IL PERTURBANTE ARTIFICIALE
LIVELLI DI COMPLESSITÀ E INTERFACCE POSSIBILI FRA UMANO E IA

di Aldo Pisano

Sei solo un accenno di ciò che era lui. Non hai nessuna storia.
Sei l'interprete di qualcosa che lui faceva senza pensare, non può bastarmi ciò che sei!
Black Mirror, 2x01, Torna da me

Abstract
Aim of this paper is to analyse interface's different levels between humans and artificial intelligence, assuming human being as a cognitive system organized on many levels of complexity. Starting from a first interface level linked to the concept of "intelligence", we'll try to figure out why AI can't reach the same condition of human intelligence and complexity. Nevertheless, AI can be assumed as a different cognitive system. At the same time, we'll try to understand how an increased level of complexity implies enhanced AI's interfaces in order to create a relationship with humans. However, if this interface is too similar to humans (e.g. AI body-endowed) and according to Masahiro Mori's uncanny valley, this machine can create a distance between human and AI. If every AI has its own objective in different fields of application, an "uncanny" interface can be unproductive for human's purposes.

1. *La spinta mimetica*

Fino a che punto si spinga oggi l'imitazione dell'essere umano con lo sviluppo delle intelligenze artificiali diventa sempre più evidente. Giornalmente si palesano gli sviluppi che inducono a una riflessione sulle conseguenze, oscillante fra il principio speranza (Bloch 2019) e il principio disperazione (Anders 2007), da intendersi come approcci polarizzati rispetto alle proiezioni sul futuro avanzamento della tecnica. Il futuro diventa sempre più incerto e mentre i fatalisti dipingono una situazione in cui l'umanità soccombe alle super-intelligenze artificiali, gli integrati si mostrano più moderati nelle proiezioni, continuando a sostenere l'idea che l'IA manterrà una funzione di supporto rispetto all'attività umana.

In questi termini, l'analisi del rapporto umano-IA diventa oggetto di indagine per andare incontro a un disciplinamento etico-giuridico ma anche per evitare forme di riduzionismo. Affinché questo sia possibile, dato che l'intelligenza umana rappresenta il modello di partenza per la costruzione delle intelligenze artificiali, è necessario tracciare una linea di confine fra ciò che è imitabile e ciò che non lo è. Questo, in riferimento al valore della complessità sistemica (Capra & Luisi 2020) dell'essere umano, articolabile sui livelli di (i) relativa, (ii) totale e (iii) assoluta. Il primo livello si riferisce in senso disgiunto alla dimensione teoretica o alla dimensione pratica. Il secondo livello è da riferirsi congiuntamente alla sfera teoretico-conoscitiva *e* alla sfera etico-pratica, quindi valutando la relazione sussistente fra la conoscenza e l'esperienza del mondo, soprattutto se la si declina in senso etico. Il terzo livello considera sempre congiuntamente la sfera teoretico-conoscitiva e la sfera etico-pratica, ma con l'aggiunta di due componenti propriamente umane: le emozioni e il corpo. Con l'aumento della complessità,

Data di pubblicazione: 15.09.2021

ovviamente, si ha un aumento delle variabili poste in gioco sia nei processi conoscitivi, sia in quelli pratico-etici.

Qui si prenderà in considerazione soprattutto la dimensione teoretica del primo livello, partendo dal pensiero computazionale come primaria analogia della macchina rispetto ai processi conoscitivi umani. Si sottolinea, infatti, che l'analisi dei meccanismi di codifica ed elaborazione dell'informazione non rappresenta altro che un primo passo verso un'intelligenza imitatrice. È possibile fare questo rintracciando certo delle analogie, ma con il pericolo di un appiattimento della complessità cognitiva; per cui «il machine learning è la spada con cui uccidere il mostro della complessità» (Domingos 2020, 29).

