sue implicazioni, l'opportunità di rimanere ancorati alla dimensione antropocentrica ed evitare, come di recente ha scritto Gianfranco Ravasi, che si tagli con un colpo di spada tecnologico il nodo che vincola il soggetto umano all'avere un corpo e all'essere un corpo.

LE FUNZIONI DEGLI ALGORITMI E IL DISCERNIMENTO UMANO

Leonardo Becchetti

1. L'UOMO, IL PROGRESSO TECNOLOGICO E L'IA: L'APPROCCIO TAM QUAM

Di fronte all'ondata impetuosa del progresso tecnologico, accelerata dall'aumento di velocità di circolazione delle conoscenze in rete, l'atteggiamento a mio avviso più corretto da adottare è quello suggerito dalla spiritualità di S. Ignazio che usava l'espressione "tam quam", ovvero suggeriva di usare dei beni tanto quanto essi possono contribuire alla nostra realizzazione. Il progresso delle tecnologie aumenta e potenzia senza dubbio le capacità dell'umano ma questo accade da sempre, sin da quando abbiamo inventato il fuoco, la ruota e gli occhiali. Cosa c'è di diverso dunque oggi? Senz'altro il fatto che nella crescente perfezione delle macchine emerge l'ambizione di simulare e riprodurre l'intelligenza umana e che molte delle scelte e dei "comportamenti" della tecnologia rispondono ad algoritmi programmati con strategie e funzioni obiettivo che non sono immediatamente comprensibili dall'uomo comune ma le cui decisioni influiscono sulla sua vita. Tornando a S. Ignazio, il criterio chiave per un loro corretto utilizzo è quello del discernimento, ovvero della capacità di discriminare tra ciò che fa bene o non fa bene all'autorealizzazione personale. Andrebbe ovviamente chiarito a questo punto di che tipo di auto-realizzazione personale stiamo parlando. Un criterio di riferimento può essere quello dei milioni di dati individuali di cui ormai disponiamo in materia di soddisfazione e ricchezza di senso di vita. La letteratura che ne studia le determinanti indica che, al di là di reddito, salute e istruzione, una variabile chiave per la soddisfazione e ricchezza di senso della nostra vita è quella della generatività, studiata dallo psicologo sociale Erik Erikson ed approfondita in campo sociologico in Italia da Mauro Magatti. I lavori empirici più recenti sul principio della generatività suggeriscono che le persone sono felici nella misura in cui la loro vita incide positiva-

mente sulla vita di altre persone. La generatività è combinazione di creatività e capacità d'impatto positivo sulle vite altrui e si può esprimere in diversi modi. La generatività può dunque essere biologica, economica, artistica, sociale, culturale, relazionale. Il progresso tecnologico è dunque strumento potente da mettere al servizio della generatività personale e, quando questo avviene, può contribuire significativamente all'autorealizzazione e alla fioritura della vita umana.

Ovviamente ci sono molte insidie ed ostacoli che si possono frapporre a questo quadro ideale.

2. IL PROBLEMA DELLA FUNZIONE OBIETTIVO DEGLI ALGORITMI

Il primo punto fondamentale è quello di tener conto che la gestione del progresso tecnologico non è neutrale ma guidata da imprese che hanno come obiettivo quello di massimizzare il loro profitto senza necessariamente prestare particolare attenzione ai risvolti di sostenibilità sociale ed umana della loro azione. Una delle caratteristiche dominanti dei nostri tempi è la diffusione dei social media e delle tecnologie in rete. I nostri giovani, ma non solo, vivono immersi nella rete dei social che faremmo un errore a considerare neutrale. Massimizzare il profitto per una società che gestisce social media (Twitter, Facebook, Instagram, TikTok) vuol dire aumentare il traffico di interazioni che si realizza sulla propria piattaforma in modo da attirare maggiori finanziamenti pubblicitari proporzionali al cosiddetto costo/contatto (costo per numero di persone che si collegano). Per aumentare il traffico d'interazioni bisogna saper attirare una risorsa scarsa che è quella dell'attenzione degli utenti. Tra le strategie migliori per attirare l'attenzione ci sono quelle del creare conflitti tra gli utenti o favorire il relax che tutti cerchiamo durante il tempo libero. Se dunque, da un punto di vista positivo, le piattaforme diventano luoghi spesso divertenti di ironia ed ilarità (uno degli sviluppi più sorprendenti in questi ultimi tempi è quello dell'interazione tra tifosi e dei commenti durante e dopo gli eventi sportivi in cui le comunità che s'incontrano sulle piattaforme social rivaleggiano con gli opinionisti professionisti accreditati), da un punto di vista meno positivo le stesse piattaforme hanno interesse a creare e stimolare conflitti che attirano attenzione e traffico di utenti.

Per questo motivo le sparate, le violenze verbali spesso di minoranze diventano enormemente attrattive e occupano un ruolo centrale nell'interazione social. Molto spesso questi processi non sono spontanei perché i gestori delle piattaforme attraverso la programmazione degli algoritmi dispongono di molti trucchi e segreti per far prevalere un tema, orientare il dibattito in una certa direzione e produrre l'effetto desiderato. La conseguenza di tutto questo è il sovrappeso di opinioni estreme e minoritarie. Solo guardando alla storia recente del nostro paese abbiamo vissuto prima la storia della "bestia" che ha organizzato la campagna di ostilità contro migranti e stranieri utilizzando falsi account che moltiplicavano l'eco degli interventi e successivamente il fenomeno "no vax", che ha avuto una cassa di risonanza e un peso specifico sui social molto superiore a quello effettivo statistico nel paese. Usando un'immagine è come se in una stanza ci fossero nove persone adulte e sagge e un bambino. Il bambino però si è impossessato del microfono e monopolizza l'attenzione. Uno dei temi centrali per il prossimo futuro è dunque quello delle regole e della *governance* delle piattaforme. Per fare ancora l'esempio di Twitter, sotto la pressione dell'opinione pubblica che accusa la società ed altre che gestiscono simili piattaforme di aver creato condizioni favorevoli per la diffusione dell'*hatred speech*, la società ha dovuto modificare alcune delle regole del gioco volte al ridurre il potenziale inquinante delle "macchine del fango". Per fare alcuni esempi delle "riforme" più note il numero massimo dei caratteri utilizzabili è raddoppiato, la compagnia identifica con degli algoritmi potenziali messaggi di odio e li oscura automaticamente ed è ora possibile decidere chi può rispondere pubblicamente ad un messaggio che postiamo sulla piattaforma.

La storia recente dei social media con la parziale rettifica dei loro comportamenti originari ci fa intuire un percorso che probabilmente si ripete e si ripeterà in casi analoghi. I programmatori delle macchine e degli algoritmi hanno le loro funzioni obiettivo che non coincidono con quelle che creano le condizioni per il bene comune. Accade pertanto quello che tipicamente nei modelli economici è una differenza tra l'ottimo privato (il risultato a cui il modello arriva per effetto delle scelte individuali e, in questo caso, del programmatore dell'algoritmo) e l'ottimo sociale. Situazione poi risolta nei casi di scuola da un "benevolent

planner” che interviene con tasse o regolamentazione per “imbrigliare” le energie private e far coincidere ottimo privato e ottimo sociale. La differenza rispetto al modello standard è che la situazione è in questo caso molto più “opaca” perché non è facile risalire da quanto osserviamo alle strategie e alle funzioni obiettivo dell’algoritmo. Per questo è necessaria maggiore attenzione, vigilanza e spirito critico che alimenti attenzione e dibattito sul tema proprio come sta accadendo oggi attorno alle scelte di alcuni social media come Twitter e Facebook.

La morale di questa storia è che dobbiamo comprendere che la tecnologia non è neutrale, conoscere le sue funzioni obiettivo e le conseguenti strategie di chi ce le propone e contrastarle con opportune regole di *governance* quando sono contrarie al bene comune.

La questione etica discriminante per la quale dobbiamo batterci in futuro sarà quella di discutere pubblicamente le funzioni obiettivo che regolano i comportamenti degli algoritmi e stabilire se siano o no compatibili con le regole che come comunità vogliamo darci e con i nostri principi morali.

3. FIN DOVE ARRIVA L’IA: LA DIFFERENZA DELL’UMANO

Molte delle applicazioni dell’intelligenza artificiale stanno migliorando la nostra vita. Applicazioni dei cellulari ci segnalano il percorso più rapido per raggiungere la nostra meta nel traffico urbano (e se le usiamo bene come strumento di apprendimento ci insegnano nuovi percorsi che non conoscevamo che collegano la nostra abitazione con luoghi in cui ci rechiamo abitualmente), ci ricordano dove abbiamo parcheggiato la macchina. Siamo passati dai mangianastri dove registravamo dagli Lp le nostre canzoni preferite alla possibilità, attraverso Youtube, di vedere in qualunque momento una canzone o un video che ci interessa. E l’algoritmo di Youtube, che apprende dalle nostre preferenze rivelate nelle scelte passate, ci suggerisce sorprendendoci e cogliendo nel segno quale potrebbe essere la prossima canzone o il prossimo video che ci interessa.

Anche in questo caso è sempre bene domandarci dove l’algoritmo ci vuole portare e se il suo obiettivo coincide con il nostro. Non ci sono problemi quando, conoscendo ormai i nostri gusti, ci suggerisce le can-

zioni che vorremmo ascoltare leggendoci nel pensiero. Ci sono maggiori problemi quando ci spinge ad acquistare alcunché a partire dalle nostre preferenze a lui ormai note.

Ancora una volta la conclusione è che l'equilibrio e il discernimento sono virtù ancora più centrali ed importanti di fronte a questo aumento di potenza dell'umano che potrebbe travolgerci.

Altro tema centrale del dibattito è fin dove l'intelligenza artificiale potrà arrivare. La fantascienza, ma non solo, prefigura un mondo dove le "macchine" potranno sviluppare tutte le abilità, funzioni e capacità umane dominandoci. In realtà, riprendo su questo il punto sollevato da Ugo Amaldi, siamo da questo punto di vista vittime di un antropocentrismo debole che sottovaluta la ricchezza, potenza e differenza dell'umano, e forse di una comunicazione che ha interesse a sopravvalutare le potenzialità dell'intelligenza artificiale anche nell'intento di attirare risorse e fondi per finanziare iniziative e progetti.

Al momento nelle applicazioni esterne, l'intelligenza artificiale che conosciamo funziona molto bene come memoria esterna ed internauta aiutandoci a rintracciare nella rete ciò che vogliamo cercare. Le sue potenzialità di dialogo e le sue capacità relazionali sono al momento piuttosto primitive ma non è affatto detto che non possano aumentare significativamente offrendoci uno strumento di compagnia che possa rivaleggiare con quella di altri esseri umani.

Dobbiamo dunque domandarci se esistono limiti invalicabili che l'intelligenza artificiale non può varcare.

Tra questi sicuramente la libertà (le macchine non sono libere) e l'amore che sono caratteristiche proprie degli esseri umani. Le macchine diventeranno sempre più sofisticate e dovremo essere capaci di decifrare le loro strategie ed obiettivi di cui dovremo pretendere trasparente rivelazione da parte dei loro costruttori. Ma difficilmente potranno raggiungere la bellezza e complessità dell'umano.

È proprio sulla libertà che si gioca e si giocherà una parte centrale del dibattito. Visioni meccaniciste/deterministe affrontano il tema dell'intelligenza artificiale insistendo sul fatto che siamo e saremo sempre più eterodiretti e determinati da forze esterne al nostro libero arbitrio e che gli spazi dell'umano tenderanno a ridursi sempre di più. Il

futuro che sembra prefigurarsi può essere in un certo senso opposto. La potenza della tecnologia, le opportunità di umanità aumentata e l'intelligenza artificiale aumenteranno la potenza dell'umano e renderanno sempre più importante la nostra capacità di scelta tra potenti opzioni alternative. Ovviamente, in assenza di esercizio del discernimento, sarà possibile farsi trasportare da stimoli e sollecitazioni degli algoritmi perdendo i nostri spazi di libertà e di scelta. L'insidia forse più grave è un'altra. Con l'aumento della nostra potenza, la funzione con-creatrice dell'umano (prevista e considerata positivamente nella visione teologica della "creazione progressiva" dove l'uomo con la propria operosità completa e fa avanzare ogni giorno la creazione divina) rischia di farci perdere il senso di creaturelità, ovvero quella dimensione di sana "povertà" che ci mette nella giusta proporzione e relazione di fronte alla creazione e al Creatore riducendo progressivamente lo spazio del religioso e della fede. Siamo, saremo e ci sentiremo sempre più potenti, ma non potremo sconfiggere i nostri limiti strutturali che ci metteranno sempre di fronte a malattia e al termine della nostra esistenza terrena. Di fatto però, mentre noi discutiamo di questi possibili rischi, la storia recente sembra suggerirci che la percezione dei nostri limiti resterà ben chiara di fronte a nuove sfide (si pensi alla pandemia e al riscaldamento globale) che sembrano essere generate proprio dagli errori di utilizzo della nostra accresciuta potenza.

Non mancherà dunque oggi e in futuro la possibilità di interpretare con intelligenza e saggezza le condizioni del vivere. Le virtù principali del futuro saranno pertanto due. Quella individuale del discernimento e quella sociale di una cittadinanza attiva che dovrà esercitarsi nella vigilanza e nella richiesta di piena *accountability* dell'intelligenza artificiale di cui dovremo conoscere e discutere pubblicamente strategie e funzioni obiettivo.

L'IA: SOSTITUZIONE O SOSTEGNO DEL MEDICO?

Giuseppe R. Gristina e Luciano Orsi

1. DOCUMENTI SULLA IA IN MEDICINA

Nel 2018 la Commissione europea ha nominato un panel di esperti (The Artificial Intelligence High-Level Expert Group – AI HLEG) per elaborare un documento sull'*intelligenza artificiale* (IA) con lo scopo di garantire una più precisa e diffusa conoscenza di questa disciplina e della tecnologia correlata¹.

Il termine IA contiene dunque un riferimento esplicito alla nozione di *intelligenza*, anche se i ricercatori gli preferiscono quella di 'razionalità'², sottolineando la capacità di tutti i sistemi di IA di scegliere l'azione migliore da intraprendere per raggiungere un determinato obiettivo con criteri predefiniti.

Gli esperti dell'AI HLEG hanno quindi definito l'IA come l'insieme di sistemi progettati dall'uomo che, dato un obiettivo complesso, agiscono nella dimensione digitale o fisica percependo l'ambiente attraverso l'acquisizione e l'interpretazione di dati raccolti in maniera strutturata o non strutturata, elaborando le informazioni derivate da questi dati e scegliendo tra più azioni possibili quella da intraprendere per raggiungere nel modo migliore l'obiettivo prefissato nella situazione data.

Tuttavia, rispetto all'obiettivo di concepire sistemi di IA in grado di svolgere la maggior parte delle attività che gli esseri umani stessi svolgono ci sono ancora molte sfide aperte, sia scientifiche che etiche: queste

¹ The European Commission, *Independent High-Level Expert Group On Artificial Intelligence. A definition of Artificial Intelligence: main capabilities and scientific disciplines*, 08.04.2019: <https://bit.ly/3LxvSas> (ultimo accesso 02.04.2022).

² Russell S., Norvig P., *Artificial Intelligence. A Modern Approach*, Edimburgh, UK. 3rd edition, 2016.

ultime con un ovvio impatto in termini di affidabilità, rendicontazione, sicurezza, responsabilità.

In un recente rapporto della National Academy of Medicine³, dedicato all'utilizzazione dell'IA nell'ambito sia delle attività cliniche di diagnosi e cura che in quelle della ricerca, particolare enfasi è stata posta sulle opportunità che l'IA può offrire se utilizzata per l'implementazione della *medicina di precisione* (MP).

Secondo il National Research Council⁴ la MP può essere definita come un processo mirato a conseguire un adattamento delle terapie mediche alle caratteristiche biologiche e genetiche di ogni singolo paziente o di popolazioni di pazienti che differiscono nella loro suscettibilità a una particolare malattia o nella loro risposta a un trattamento specifico.

2. APPLICAZIONI DELLA IA ALLA SALUTE

Il potere della MP risiede pertanto nella sua capacità di guidare le decisioni sanitarie verso il trattamento più efficace per uno specifico paziente migliorando quindi la qualità dell'assistenza e riducendo al contempo le terapie inutili perché inefficaci, i relativi effetti collaterali, i test diagnostici superflui.

Il processo della MP prevede che le informazioni su singoli pazienti (stili di vita, storia clinica, esame fisico, ematochimica, *imaging*, diagnostica funzionale, immunologia/istologia, dati “-omics”: microbioma, genoma, proteoma, metaboloma) siano riunite in voluminosi set di dati (*big data*) il cui pieno utilizzo richiede lo sviluppo di metodi standardizzati di aggregazione, analisi e traduzione interdisciplinare tramite tecniche computazionali quali quelle dell'IA.

L'uso di questi nuovi metodi analitici potrà consentire di definire modelli di salute, malattia e assistenza più efficienti, sostenibili, guidati da dati e tecnologia.

³ Matheny M., Thadaneys Israni S., Ahmed M., Whicher D. (a cura di), *Artificial Intelligence in Health Care. The Hope, the Hype, the Promise, the Peril*, Washington DC, 2019.

⁴ National Research Council, *Committee on a Framework for Developing a New Taxonomy of Disease. Toward Precision Medicine. Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*, Washington DC, 2011.

Poiché l'IA va assumendo un ruolo sempre più rilevante nella pratica clinica, è plausibile ritenere che in futuro molte specialità mediche potranno essere influenzate nel modo di gestire ogni specifico processo clinico.