Il grafico di Masahiro Mori (1970) tenta di restituire l'andamento della complessità, partendo dalla relazione analogica fra umano e IA e in riferimento alle modalità di interfaccia: dalla semplice riproduzione vocale alla robotica sociale che permette di dare corpo all'IA. Il grafico di Mori tenta di disegnare il limite ultimo fra la familiarità rassicurante e la deriva perturbante (Freud 1989)[1], muovendo dalle intelligenze artificiali deboli a quelle forti, qui restituibile nei termini di un'articolazione fra la complessità relativa e quella assoluta.

Questo determina una necessaria rimodulazione di stili cognitivi eterogenei, validi in ambiti applicativi diversi: dalla domotica all'impiego bellico dell'IA. La *uncanny valley* (valle del perturbante) di Mori, di fatto, rappresenta il limite ultimo in cui l'automa diventa autonomo, destrutturandosi dalla funzione assegnata e presentandosi come una pseudo-individualità, da cui la cessazione di ogni sua attività semplicemente supportiva.

Il presupposto è che una lenta e graduale acquisizione di abilità diverse comporta un sempre maggiore avvicinamento della struttura e della forma dell'intelligenza artificiale all'essere umano, fino a produrre un paradossale allontanamento. Inoltre, ogni tentativo di imitazione o simulazione si presenta come una riduzione epistemologica, in cui il fattore complessità rimane ontologicamente tale tracciando il termine dell'inimitabilità.

Si procede, dunque, da un primo livello di analisi che riguarda le analogie tra processi cognitivi umani e artificiali, partendo dal pensiero computazionale, dunque dalla definizione di IA per come proposta da Russell e Norvig in riferimento al rapporto conoscenza-azione-obiettivo (Russell & Norvig 2010) e il ruolo dell'essere umano nei processi di programmazione. Questo in riferimento all'analisi del livello di complessità relativo-teoretico, disgiunto da quello pratico.

Segue un'analisi della situazione che si prospetta con l'aumento della complessità della macchina. In questo senso, si passa da uno stile cognitivo base che opera secondo codice binario e in termini statistici (*machine learning*) a uno più complesso (*deep learning*), in cui i livelli e le connessioni utili alla rielaborazione aumentano, in quanto incrementa il numero di variabili e quindi il livello di complessità della risposta.

La macchina arriva così a processi di auto-modifica, lanciando l'IA verso un'autonomizzazione che la rende sempre più simile all'umano e, paradossalmente, sempre più distanze. Si prenderà in considerazione il pensiero incarnato (*embodied cognition*; Fuchs 2021), come e se quest'ultimo sia possibile da replicare come simulazione di una complessità assoluta e quindi come intelligenza in senso pieno.

Il livello di autonomia, dunque, implica una sorta di individualizzazione che fa della macchina un "altro" in senso assoluto, soprattutto considerando l'intervento dell'ingegneria e della robotica che conferiscono corporeità e tratti somatici simili a quelli umani (Rossi 2019). Un approccio in questo stile permette di comprendere che il livello di interazione fra uomo e macchina aumenta con la successiva articolazione dell'interfaccia fra i due. Anche la sola riproduzione vocale, nella forma di una dizione

[1] A questo proposito si rinvia a Carotenuto 2002; Rank 1971.

affettata, comporta un primo livello di dissonanza che produce straniamento. Il grafico di Mori può essere dunque interpretato come descrittore del graduale aumento del perturbante.