Sebbene alcuni credano che l'IA potrà addirittura sostituire il medico⁵, è più probabile che queste tecniche saranno principalmente di *supporto all'attività dei clinici*, ma è evidente che comunque più livelli di controllo si renderanno necessari.

3. LA NECESSITÀ DELLA SUPERVISIONE UMANA

Tuttavia, anche con un'adeguata supervisione, la fiducia in queste tecnologie non dovrebbe essere sopravvalutata. In questo senso sono sorti importanti quesiti in merito al livello di fiducia che dovrebbe essere concesso ai sistemi di IA⁶.

Ad esempio, alcune tecniche di apprendimento automatico, sebbene molto accurate, non assicurano la necessaria trasparenza circa le modalità con cui le decisioni vengono prese (*decision-making algorithms*). La nozione di *black-box* si riferisce appunto a quella parte dei sistemi di IA che, essendo inaccessibile, non permette di conoscere il motivo di determinate decisioni. Si tratta della cosiddetta *opacità*⁷ del modello che è tale sia per la sua intrinseca complessità, sia per *policies* strategiche o di segretezza correlate alla proprietà del brevetto o agli interessi industriali.

A fronte di enormi capacità già realizzate quali la sintesi di una sterminata letteratura scientifica che si moltiplica in maniera esponenziale o la disponibilità comparativa di enormi database di immagini diagnostiche radiologiche e istopatologiche, sono finora scarse le evidenze scientifiche della reale applicabilità alla popolazione generale dell'IA e la conseguente utilità nella pratica quotidiana.

⁵ Alvarado R., *Should we replace radiologists with deep learning? Pigeons, error and trust in medical AI*, *Bioethics*, 2022, 36(2):121-133.

⁶ Symons J., Alvarado R., *Epistemic entitlements and the practice of computer simulation*, *Minds*, 2018, 29:37-60.

⁷ Collecchia G., *Apriamo la scatola nera: l'intelligenza artificiale spiegabile*, *Recenti Progressi in Medicina*, 2021; 112(11):709-710.

Per accertare che tali capacità producano reali *outcome* positivi saranno comunque necessari *trial* clinici molto robusti (*Randomized Clinical Trials – RCT*) su campioni rappresentativi di malati. Finora, infatti, le ricerche sull'uso dell'IA in medicina sono state prevalentemente condotte su piccoli campioni di pazienti, selezionati con criteri discutibili e con *end-point* limitati. Potrebbe allora essere utile introdurre nuovi approcci di studio come confrontare i risultati di *équipe* sanitarie che utilizzano sistemi di IA con quelli di *équipe* sanitarie che non li utilizzano, invece che confrontare sanitari e IA sui risultati prodotti.

A tutt'oggi rimane da chiarire e normare, il ruolo degli enti che finanzieranno la ricerca di settore (enti pubblici, privati, *joint-venture* pubblico-privato) rispetto alla possibilità di svolgere 'funzioni di controllo' sui processi di utilizzazione dei dati, di strutturazione dei sistemi algoritmici e sui prodotti finali. Gli interessi commerciali potrebbero infatti influenzare l'approccio diagnostico a una patologia allargando o riducendo impropriamente le popolazioni dei malati o dei sani per orientare le decisioni nella pratica clinica e determinare, in ultima istanza, i profitti.

Ancora senza risposta rimangono poi almeno tre altri rilevanti interrogativi.

Il primo relativo alle modalità necessarie a conseguire e mantenere un equilibrio fra le *soft skills* umane dei sanitari (astrazione, intuizione, flessibilità, empatia) e le *hard skills* (enormi capacità di calcolo e gestione dei database) dell'IA. Il secondo riguarda la ripartizione della responsabilità (morale prima che giuridica) fra il sanitario che prende una decisione in base alle indicazioni fornite dall'IA (aderendovi o discostandosene) e il produttore dell'algoritmo che fornisce tali indicazioni; l'ultimo che attiene alla gestione del consenso informato del malato e/o del fiduciario nelle scelte di cura basate su dati forniti dall'IA.

La nostra società sta vivendo un momento storico senza precedenti in cui la disponibilità di grandi quantità di dati che riguardano l'essere umano, combinati con i progressi dell'IA, consentiranno in tempi brevi di affrontare problemi complessi attraverso l'uso di processi decisionali algoritmici. In ambito sanitario l'IA offre l'opportunità di migliorare significativamente sia le strategie di definizione dei programmi sanitari sia la gestione dei processi assistenziali e di cura di singoli pazienti.

Rimangono tuttavia ancora aperte le sfide cui la medicina deve fornire una risposta: le aspettative irrealistiche, la gestione dei dati, l'inadeguatezza dei criteri di equità e inclusione, il rischio di esacerbare le disparità sanitarie, le discriminazioni e i pregiudizi, i bassi livelli di affidabilità, l'incertezza normativa e giuridica, l'asimmetria di potere e informativa, la mancanza di trasparenza e responsabilità.

Pur ritenendo, come medici, che non si debbano perdere le opportunità offerte dall'IA, dovremmo concentrare la nostra attenzione sui rischi legati all'uso di processi decisionali algoritmici accettati pragmaticamente e non sottoposti a vaglio scientifico. Solo adottando un approccio scientifico, alimentato da un sano scetticismo in grado di non cadere in una acritica diffidenza, i medici potranno evolvere assieme alla tecnologia evitando il rischio di esserne dominati, o, peggio, di essere sostituiti dall'IA.

Jean-Pierre Darnis e Mons. Carlo Maria Polvani

L'Intelligenza Artificiale sta investendo l'insieme delle componenti della vita dell'individuo. Possiamo partire dall'esempio dell'introduzione delle IA nelle cause penali – sia in termini di analisi dei comportamenti passati, sia e ancor più, in termini di previsione di comportamenti futuri – che, oltre a creare dilemmi considerevoli nei tribunali, prova che l'IA sta riconfigurando uno degli assi principali dei nostri sistemi giuridici, quello della relazione e dell'equilibrio tra la tutela delle libertà individuali e la promozione del bene comune e dell'interesse pubblico. L'Unione europea sta introducendo le categorie di basso rischio, di medio rischio e di alto rischio nell'uso delle IA, ma non è del tutto chiaro se la proposta europea riuscirà ad imporsi efficacemente.

Va considerata, infatti, la rapidità dello sviluppo e la pervasività dell'utilizzo delle IA anche in campi tipici della creatività umana. A un orecchio non esperto è già difficile distinguere le parti della *Incompiuta* composta dell'autore originario da quelle completate da un algoritmo che ha studiato l'intera opera di Schubert per imitarne lo stile. Appare ragionevole supporre che, in breve tempo e magari anche con l'arrivo di computer quantistici, l'IA si trasformi in uno strumento di risoluzione dei problemi ben più competitivo dell'intelligenza e dell'intuizione umana. È prospettabile, per esempio, che una IA che avesse a sua disposizione tutti i dati genetici e medici di una persona e una lista completa di tutti i fattori medio-ambientali alla quale questa persona è stata sottoposta durante la sua vita, possa effettuare dei diagnostici molto oculati, per il semplice fatto che i dati a sua disposizione sarebbero immensamente più ampi di quelli che un dottore possa mai raccogliere e analizzare.

In alcuni campi, l'attendibilità delle soluzioni proposte da una IA potrebbe portare persino all'obsolescenza dell'uso delle intelligenze e delle capacità umane. Questo è già accaduto per alcuni compiti (quali la verni-

ciatura di automobili su una catena di montaggio) e si deve anche al fatto che le IA sono più affidabili degli esseri umani nell'esecuzione di incarichi precisi anche di natura non ripetitiva.

Questa notevole affidabilità delle IA invoglierà i settori pubblici e privati a farne un uso sempre più ampio che potrebbe in prospettiva escludere gli umani dai processi decisionali. Un possibile esempio premonitore è riscontrabile nel campo della difesa. Esistono già gli *Autonomous Weapons Systems* (AWS). Alcuni AWS sono capaci di analizzare uno specifico teatro di conflitto (per esempio, soldati alleati che entrano in una zona ostile), identificare i possibili bersagli (le posizioni nascoste dei soldati nemici) e sottoporre un'azione per neutralizzarli (un tiro di mortai sulle postazioni avverse) a un essere umano, che mantiene, per ora, la facoltà di prendere una decisione finale (dare l'ordine all'artiglieria di aprire il fuoco). Con il miglioramento delle IA, questo ultimo passaggio potrebbe rivelarsi fatale in una situazione di conflitto reale, in quanto il tempo di reazione e di decisione di un umano, potrebbe invogliare a lasciare alle IA l'intera gestione dello scontro onde evitare effetti nefasti. Si noti, aneddoticamente, che questi sistemi di delega decisionale alle IA sono già operativi sugli aerei da caccia, nei quali la IA può entrare, nel caso di minaccia imminente, in modalità *override*, intervenendo prima e senza il consenso del pilota.