2. *Le analogie computazionali*

Il primo livello di complessità (relativo), anche se circoscritto alla dimensione teoretica – da riferire qui ai processi cognitivi - costituisce già di per sé un sistema olistico complesso che lavora su più livelli: recezione, immagazzinamento, elaborazione delle informazioni e risposte. Partendo da un modello di tradizione empirista, la prima articolazione interna ai processi cognitivi di base – nel momento in cui tenta di accostare il modello umano a quello artificiale – viene ridotta a processi computazionali. Si potrebbe dire, in lessico kantiano, che la produzione di giudizi viene qui arginata al semplice piano aprioristico (Kant 2007)[2]. Il livello qui proposto, infatti, è quello che guarda alla macchina come un agente analitico, quindi consistente in «alcune linee di codice, un programma per computer che può esistere in moltissime copie […]. Un simile agente artificiale, non essendo un individuo o un oggetto fisico tridimensionale, in un certo senso non è nemmeno una cosa» (Dumouchel & Damiano 2019, 190). Questi agenti analitici, in definitiva «corrispondono a un momento dell'analisi di un processo di azione complessa» (*ibidem*). I *learners*, ossia le unità-base che compongono l'algoritmo, sono legati alla conoscenza e non alle operazioni concrete: «nel machine learning la conoscenza assume spesso la forma di modelli statistici, perché gran parte della conoscenza è di natura statistica» (Domingos 2020, 30). Se lavorare sull'IA «costituisce anche un'indagine sperimentale sulla natura della mente» (Dumouchel & Damiano 2019, 25), di certo, il riduzionismo eccessivo della dimensione teoretico-conoscitiva è una delle potenziali derive verso cui sta conducendo l'ipertrofica diffusione del modello computazionale. Eppure, questo è il modello di partenza su cui si costruisce l'analogia intellettiva fra uomo e macchina. Da questo deriva un'idea di antropomorfismo «che cerca, sì, di modellarsi sulle nostre capacità cognitive, ma le usa come leve al fine di elaborare meccanismi che, traendo ispirazione dai nostri schemi cerebrali, son destinati a essere più rapidi, efficaci e affidabili di quelli che ci costituiscono, rimanendo tendenzialmente inalterati» (Sadin 2019, 11). Questo livello di partenza degli agenti analitici mira a implementare il *coding* abbinandolo a capacità esecutive (come si vedrà a breve) per raggiungere determinati obiettivi: da cui l'evoluzione dall'idea di pensiero computazionale a intelligenza vera e propria.

Il punto, però, rimane l'intrinseca difficoltà di rendere spiegabili questi stessi processi che guidano la macchina; questa, infatti, sarà sempre condannata a un'«opacità epistemica» (Dumouchel & Damiano 2019, 81). La possibilità di definire la macchina un'intelligenza decade di fronte al fatto che essa è capace di conoscere (statisticamente), ma non di pensare consapevolmente con tutto ciò che questo comporta: il volere, il porsi un obiettivo, lo spiegare le proprie azioni. Rimane, dunque, uno scarto teoretico non indifferente fra il conoscere e il pensare[3]. Quest'ultimo si configura come un'attività

[2] A questo proposito si rinvia a Sacchetto 1989; Marcucci 1997; Dicker 2004; Ciafardone 2007; Arendt 2009.

[3] L'analisi sull'attività del pensiero in senso arendtiano riconduce la riflessione verso il nesso indistricabile fra conoscenza ed esperienza del mondo, come modello di etica sintetica per cui il robot sociale può diventare, di fatto, un attore che co-evolve con l'essere umano, permettendo una chiarificazione delle questioni conoscitive e morali ancora oscure. Scrivono Dumouchel e Damiano: «Proponiamo di chiamare "etica sintetica" l'indagine sulle sfide morali inerenti alla coevoluzione di attori umani e attori artificiali. Si

squisitamente umana, qui recuperabile nella precisa lettura arendtiana di Platone: «il filosofo, anche se decide con Platone di lasciare la "caverna" delle faccende umane, non deve nascondersi da sé stesso; al contrario, sotto il cielo delle idee egli non solo trova la vera essenza di tutto ciò che è, ma anche sé stesso, nel dialogo tra "me e me stesso" (*eme emautō*) in cui Platone individuò evidentemente l'essenza del pensiero» (Arendt 2014b, 55)[4].

Il *machine learning* oggi inverte la tendenza. Il punto non è più avvicinare la macchina all'uomo, ma quanto l'uomo sia simile alla macchina per derivarne i processi intellettivi imitabili.

L'effetto di quest'inclinazione non è l'implementazione della macchina, ma un piegare il pensiero umano sul livello computazionale e analitico, in una parola: semplificarlo. La conoscenza analitica produce maggiore prevedibilità, sicurezza, possibilità di riuscita. L'algoritmo, tuttavia, nella sua tendenza conoscitivo-statistica-imitativa rimane solo una serie di istruzioni che dice a un computer cosa fare (Domingos 2020). Infatti, un algoritmo semplice, come processo mentale base viene ricondotto a tre operazioni "AND", "OR" oppure "NOT" (*ibidem*). Proprio qui entra in gioco il criterio di prevedibilità, con l'aggiunta, però, di un calcolo non più deterministico ma probabilistico: questo è anche il passaggio dall'informatica al *machine learning*.

Dunque il primo livello di connessione umano-IA si configura nei processi computazionali che costituiscono l'operatore logico primario su cui fondare le possibili analogie fra stili cognitivi diversi, seppure simmetrici. Si consideri che il pensiero computazionale trova una sua prima importante strutturazione nel modello dell'HIP (*Human Information Processing*), un quadro di riferimento teorico che si riferisce a quattro macro processi: elaborazione delle informazioni, analisi del compito, uso di una metodologia sperimentale e auto-modificazione (Lindsay & Norman 1987). Come si diceva, una tale attribuzione ritaglia una componente del sistema-complessità rintracciabile nella sua disgiunzione dalla sfera etico-pratica, in quanto non contaminata da questa, dunque forte di meccanismi che rinviano a una logica univoca, di stimolo-ricezione dell'informazione e alla sua successiva elaborazione in termini di logica binaria, analitica, aprioristica.

I processi cognitivi che si reggono sul pensiero logico-matematico, seriale, calcolante e basato su una statistica, si configurano anche in base alla distanza che prendono da altre forme di pensiero non-strutturato come quello narrativo (Bruner 1992), qui da intendersi come attribuzione significativa al mero dato, così che possa essere interpretato o perché gli venga conferito valore.

Quest'architettura della complessità relativo-teoretica, dunque, è qui ritagliata per essere oggetto di imitazione, non potendo però essere ricostruita nella sua valenza attributivo-esistenziale, quindi come pensiero simbolico-valoriale. Dunque, partendo da questo livello si riconosce il valore dell'IA come intelligenza non sintattica: la macchina può conoscere, ma non pensare (Arendt 2014a) o attribuire valore ai fatti (Nida-Rümelin & Weidenfeld 2018). Rimanendo su questo profilo, lo scambio è minimo – a livello commerciale queste tipologie di IA sono le più ricercate in quanto realizzano meglio «il

---

tratta di un progetto, insieme etico ed epistemologico […] etica sintetica perché – (anche) in questo emergente contesto di coevoluzione – il conoscere è inseparabile dal fare» (Dumouchel & Damiano 2019, 30). Questa linea dell'etica sintetica, di fatto, si origina spontaneamente accedendo al livello di complessità totale; qui essenziale è l'intreccio fra dimensione teoretica e dimensione etica.

[4] Si ricorda qui che, la stessa Hannah Arendt, nell'Introduzione a *The Human Condition* sminuiva il pericolo della tecnica, confinando la macchina proprio sul livello della conoscenza (accumulo quantitativo di dati) e non su quello del pensiero, qui inteso come attività produttiva e libera (Arendt, 2014a). Sul tema si rinvia a Arendt 2005, 2006, 2009, 2010, 2014b; Forti 2006.

duplice obiettivo dell'elevata prestazione tecnica e della massima evanescenza fisica» (Dumouchel & Damiano 2019, 44). Una constatazione di questo genere diventa rilevante, in quanto mostra che l'IA parte da un livello di interfaccia/presenza con l'essere umano che è completamente dis-incorporata o non-incarnata.