Questi esempi sono direttamente collegati con il primo – e fondamentale – principio sollevato da Giuliano Amato quando stipula che l'uso delle IA debba aiutare gli umani e mai sostituirsi a loro nella presa di decisioni importanti. Come è emerso nella nostra discussione, per difendere questo principio si rivela indispensabile la categoria della libertà umana, senza la quale le decisioni degli esseri umani non sarebbero esse stesse che il prodotto di un algoritmo dal funzionamento biochimico su una base neuronale al posto di un algoritmo elettrico su una piattaforma formata da semiconduttori.

Non dovrebbe quindi sorprendere che la categoria di libertà umana, a livello filosofico, antropologico ed etico, potrebbe essere quella che costituisce un parametro critico fondamentale. Sia concesso un *Gedankenexperiment* dal sapore huxleyiano, ma non così tanto fantascientifico se si considerano i recenti avanzamenti delle IA (per esempio, AlphaFold della DeepMind).

Si immagini che una IA futuristica abbia a sua disposizione i dati di un carcerato condannato per violenza sulla convivente e se ne debba decidere il rilascio per buona condotta, con il rischio di reiterata aggressione sulla vittima; tali dati includerebbero: il patrimonio genetico completo del reo (identificando tutti i geni che lo predispongono alla violenza), pienamente integrati con altri dati della sua famiglia e una miriade di fattori medio-ambientali più disparati (livello di studi, amicizie e relazioni sociali, pratica religiosa, condizioni economiche, precedenti penali in famiglia e via dicendo) e poi analizzati e collegati su base statistica, tenendo conto di una database proveniente da migliaia di individui condannati per violenze domestiche e poi rilasciati per buona condotta. Immaginiamo, inoltre che, affinando sempre le sue capacità a partire da un miglioramento continuo dell'algoritmo e dalla costante introduzione di sempre più dati, la IA in parola raggiunga una prevedibilità di recidiva con una precisione quasi ineccepibile da un punto di vista probabilistico.

A parte la delicata questione sollevata da Luca Antonini su quale giudice si arrischierebbe nel pronunciare una sentenza contraria alle previsioni di questa IA avanzatissima, a parte la questione più inquietante sollevata da Eugenio Mazzarella sull'eticità di correzioni eugenetiche sui tratti ereditari responsabili di comportamenti sociali pericolosi, emergerebbe un'altra domanda più a monte: se un algoritmo matematico potesse prevedere un comportamento umano in forma affidabile, si potrebbe allora, ancora parlare di libertà umana? O detto in altre parole: la putativa prevedibilità di un comportamento umano sulla base di un complesso calcolo matematico-probabilistico, non sarebbe la miglior prova per dubitare dell'esistenza – o comunque ridimensionare la portata – della categoria stessa di libertà umana?

L'IA è un prodotto dell'uomo e quindi fa parte dello spazio cognitivo e sociale dell'uomo. Ed è spesso difficile, se non impossibile, fare la cesura fra l'uomo e la tecnologia che produce, in quanto rappresenta da sempre un oggetto integrato nel proprio campo sociale, nelle proprie rappresentazioni. Basti pensare agli strumenti, alla possessione e l'uso delle armi nelle civiltà antiche. Riprendendo il pensiero che Antonino Raspanti espone nel suo saggio, è utile e doveroso esaminare aspetti e dinamiche delle creature, compresa quella umana, che ci sono ignoti, anche per misurarne i rischi. Ma

se l'IA, in quanto artefatto, deve essere valutata come un potenziale elemento dirompente nella sfera individuale e sociale, va però riaffermata la sua distinzione dell'umano con una dignità che in ultimo si fonda nella sua destinazione alla comunione con Dio, sempre seguendo la visione di Raspani.

Ci sono poi problematiche preoccupanti, se non pericolose, che vanno trattate nella sfera umana e dalla sfera umana. Una volta detto questo si può accogliere la discussione sul carattere antropomorfo della tecnologia, ma anche sul desiderio degli uomini di produrre una tecnologia a loro immagine (robot antropomorfo, sistemi informatici con assistenza vocale che adoperano un vocabolario gradito e umanizzante...). La stessa espressione "intelligenza artificiale" per descrivere procedure basate su algoritmi introduce una trasgressione, indicando la "tecnologia" (ovverosia la macchina) come intelligente, il che è una proprietà umana (cfr. Carla Colicelli quando parla del "colore dell'intelligenza") e quindi crea una categoria nuova, un golem o frankenstein tecnologico che sarebbe in qualche modo un potenziale rivale dell'uomo (aspetto sottolineato da Stefano Zamagni quando parla di *playing gods*) E bisogna stare attenti al paradigma *HAL 2001 Odissea dello Spazio / Skynet-Terminator 3* in cui si descrivono macchine che per proseguire i loro fini si mettono ad eliminare in modo metodico gli uomini. Queste visioni contengono elementi millenaristi e apocalittici che vanno analizzati in quanto tali, sottolineando la persistenza di alcuni discorsi storici e culturali anche nei contesti moderni: la macchina diventando l'ultima *ratio* di un umanissimo desiderio di annientamento apocalittico. Possiamo quindi cercare di esprimere la tecnologia come fatto sociale (e dunque umano), e quindi contenitore (e semmai amplificatore) delle passioni e nevrosi individuali e collettive. Ma anche sottolineare i progetti sottostanti nonché le costruzioni relative all'espressione "intelligenza artificiale". Per ragionare meglio, si dovrebbe pensare all'intelligenza umana nei suoi rapporti con l'artefice (e quindi la tecnologia basata su algoritmi), il che cambia la prospettiva del problema (un approccio che si congiunge con le conclusioni di Stefano Zamagni sulla possibilità o no di attribuire alla macchina lo *status* di agente morale).

Da questo punto di vista le conclusioni sviluppate da Laura Palazzani nel testo trasmesso ci aiutano a focalizzare sulle priorità: «rifondare un umanesimo tecnologico significa evitare da un lato di "antropomorfizzare"

la macchina (umanizzare la tecnologia) o artificializzare l'uomo a "macchina biologica" (tecnologizzare l'uomo). La "macchina" può essere una alleata dell'uomo e l'integrazione tra naturale ed artificiale può divenire complementare, sempre nella priorità del rispetto dell'umano, salvaguardando ciò che "conta" dell'umano, nel quadro dei diritti umani fondamentali». Va rilevato che Laura Palazzani introduce il concetto di "diritti umani fondamentali" come uno strumento/concetto da adoperare per la salvaguardia nel rapporto uomo/macchine.

La questione dei diritti è affrontata in modo estremamente efficace da Luca Antonini che da questo punto di vista pone giustamente al centro della riflessione l'operato della Commissione europea, un'istituzione che appare all'avanguardia per la produzione di regolamentazione dei dati (GDPR) con la proposta di regolamentazione dell'IA. Le istituzioni europee stanno svolgendo un lavoro sia di posizionamento dei diritti sia di ricerca di un sistema di tutela distinguendo le IA secondo la loro pericolosità, ed è ovvio che in qualche modo eventuali raccomandazioni debbono partire da lì, anche riferendosi al concetto di "algor-ethics" che si ritrova nel *Rome call for AI ethics* firmato nel 2020 dalla Ponteficia Accademia per la Vita (aspetto anche analizzato da Carla Colicelli). Dal canto suo Leonardo Becchetti ben descrive meccanismi ed effetti perversi del "capitalismo di sorveglianza".

Possiamo anche prolungare questa riflessione cercando di qualificare alcuni problemi legati all'uso di tecnologie algoritmiche.

L'IA potrebbe rappresentare un pezzo di tecnologia, che può essere ritenuto come fondamentale, ma che non raffigura un pericolo "a sé". Nell'evoluzione storica dell'IA si pongono però una serie di questioni sulle implicazioni dell'utilizzo di una cosiddetta IA "forte", aspetto sottolineato da Paolo Benanti nel suo saggio. Alcuni scenari di uso dell'IA possono portare a una diminuzione se non un annientamento della libertà positiva. Per questo motivo Paolo Benanti invoca l'adoperare di un concetto di "sostenibilità digitale", riprendendo la riflessione iniziata con il concetto di "algor-ethics" per portarla a una visione più larga.

Nella logica descritta da Leonardo Becchetti, osserviamo una crescita delle capacità di trattamento dei dati che avanza di pari passo con le capacità di raccolta e trasmissione di dati. La dinamica è quella di una crescita esponenziale di smartphone, *cloud* e capacità dei computer. Da questo pun-

to di vista va sottolineato che è la diffusione di massa degli smartphones che costituisce il primo anello fondamentale della catena di trattamento e uso dei dati. Lo smartphone appare oggi come un fantastico sensore che accompagna l'individuo 24 ore su 24 ed è capace di raccogliere permanentemente informazioni attive (le interazioni dell'individuo con alcune piattaforme) o passive (il posizionamento, il non uso che rivelano altrettante informazioni...).