Questo sottolinea come il perturbante inizi già qui a manifestarsi, ma su una soglia minima come nel caso degli assistenti vocali. Anche solo il valore ergonomico dello *smartphone* muove oltre il limite della complessità relativa costituendo una prima forma di «contatto carnale» (Sadin 2019, 51). Gli stessi assistenti vocali, così vicini e così lontani dall'essere umano, si trovano spinti verso una complessità che è drasticamente ridotta, annichilita da processi ancora eteronomi. C'è, infatti, una completa dipendenza cognitiva dall'essere umano. Come scrive Stuart Russell, in riferimento al *PDA* (*Persona Digital Assistant*): «gli input includono non solo il segnale acustico dal microfono […] e gli input dal *touch screen* ma anche il contenuto da ogni pagina Web a cui il dispositivo accede, mentre le azioni includono sia parlare che mostrare materiale sullo schermo (Russell 2020, 43).

La macchina compie quindi un lavoro che parte dall'immissione di dati, li analizza e stabilisce relazioni statistiche formulando conoscenze analitiche. Ora, rimane il punto di un'impossibilità di sganciamento della macchina dall'essere umano o da altro, in quanto i dati sono sempre etero-riferiti, non auto-prodotti – anzi nel caso della complessità totale anche estrapolati dall'esterno da strutture recettive artificiali. È la conoscenza sintetica, invece, che costituisce un processo del tutto autonomo e che, a priori o a posteriori, comporta la produzione di nuove conoscenze. In questo senso non esiste intelligenza che per dirsi autonoma non derivi da una condizione eteronoma[5]. Un discorso simile vale sia dal punto di vista teoretico/cognitivo sia dal punto di vista etico/pratico.

L'attività di sintesi contraddistingue quella forma del pensiero vista in precedenza con Arendt e che spinge avanti le conoscenze e le azioni; essa coincide con autonomia e libertà: è esperienza di cominciamento. Per rimanere sempre nel registro arendtiano, si può dire che la conoscenza (accumulazione quantitativo-statistica di dati) sta al lavoro, come il pensiero (attività produttiva e libera) sta all'azione (Arendt 2014b).

Si consideri, inoltre, che la macchina non può auto-conferirsi degli stati mentali, per cui l'autonomia rimane sempre relativa e relazionale rispetto al sistema individuo-ambiente (Dumouchel & Damiano 2019)[6].

In questo processo di sintesi dei dati, mentre i sistemi cognitivi, di fatto, commettono degli errori, i sistemi meccanici, invece «vanno in panne, non fanno errori» (Dumouchel & Damiano 2019, 87). Questo è il primo elemento che contraddistingue l'errore antropocentrico che si commette sul livello di complessità relativa. La mente computazionale costituisce un modello possibile, ma non un marchio da imprimere e in cui esaurire il concetto di intelligenza. Questo permette già di definire un distanziamento tra il modello macchina e il modello umano. Di certo, l'IA, anche nella sua forma più

[5] Analizzando la questione dal punto di vista etimologico, così si esprime Benveniste: «Al suo appartenere al gruppo – di nascita o di amici – l'individuo deve non solo l'essere libero, ma anche l'essere *sé stesso*: i derivati del termine **swe*, gr. *idiōtes* "privato", lat. *suus* "suo", ma anche gr. *étēs*, *hetâiros* "alleato, compagno", lat. *sodalis* "compagno, collega", lasciano intravvedere nello **swe* primitivo il nome di un'unità sociale in cui ogni membro scopre "sé stesso" solo nel suo "essere con gli altri". È quindi chiaro che la nozione di "libertà" si costituisce a partire dalla nozione socializzata di "crescita", crescita di una categoria sociale, sviluppo di una comunità. Tutti quelli che sono usciti da questa "matrice" da questo "ceppo", hanno la qualità di **(e)leudheros* (da cui il gr. Eleútheros) […]. Il senso primitivo non è, come si sarebbe tentati di pensare, "liberato da qualcosa"; è quello di appartenenza a una razza etnica designata con una metafora di crescita vegetale. Questa appartenenza conferisce un privilegio che lo straniero e lo schiavo non conoscono mai» (Benveniste 2001, 247-249).

[6] I due autori rinviano a Maturana & Varela, 1984; Ceruti, 2007; Damiano, 2009.

debole, ha la sua funzionalità, la sua velocità e i suoi risvolti positivi ma non per questo va catalogata immediatamente come intelligenza – questo tipo di riflessione è il residuo antropico nell'opera di implementazione delle intelligenze artificiali. Allo stesso tempo, però, si presenta come limite necessario: pensare e creare un nuovo sistema intelligente non può prescindere dal modello umano al quale continuamente viene adattato.

Questo rimodellare l'IA parte sempre da esigenze umane e sul modello umano ritorna, toccando quindi la soglia del perturbante. Scrive Stuart Russell, riproponendo il primo approccio antropomorfo all'IA: «una macchina è intelligente nella misura in cui agendo riesce a ottenere ciò che vuole, in base a ciò che ha percepito» (Russell 2020, 41).

Attenzione, ancora qui interviene il fattore eteronomo: il punto non è cosa *è voluto*, ma cosa *è dato* come percezione; questo significa che l'aspetto volitivo della macchina è assente, il che è già un motivo che non permette di catalogarla come intelligenza (Borghini & Casetta 2013). La tensione desiderativa è una delle condizioni proprie dell'essere vivente nella classificazione delle specie (Aristotele 2001). La successiva articolazione in finalità più complesse (bisogno di socialità, bisogno di espressione, di educazione) che si può ricondurre a Maslow (Allport et al. 1970), caratterizza propriamente la specie umana. È la "spinta interna" che guida e stabilisce la relazione fra azione e obiettivo – da cui la definizione di intelligenza offerta da Russell: «Le macchine sono intelligenti nella misura in cui dalle loro azioni ci si aspetta che raggiungano i propri obiettivi» (Russell 2020, 9). La macchina non elabora degli obiettivi da raggiungere per cui predisporre la propria azione. Da questa constatazione, la nuova definizione che riporta la macchina verso lo *status* di entità non-autonoma, ma da cui l'uomo può trarre beneficio: «Le macchine sono utili nella misura in cui ci si può aspettare che le loro azioni raggiungano i *nostri* obiettivi»[7] (Russell 2020,11).

Se, dunque, un'entità è priva della spinta desiderativa (nelle sue varie forme) utile a fare sì che si prefigga quest'obiettivo e fino a che il processo esecutivo rimane opaco (Tamburrini 2020), allora non si può ricondurre tutto questo al concetto di intelligenza. Lo si può, invece, riportare su quel piano dell'analogia conoscitivo-computazionale che parte da un accumulo statistico di dati, tale da fornire strumenti più o meno idonei perché la macchina possa raggiungere degli scopi imposti dall'esterno, dal programmatore o dall'utente.

Tuttavia, aumentando il sistema-ambiente in termini di complessità, la possibilità che l'obiettivo da raggiungere vada a scemare è altissima: «la logica richiede certezze, e il mondo reale semplicemente non le fornisce» (Russell 2020, 42). Per questo la complessità rimane il criterio di misurazione più rilevante nella discriminazione uomo-macchina e nel limite imposto al lemma "intelligenza", come residuo di una tradizione che porta con sé un'analogia assolutamente parziale. L'intelligenza è attribuzione di complessità alla relazione fra individuo e ambiente, che non può esaurirsi nel pensiero calcolante. Questo non solo non rende la macchina simile all'uomo ma delegittima l'utilizzo del lemma "intelligenza".