Siamo quindi di fronte all'accumulo di una larga massa di dati legati a tutte le nostre interazioni online. Come detto da Leonardo Becchetti, gran parte di questi dati viene raccolta dalle reti sociali che adoperano un loro preoccupante modello di massimizzazione del profitto. Il successo delle reti sociali poggia su un'umanissima inclinazione, ovverosia sull'imperioso e intimo desiderio da parte di ognuno di socializzare, indole che si riversa sulle reti. Detto questo, alcune aziende statunitensi hanno creato un meccanismo di "capitalismo di sorveglianza" nel quale la crescita del traffico rappresenta l'obiettivo primario. Tutto questo spinge ad elaborare, con largo ricorso alle IA, meccanismi di marketing predittivi, in grado anche di orientare le azioni dell'individuo/consumatore, il che poi spinge anche a tentare di ampliare l'interesse, per non dire la dipendenza, tramite tecniche di profilazione (*profiling*) e di aumento della permanenza dell'individuo sulle reti sociali, suscitando il suo interesse costante e orientando alcune scelte, in quanto più resta online, maggiore è il valore per la stessa piattaforma. Già il connubio fra profilazione e capacità predittive lascia intravedere scenari inquietanti se la tecnologia non viene regolamentata.

Gli algoritmi sono delle opinioni iscritte nei codici che vano a stanare quello che negli Usa viene chiamata "la tana del coniglio" (*rabbit hole*), ovvero un punto focale della nostra attenzione/gusti, su cui poi si costruisce e si elabora. Le esagerazioni/false informazioni producono più traffico, e dunque più guadagno per le piattaforme rispetto a quelle vere o verificate, un vero e proprio modello di estrazione dell'attenzione. Il che può essere anche tradotto in manipolazioni politiche. Il caso della manipolazione politica basata su Facebook fornisce il punto di partenza della riflessione di Eugenio Mazzarella quando analizza i pericoli di una traslazione della vita sociale dentro l'infosfera.

Questo spiega anche come poi vengono alimentati cicli di notizie false fatte per rafforzare le convinzioni di questo o quel gruppo aumentando le loro comunicazioni (sviluppo di informazioni in “silos” che si auto-alimentano e rinforzano l'esistenza di mondi cognitivi paralleli).

Un altro aspetto preoccupante sta nelle forme di “dipendenza tecnologica”, ovvero sia l'impossibilità di vivere senza smartphone, percepito come una necessaria stampella individuale e sociale. Il che non è direttamente legato alle tecniche di intelligenza artificiale, bensì all'uso che ne fanno le piattaforme globali/reti sociali che hanno sviluppato un modello economico fuorviante. Modello che non riesce a cambiare malgrado l'importanza delle problematiche e spesso il carattere “illegale” di alcuni effetti. Si pone quindi un enorme problema di politica pubblica nei confronti di queste aziende, perché esercitano una specie di monopolio inedito che rappresenta un potenziale pericolo per la vita civile, la democrazia, i diritti. In sintesi, si può tentare di differenziare le aziende a seconda delle problematiche etiche/civili che producono?

Inoltre se si considera che Facebook si è recentemente ribattezzata Meta anche per significare la sua accelerazione verso il metaverso, un universo digitale/rete sociale potenziato dalla realtà virtuale che offre vertiginosi aspetti di mondo parallelo, con potenziali effetti di ulteriore dipendenza e manipolazione dei comportamenti, dobbiamo poi ulteriormente interrogarci, come fa già Eugenio Mazzarella nel suo testo, sulle implicazioni di quest'accelerazione, da un punto di vista filosofico ma non soltanto.

Una delle ipotesi potrebbe quindi essere di identificare le reti sociali/piattaforme come un vero problema, ma non necessariamente la tecnologia dell'IA usata dalle stesse reti. Il contributo di Giuliano Orsi indica giustamente le valutazioni in corso sui benefici dell'intelligenza artificiale nel campo della medicina, fornendo potenziali esempi positivi.

Più che l'intelligenza artificiale in quanto tecnologia basata su algoritmi, bisogna pensare a un divieto del *profiling* e delle attività di sviluppo di modelli predittivi. Gli algoritmi di IA sono soltanto lo strumento di questa logica. Certamente gli algoritmi IA rappresentano una tecnologia fondamentale per queste operazioni, ma bisogna interrogarsi sulla parte di “responsabilità” all'interno di queste catene aziendali e sul ruolo delle istituzioni sovrane (Stati).

Il mercato delle connessioni è “gratuito”, il che significa che ci affidiamo a piattaforme che per esistere sviluppano questi servizi di “tramite,” con esiti problematici in termini di rispetto dell’individuo/dell’umano. Al centro di molti interventi troviamo una preoccupazione per la libertà, e quindi una preoccupazione per le varie forme di condizionamento che possono risultare dall’uso delle tecnologie. Il connubio fra profilazione e previsione dei comportamenti può rapidamente portare a esiti distopici preoccupanti, in quanto possono aumentare il controllo delle istituzioni sugli individui, ma anche producono saperi e tecnologie in grado di influenzare pesantemente il corso della vita degli individui, il che può avere risvolti su tutte le libertà, politiche o religiose.

Alcune domande si fanno strada: soluzione solvibile dall’IA stessa? Migliorare la tecnologia? Rendere etica la tecnologia sviluppando il concetto di “sostenibilità digitale”. Oppure in alcuni casi di piattaforme troppo globali (Facebook) non è più pensabile poter curare e rendere “etica” la piattaforma? (e quindi bisogna pensare a un’opposizione, smantellamento o addirittura divieto?). Da questo punto di vista bisogna tenere in mente l’esperienza cinese dove il potere centrale pone divieti sull’uso di alcune piattaforme perché non corrispondono ai valori che intende inculcare nella popolazione. Sicuramente abbiamo una scala valoriale diversa, ma la questione del divieto come possibilità di azione va seriamente presa in considerazione.

Qual è la minaccia per la nostra società? La tecnologia o l’amplificazione tramite la tecnologia dei nostri comportamenti peggiori? La regolamentazione europea è fondamentale. E il tentativo di classificare le IA secondo la loro pericolosità va perseguito, ma non pone il problema delle IA già esistenti in modo “massiccio e totale”.

Bisogna agire pesantemente sul controllo, la privacy ma anche il regolamento economico delle attività basate sui dati, con le sovranità degli Stati che si debbono affermare sugli attori privati (google tax, ma anche regolamentazione e tasse per l’accumulo dei dati.)

Al di là della regolamentazione esiste anche uno spazio per investimenti diversi da quelli delle maggiori piattaforme. Per gli Stati sovrani, ma anche l’Unione Europea, potrebbe essere fondamentale non soltanto legiferare, regolamentare, tassare o vietare, ma anche creare o certificare

i propri algoritmi per funzioni pubbliche, in modo tale da non lasciare lo sviluppo di IA soltanto alle piattaforme digitali ma di aprire spazi regolati e regolamentati, anche nel futuro metaverso. Il che significa investire nelle capacità tecnologiche e umane relative, una politica che sembra opportuna a livello dell'Unione Europea, anche per mantenere una società "propria" e non ritrovarsi schiacciati fra i "meta universe" americani e cinesi.

Il caso di "ChatGPT" può fornire ulteriori spunti di riflessioni utili. Nel novembre del 2022 la società Open AI ha rilasciato un prototipo di agente di conversazione ("Chatbot") basato su un programma di linguaggio (*Generative Pre Train Transformer*, GPT3) che poggia su un allenamento sorvegliato dall'uomo e che comporta anche l'apprendimento di divieti per contenuti problematici segnalati nell'apprendimento "umano". La logica fondamentale di quel programma è di predire le associazioni di parole ai concetti posti come domanda, con una capacità che proviene dalla vastità delle banche dati a disposizione. Questo programma si è dimostrato particolarmente interessante sia per elaborare testi scritti (ad esempio risposte a questioni scolastiche) ma anche per poter generare programmi (linee di codici). Come in ogni applicazione tecnologica, appaiono anche dei risvolti non previsti, ovverosia delle applicazioni (programmazione, traduzione, identificazione di immagini...) che possono essere di grande utilità. ChatGPT3 non rappresenta un salto tecnologico in quanto usa tecnologie già disponibili, ma la sua messa a disposizione in modo gratuito e aperto ha creato un'accelerazione nell'utilizzo, con poi prospettive di concorrenza (Google/Bard) o di integrazione in piattaforme a pagamento (Microsoft). Uno degli interrogativi più banali riguarda la capacità di questo Chatbot di produrre testi in risposta alle domande, funzionalità subito utilizzata nei contesti scolastici dove è diventato un "caso" sul come evitarne il ricorso sistematico da parte degli studenti. Tutto questo ha innescato un dibattito sui pericoli di una tecnologia che non soltanto permette a molti di evitare lo sforzo di apprendimento e di scrittura, ma che rischia poi di riprodurre all'infinito ragionamenti omogeneizzati per non dire pregiudizi. E di conseguenza vediamo già apparire i difensori di capacità algoritmiche di controllo dello strumento ChatGPT (ad esempio per la detezione degli scritti prodotti da algoritmi, e quindi evitare di copiare) mentre altri rivendicano modalità di insegnamento e di controllo che non lascino spazio a queste

possibilità (ad esempio con interrogazioni orali). Questi piccoli esempi illustrano come ChatGPT ci permette di ragionare sulle varie problematiche legate al trattamento algoritmico delle informazioni, riprendendo vari elementi che avevamo già evocato. Da un lato l'emergenza di ChatGPT potrebbe anche scalfire il quasi monopolio di Google come motore di ricerca, minacciandone anche il relativo "universo". Il che potrebbe introdurre un paradossale fattore di concorrenza e di relativo pluralismo. Detto questo, le elaborazioni di ChatGPT pongono una serie di problemi come quello dei diritti d'autore relativi al corpus di dati (database) che "nutre" il servizio, ponendo l'esigenza di pensare ad un'estensione del regime di "diritto connesso" degli editori di stampa per l'insieme dei testi utilizzati. Avendo poi in mente che l'integrazione di questo tipo di Chatbot dentro un motore di ricerca – ad esempio Bing di Microsoft – ne amplifica la capacità di raccolta di informazioni personali tramite la navigazione "integrata" e l'elaborazione di testi o prodotti. La natura di "robot di discussione" potrà nutrire qualche svista antropomorfa, ovvero dove la "risposta" elaborata su serie di probabilità offre sembianze "umane", e quindi di fatto rilancia la qualificazione di "intelligenza" da parte dell'artefice, ma anche questo ormai rappresenta un corollario tutto sommato classico dell'evoluzione della tecnologia e delle sue percezioni.

La sfida delle IA quindi, si dirigerebbe lungo una polarità composta da due vertici contrapposti: l'IA, da una parte obbligherebbe alla difesa della dignità dell'umano e dall'altra, inevitabilmente, porterebbe a questionare le basi stesse sulle quali essa si fonda. Ammettendo che questo sia vero, l'uso delle IA porterebbe ineludibilmente a una revisione di tante categorie che hanno contribuito alla formazione della percezione dell'uomo nelle nostre civiltà: la coscienza morale, la responsabilità e l'imputabilità dei propri atti, il rapporto fra innato e acquisto, per menzionare solo alcuni di quelli che sono emersi, direttamente o indirettamente, nel nostro incontro. Le conclusioni che si trarranno, man mano che le IA diventeranno sempre più performanti e sempre più invasive di ogni sfera dell'agire umano, dovranno ricercare un bilanciamento fine, in modo tale da congiurare i pericoli (la sostenibilità digitale) ma allo stesso tempo rifondare un "umanesimo tecnologico" come invocato spesso nel nostro gruppo. Questo potrebbe essere il *brave new world* che ci aspetta.

I membri della Consulta Scientifica del Cortile dei Gentili

UGO AMALDI

È stato Dirigente di ricerca nel Laboratorio di Fisica dell'Istituto Superiore di Sanità e al CERN. Ha pubblicato 600 lavori di fisica subatomica e medica. Ha insegnato fisica delle particelle e fisica medica alle Università di Firenze e Milano. Due milioni di allievi delle scuole superiori hanno studiato fisica sui suoi testi. È Distinguished Affiliated Professor della Technische Universität München. Al CERN ha fondato e diretto per 13 anni la collaborazione internazionale DELPHI. Dal 1992 è Presidente della 'Fondazione per Adroterapia Oncologica' TERA, che ha progettato il Centro Nazionale di Adroterapia CNAO di Pavia dove, nel quadro del SSN, sono stati irradiati 4500 pazienti.

GIULIANO AMATO

Professore emerito di Sapienza Università di Roma e dell'Istituto Universitario Europeo di Firenze, è stato Presidente della Corte Costituzionale. Ha ricoperto in passato incarichi pubblici diversi, in Italia e in sede europea.

LUCA ANTONINI

È Professore ordinario di Diritto costituzionale presso la facoltà di Giurisprudenza dell'Università degli studi di Padova. Il 19 luglio del 2018 è stato eletto dal Parlamento in seduta comune giudice della Corte Costituzionale. Ha ricoperto in precedenza diversi incarichi scientifici e istituzionali.

ANDREA ARCANGELI

È l'attuale Direttore della Direzione di Sanità e Igiene dello Stato della Città del Vaticano. È Professore associato di Anestesia e Rianimazione

presso l'Università Cattolica del Sacro Cuore. È stato responsabile del reparto di Terapia Intensiva Post-Operatoria del Policlinico Agostino Gemelli di Roma fino al 2019.

LEONARDO BECCHETTI

È Professore ordinario di Economia Politica presso l'Università Tor Vergata di Roma dove è Direttore del Master in Economia dello Sviluppo e Cooperazione Internazionale (MESCI). È Co-fondatore di Next e Gioisto e autore di circa 560 lavori tra pubblicazioni scientifiche, *working paper* e numerosi volumi divulgativi. È consigliere economico del MEF, membro del comitato di Bioeconomia del MASE e del comitato scientifico per la valutazione delle misure di contrasto alla povertà del ministero del lavoro. È tra i promotori della Scuola di Economia Civile, direttore del festival di Economia Civile, editorialista di *Avvenire* e del *Sole 24 Ore*, membro del Comitato scientifico del *Corriere della Sera* *Buone notizie*. È Presidente del comitato Etico di Etica SGR.

PAOLO BENANTI

Francescano del Terzo Ordine Regolare (TOR), teologo, si occupa di etica, bioetica ed etica delle tecnologie. I suoi studi si focalizzano sulla gestione dell'innovazione: internet e l'impatto della *Digital Age*, le biotecnologie per il miglioramento umano e la biosicurezza, le neuroscienze e le neurotecnologie. Dal 2008 è docente presso la Pontificia Università Gregoriana. È membro ordinario della Pontificia Accademia per la Vita. A fine 2018 è stato selezionato dal Ministero dello sviluppo economico come membro del gruppo di trenta esperti che a livello nazionale hanno il compito di elaborare la strategia nazionale sull'intelligenza artificiale.

MAURIZIO BETTINI

È Professore emerito di Filologia Classica presso l'Università di Siena, dove ha fondato il Centro "Antropologia e Mondo antico", che dirige. Tiene seminari presso la University of California at Berkeley, è stato Directeur d'Études associé presso l'École des Hautes Études di Parigi, ha insegnato al Collège de France e ha tenuto corsi in varie università del Regno Unito, Nord America, Giappone e Cina. Con l'editore Einaudi cura

Mythologica; presso l'editore Il Mulino è responsabile della collana Antropologia del Mondo Antico. Collabora con la pagina culturale de La Repubblica ed è autore di romanzi e racconti. Il suo principale campo di studi è la riflessione antropologica sulla cultura greca e romana, spesso in rapporto con l'esperienza della modernità.

CINZIA CAPORALE

Coordina il Centro per l'Etica e l'Integrità nella Ricerca del Consiglio Nazionale delle Ricerche e l'omonima Commissione. È membro del Comitato Nazionale per la Bioetica (PDCM) dal 2002 e del Comitato Etico Nazionale per le sperimentazioni cliniche relative alle terapie avanzate. È presidente onorario del Comitato Etico della Fondazione Umberto Veronesi. È stata Presidente eletta del Comitato Intergovernativo di Bioetica (IGBC) dell'Unesco per due mandati, capo della Delegazione italiana sull'Etica, membro della World Commission on the Ethics of Scientific Knowledge and Technology (COMEST) e del Legal experts Group for UNESCO GEObs-law database. Ha presieduto il Comitato Etico dell'INMI L. Spallanzani e il Comitato etico unico nazionale per le sperimentazioni su Covid-19. È stata membro del Comitato Tecnico Scientifico (CTS) del Governo Draghi per la gestione pandemica. È membro della Consulta scientifica del Cortile dei Gentili dal 2016.

AMEDEO CESTA

È dirigente di ricerca presso il CNR, all'Istituto di Scienze e Tecnologie della Cognizione. Si occupa da sempre di Intelligenza Artificiale, in particolare di presa di decisioni autonoma, risoluzione di *problemi complessi di pianificazione e scheduling*, di integrazione di tecnologie cognitive e IA, interazione umani - robot, sia in fabbrica sia per supporto a popolazione fragile. Alcuni prodotti di ricerca del suo gruppo sono in uso presso l'Agenzia Spaziale Europea. È socio fondatore di AIxIA (Associazione Italiana per l'Intelligenza Artificiale) e ne è stato presidente (2013-2017). È stato componente delegato del CNR presso il Comitato Nazionale per la Bioetica. È membro della Consulta scientifica del Cortile dei Gentili dal 2019.

CARLA COLLICELLI

Sociologa del welfare e della salute, Associate Researcher presso CNR – CID Ethics e senior expert ASviS per le relazioni istituzionali. Ha studiato alla Sapienza ed al Diipf di Francoforte sul Meno e tra 1980 e 2016 ha diretto ricerche e studi presso il Censis, tra il 1993 ed il 2015 come Vice Direttore Generale. Ha insegnato Sociologia della Salute alla Sapienza e Sociologia dei Servizi sociali a Roma3 e attualmente è docente di Comunicazione biomedica e scientifica presso il Dipartimento di Sanità Pubblica della Sapienza. È socio fondatore della Rete europea “Social Monitoring and Reporting” e della Fondazione Humanitas Anaste, Vice-presidente dell’Osservatorio Giovani e Alcol e membro della Consulta scientifica del Cortile dei Gentili della Pontificia Accademia delle Scienze. Oltre ai saggi ed articoli, ha pubblicato: *One Health: salute e sanità alla luce della pandemia* (2022 Editoriale scientifica), *Big data, privacy e comunicazione in epoca pandemica* (2021 L’Arco di Giano), *L’innovazione in sanità* (2019, L’Arco di Giano), *50 anni di Censis* (2014, Censis), *Salute come processo sociale* (2011, FrancoAngeli), *Le transizioni sommerse degli anni 90* (2004, Rubettino), *Welfare fai da te* (2000, Rubettino), *Benessere e tutela* (1998, FrancoAngeli).

FRANCESCA MARIA CORRAO

È Professore ordinario di Lingua e cultura Araba all’Università Luiss di Roma dove dirige il Master Mislam. È membro delle Associazioni di orientalistica UEAI e EURAMAL e dell’Institute of Oriental Philosophy (Soka University Tokyo). Presiede il Comitato Scientifico della Fondazione Orestyadi di Gibellina. Tra i suoi libri: *Nuri al-Jarrah. Esodo dagli abissi del Mediterraneo* Le Monnier 2023; *States, Actors and Geopolitical Drivers in the Mediterranean*, a cura di Corrao F.M. e Redaelli R., Palgrave; *I cavalieri, le dame e i deserti. Storia della poesia araba* (Aseq 2020); *L’Islam non è terrorismo* (Il Mulino 2018); *Islam Religion and Politics* LUP 2017; *Le rivoluzioni arabe. La transizione mediterranea* (Mondadori 2011).

JEAN-PIERRE DARNIS

È Professore associato all’Université Côte d’Azur (Nizza), dove dirige il corso di laurea in “Lingue e Affari Internazionali, Relazioni franco-italiane”. Nell’ambito del CMMC (Centre de la Méditerranée Moderne et Con-

temporaine, Nizza) coordina il programma di ricerca “La Francia, L’Italia e i loro Mediterranei”. Insegna nella facoltà di scienze politiche della LUISS Guido Carli ed è Associate Fellow della Fondation pour la Recherche Stratégique (FRS, Parigi). Scrive per il quotidiano Il Foglio ed è membro del comitato di redazione della rivista Cahiers de la Méditerranée. Fa parte della Consulta Scientifica del Cortile dei Gentili dal 2016.

EMMA FATTORINI

Professore ordinario di Storia contemporanea dal 2000 presso la Università “La Sapienza” di Roma. Collaborato con la Freie Universität di Berlino e la LMU di Monaco di Baviera. I suoi studi vertono su la Secolarizzazione e i movimenti cattolici in Europa, la Storia politico-diplomatica della Santa Sede (Archivio Segreto vaticano) tra Otto e Novecento. Ha collaborato al Corriere della Sera e al Sole 24 Ore, curato programmi radiofonici, diretto ricerche internazionali. È stata membro del Comitato Nazionale per la Bioetica e del Comitato scientifico della Treccani. È stata Senatrice nella XVII legislatura.

ALBERTO GIANNINI

Medico, è specialista in Anestesia e rianimazione e Direttore dell’Unità Operativa di Anestesia e Rianimazione Pediatrica dell’Ospedale dei Bambini – ASST Spedali Civili di Brescia. Ha coordinato il Gruppo di Studio per la Bioetica della SIAARTI (Società Italiana di Anestesia, Analgesia, Rianimazione e Terapia Intensiva) ed è responsabile del Comitato Etico della Società. Nel corso degli anni ha approfondito lo studio di alcuni temi di etica clinica relativi alla medicina intensiva (processi decisionali di fine vita, allocazione di risorse limitate, *visiting policies*).

GIUSEPPE R. GRISTINA

Medico, è specialista in Anestesiologia-Rianimazione ed Epidemiologia Clinica. È stato responsabile della sezione di alta specialità di cure palliative presso il Centro per lo Shock e il Trauma dell’ospedale San Camillo-Forlanini di Roma. È stato componente del Comitato Etico della Società scientifica di Anestesia Analgesia Rianimazione e Terapia Intensiva (SIAARTI) della quale ha coordinato anche l’attività del Gruppo di Studio per la Bioe-

tica. Della stessa società scientifica è oggi socio onorario. È stato Relatore ufficiale SIAARTI nel corso di audizioni parlamentari presso la Camera dei Deputati e il Senato della Repubblica. Principali aree di studio: etica clinica in terapia intensiva (decisioni di fine-vita, comunicazione in sanità, problematiche etiche relative alla distribuzione delle risorse assistenziali), sepsi, trattamento intensivo del trauma cranico e toracico.

FRANCO LOCATELLI

Medico, è specialista in Pediatria ed Ematologia. Ha diretto in passato la Clinica Pediatrica dell'Università di Pavia, Fondazione Policlinico San Matteo. È attualmente Professore ordinario di Pediatria presso l'Università Cattolica del Sacro Cuore ed è Direttore del Dipartimento di Oncoematologia e Terapia Cellulare e Genica dell'Ospedale Pediatrico Bambino Gesù. È stato Presidente dell'Associazione Italiana di Oncoematologia Pediatrica (AIEOP) e coordinatore del gruppo europeo per la diagnosi e trattamento delle sindromi mielodisplastiche del bambino. È Presidente del Consiglio Superiore di Sanità e ha svolto il ruolo di coordinatore del Comitato Tecnico Scientifico (CTS) del Governo per la gestione pandemica.

NICOLA MAGRINI

Medico, farmacologo clinico, esperto di linee guida e sintesi delle evidenze scientifiche. Ha 25 anni di esperienza nel campo della valutazione dei medicinali e delle politiche sanitarie. È stato membro fondatore del Centro Cochrane Italiano, ha fatto parte dell'Editorial Board del British Medical Journal e per oltre 10 anni è stato Presidente del Comitato Etico di Reggio Emilia (2000 -2012). Assieme ad Alessandro Liberati, ha fondato e diretto il Ceveas (Centro per la Valutazione della Efficacia della Assistenza Sanitaria) dell'Azienda USL di Modena (2000 – 2012) ed è stato direttore dal 2008 del Centro Collaborativo OMS per la sintesi delle evidenze scientifiche e la metodologia delle linee guida. È stato membro del Comitato Etico dell'OMS e Segretario della Lista dei Farmaci Essenziali dell'OMS dal 2014 al 2020 a Ginevra. Direttore Generale dell'AIFA dal marzo 2020 fino al gennaio 2023. È attualmente Direttore del Governo Clinico dell'ASL Romagna e co-direttore del WHO Collaborating Centre for evidence synthesis and guideline development.

PAOLA MARION

È laureata in Filosofia, membro ordinario con funzioni di training Società Psicoanalitica Italiana (SPI) e International Psychoanalytic Association (IPA), ed è specializzata in psicoterapia dell'infanzia e dell'adolescenza. Esercita la professione privata. Svolge attività didattica presso il Training psicoanalitico della SPI e presso il Corso di psicoterapia del bambino, dell'adolescente e della coppia (ASNE-SIPSIA). È stata Segretario per le Relazioni Internazionali nell'Esecutivo SPI (2009-2013) e Chair per l'Europa dell'Outreach Committee dell'IPA (2011-2013), Direttore della Rivista di Psicoanalisi (2017- 2021). Tra le sue pubblicazioni il volume *Il disagio del desiderio. Sessualità e procreazione al tempo delle biotecnologie* (Donzelli, 2107, uscito nella traduzione inglese, Routledge, nel 2021).

GIACOMO MARRAMAO

Formatosi nelle università di Firenze e di Francoforte, è Professore Emerito di Filosofia teoretica presso l'Università degli Studi Roma Tre. È inoltre coordinatore del Comitato Scientifico della Fondazione Basso e membro del Comitato d'Onore del Collège International de Philosophie di Parigi (dove ha insegnato anche Teoria politica a SciencesPo). Visiting professor in diverse università europee, americane e asiatiche, si è concentrato nei suoi lavori degli ultimi anni su tre ambiti tematici: 1) il concetto di tempo (lungo la linea di confine tra filosofia, scienza e teologia); 2) il rapporto tra politica e secolarizzazione; 3) i mutamenti di forma del potere nell'epoca della globalizzazione.

EUGENIO MAZZARELLA

È professore emerito di Filosofia teoretica presso l'Università di Napoli "Federico II", di cui è stato preside della Facoltà di Lettere e Filosofia. Ha fatto parte della Commissione Cultura della Camera dei deputati nella XVI Legislatura. Ontologia, tecnica e antropologia sono i suoi principali interessi di ricerca.

LUCIANO ORSI

Medico, è stato Responsabile della Terapia Intensiva e del Dipartimento Emergenza e Accettazione dell'Ospedale Maggiore di Crema fino al 2002;

dal 2002 al 2010 ha diretto la Rete di Cure Palliative di Crema; dal 2011 al 2016 è stato Direttore della Struttura Complessa Cure Palliative della Azienda Ospedaliera Carlo Poma di Mantova e direttore del Dipartimento interaziendale di Cure Palliative di Mantova. Ha fatto parte del Comitato Didattico della Scuola Italiana di Medicina Palliativa ed è socio fondatore e membro della Commissione di Bioetica della Società scientifica di Anestesia, Analgesia, Rianimazione e Terapia Intensiva (SIAARTI). Dal luglio 2011 è Direttore della Rivista Italiana di Cure Palliative e dal novembre 2016-2020 è stato Vice presidente della Società Italiana di Cure Palliative.

LAURA PALAZZANI

È Professore ordinario di Filosofia del diritto presso la LUMSA di Roma. È stata membro del Comitato Nazionale per la Bioetica (2002-2007) e vicepresidente (2008-2022). È componente dell'European Group on Ethics in Science and New Technologies presso la Commissione europea (dal 2011), componente dell'International Committee for Bioethics presso l'Unesco (dal 2016) e delegata del governo italiano presso il Comitato di Bioetica del Consiglio d'Europa (dal 2015).

SILVANO PETROSINO

È Professore ordinario di Filosofia teoretica presso l'Università Cattolica del Sacro Cuore di Milano. Presso questa stessa Università attualmente insegna Teorie della Comunicazione e Antropologia religiosa e media. È inoltre titolare del corso di Antropologia del Sacro presso l'Istituto Superiore di Scienze Religiose di Milano. Tra le sue ultime pubblicazioni: *L'idolo. Teoria di una tentazione. Dalla Bibbia a Lacan* (Mimesis 2015), *Emmanuel Levinas. Le due sapienze* (Feltrinelli 2017), *Contro la cultura. La letteratura, per fortuna* (Vita e Pensiero 2017), *Il desiderio. Non siamo figli delle stelle* (Vita e Pensiero 2019), *Dove abita l'Infinito. Trascendenza, potere e giustizia* (Vita e Pensiero 2020).

ALBERTO PIRNI

È Professore associato di Filosofia morale presso la Scuola Superiore Sant'Anna di Studi Universitari e di Perfezionamento – Pisa, dove insegna Etica pubblica, Etica delle relazioni di cura, Etica ed economia,

Ethics and Security, Intergenerational Justice. È coordinatore dell'Area di Ricerca in "Ethics and Global Challenges" dell'Istituto di Diritto, Politica e Sviluppo. È stato vice-Presidente del Comitato Etico Congiunto tra Scuola Superiore Sant'Anna e Scuola Normale Superiore ed è ora coordinatore del Curriculum 4 ("Theories, Institutions and Cultures of Ecological Transition") della Scuola di Dottorato Nazionale in "Sustainable Development and Climate Change". Si occupa di filosofia classica tedesca, etica e tecnologie emergenti, etica per le istituzioni, etica e giustizia climatica.

S.E. MONS. ANTONINO RASPANTI

Vescovo di Acireale (CT) e presidente della Conferenza Episcopale Siciliana. È stato professore ordinario di Teologia presso la Facoltà Teologica di Sicilia, della quale è stato anche preside.

ANTONELLA SCIARRONE ALIBRANDI

Professore ordinario di Diritto dell'economia nella Facoltà di Scienze bancarie, finanziarie e assicurative dell'Università Cattolica del Sacro Cuore, in aspettativa dal 1° gennaio 2023 in seguito alla nomina a Sottosegretario del Dicastero per la Cultura e l'Educazione della Santa Sede. È membro dell'Autorità di Supervisione e Informazione Finanziaria (ASIF) della Santa Sede e dell'Academic Board dello European Banking Institute (EBI). Ha fatto parte dell'Expert Group ROFIEG (Expert Group on Regulatory Obstacles to Financial Innovation), costituito nel 2018 dalla Commissione europea presso DG FISMA.

GIUSEPPE VACCA

Ha insegnato Storia delle dottrine politiche presso l'Università di Bari ed è stato deputato al Parlamento eletto nelle liste del Pci nel collegio Bari-Foggia dal 1983 al 1992. Dal 1988 al 2016 è stato direttore e poi Presidente della Fondazione Gramsci di Roma e negli stessi anni membro del Consiglio di amministrazione dell'Istituto della Enciclopedia Italiana Treccani. Presiede la Commissione scientifica dell'Edizione nazionale degli scritti di Antonio Gramsci. Fra i suoi libri più noti, *Politica e filosofia in Bertrando Spaventa* (Laterza 1967), *Saggio su Togliatti e la tradizione*

comunista (De Donato 1975), Il marxismo e gli intellettuali (Editori Riuniti 1985), Pensare il mondo nuovo. Verso la democrazia del XXI secolo (Edizioni San Paolo 1994); fra i più recenti ricordiamo Vita e pensieri di Antonio Gramsci. 1926-1937 (Einaudi 2012), L'Italia contesa. Comunisti e democristiani nel lungo dopoguerra (1943-1978) (Marsilio 2017), Modernità alternative. Il Novecento di Antonio Gramsci (Einaudi 2017), La sfida di Gorbaciov. Guerra e pace nell'era globale (Salerno 2019) e Il comunismo italiano. Una cultura politica del Novecento (Carocci 2021).

STEFANO ZAMAGNI

Insegna Economia all'Università di Bologna ed è Adjunct Professor di International Economics al SAIS Europe della Johns Hopkins University. È stato presidente della Pontifical Academy of Social Sciences (Vatican City). Presiede il Comitato scientifico della SEC (Scuola di Economia Civile) e di AICCON (Associazione Italiana per la promozione della Cultura della Cooperazione e del Nonprofit). Presidente della Fondazione LUMSA Human Academy. È membro della Consulta del Cortile dei Gentili dal 2016.

Altri autori

MONS. CARLO MARIA POLVANI

Ha conseguito laurea e dottorato in Biochimica alla *McGill University* e ottenuto un *Master of Divinity* alla *Weston School of Theology*. Dopo la Licenza in Diritto Canonico alla Pontificia Università Gregoriana, si è specializzato in Giurisprudenza e Psicologia Forense e ha conseguito un dottorato in Diritto Canonico. È stato ordinato presbitero per l'Arcidiocesi di Milano e Prelato d'Onore di Sua Santità. Ha frequentato la Pontificia Accademia Ecclesiastica ed è stato ammesso nel Servizio Diplomatico della Santa Sede. Ha lavorato in Segreteria di Stato come Responsabile dell'Ufficio Informazione e Documentazione e dell'Ufficio Tecnico. Dal 2019 è Sottosegretario del Pontificio Consiglio della Cultura.

“In questo volumetto il lettore non troverà posizioni luddiste, e nessun vade retro, nei confronti dell’innovazione, e quindi della stessa Intelligenza Artificiale. Al contrario, di essa vedrà unanimemente condiviso il formidabile contributo ai nostri progressi nei campi più diversi. Ma vedrà anche indicate, sempre unanimemente, le aree di limiti possibili ed anzi necessari, con il chiarimento, non meno necessario, che alcuni di tali limiti tocca al diritto stabilirli, mentre altri non possono che stare nella coscienza dei ricercatori.”

“Se tutto questo è sufficientemente condiviso, se da un lato la società con le sue leggi, dall’altro la ricerca con il suo codice morale sapranno tenere le barre loro affidate, potremo con tranquillità fidarci dell’IA: avvalerci dei suoi benefici, non drammatizzare le nostre difficoltà nel distinguere l’Incompiuta di Schubert da quella in parte compiuta dal computer, imparare a capire, parlandoci, quando uno studente ha scritto il suo paper, quando se lo è fatto scrivere da ChatGPT. Non sarà questo a portarci nelle acque dove, secondo padre Dante, finirebbe per richiudersi sopra di noi il mare. Quelle acque sono più avanti e tocca a noi, a tutti noi, evitare che l’umanità ci arrivi.”

Dalla prefazione di Giuliano Amato

La Consulta Scientifica è un organo permanente della Fondazione Cortile dei Gentili, che lavora in stretta connessione con il Dicastero per la Cultura e l’Educazione. È composta da studiosi e personalità di spicco che, grazie alle loro professionalità, esperienze e diversi ambiti di competenza, indirizzano con argomentazioni scientifiche le iniziative del “Cortile dei Gentili” per favorire il dialogo tra credenti e non credenti. Il contributo della Consulta consente di approfondire i grandi temi della filosofia, dell’antropologia, della scienza, del diritto, dell’economia e della cultura in generale.

€ 10,00



9 788880 805588