L'intelligenza è quindi da intendersi come comprensione razionale e possibilità di spiegazione, questa non si ferma al livello della semplice descrizione e conoscenza dei fenomeni interni ed esterni al soggetto. Quando Dilthey proponeva la differenza fra *Geisteswisseshaften* e le *Naturwissenshaften*, rendeva evidente questo paradigma nel sostanziale divario fra *verstehen* ed *erklären* (Dilthey 2007) scindendo il calcolo dei fenomeni dal valore del mondo come esperienza vissuta, come intreccio simbolico irriducibile a un giudizio analitico o sintetico a priori.

[7] Corsivo mio.

Di fronte alla complessità e all'aumento delle variabili, di fronte all'incapacità di gestione, il sistema entra facilmente in panne. Esistono elementi non calcolabili pur rimanendo fermi sul livello di complessità relativa e che la logica non riesce a esaurire – infatti l'interezza delle relazioni individuo-ambiente non può essere mai calcolata completamente da un algoritmo. In questo senso, come analizza sempre Russell, l'ignoranza è il problema insuperabile di un sistema che è puramente formale (e non solo) (Russell 2020). Per questo la programmazione dell'IA ne determina lo scopo, sempre in forma strettamente dipendente dall'essere umano; ma per potere raggiungere l'obiettivo, è necessario immettere dei dati per aiutare la macchina a orientarsi nel mondo. Dunque, le variabili per la costruzione di un agente intelligente determinano lo scopo per cui è costruito e queste sono: (i) ambiente; (ii) osservazione e azione; (iii) obiettivo. Tutto questo permette di attribuire alla macchina una certa efficienza, ma di certo non una razionalità pratica (Aristotele 2013; Vaccarezza 2012).

Assumendo il paradigma dell'intelligenza come non strettamente connessa alla dimensione conoscitivo-teoretica, ma riferita anche alla relazione fra azione e obiettivi, è qui necessario uno slittamento verso un livello di complessità che supera quello relativo e muove verso la complessità totale. Assumere questa linea significa comprendere che l'IA evolve verso processi cognitivi più complessi di rielaborazione ed esecuzione, qui riferiti all'incremento dell'interazione con altri agenti e quindi all'aumento della complessità dell'interfaccia uomo-macchina.

Non è possibile non considerare la complessità anche negli aspetti più semplici che riguardano agenti analitici; questo permette di misurare il livello di compatibilità umana, di considerare quale sia il livello di interfaccia che rende simile la macchina rispetto all'uomo. Scrive Russell: «Possiamo dire che un algoritmo è generale perché abbiamo prove matematiche del fatto che esso fornisce risultati ottimali o quasi-ottimali; questo considerando una ragionevole complessità computazionale riferita a un'intera classe di problemi e proprio perché funziona nella pratica su quegli stessi problemi senza avere bisogno di [continue] modifiche per risolverli» (Russell 2020, 45). Perché possa interfacciarsi produttivamente con l'umano, l'IA deve aumentare il proprio livello di complessità, senza superare una determinata soglia.

La pluralità delle intelligenze artificiali è tale perché diversi sono gli scopi e le modalità di interfaccia per cui vengono programmate, il che significa non ricondurre l'IA all'intelligenza umana ma all'idea di pluralismo del mentale – quindi sulla possibilità di uno stile cognitivo completamente diversificato rispetto a un approccio che sarebbe quello dell'antropocentrismo cognitivo e che vede l'intelligenza umana come unico modello intelligente. Questo non solo evita di dovere riportare l'intelligenza artificiale su quella umana, ma anche di ridurre l'essere umano alla macchina.

3. *La complessità totale: verso il grafico di Mori*

Stando alla ricostruzione di Dumouchel e Damiano, il primo livello di complessità relativo corrisponde all'idea di agente analitico. Qui, invece, considerando la complessità totale in cui è possibile inserire l'idea di esperienza del mondo, di interazione base dell'IA con l'ambiente esterno per cui riesce nei processi di auto-modifica e a produrre effetti, si parla di agenti esecutivi; ossia, un agente analitico inserito in un contesto particolare. Se l'agente è isolato, il processo di elaborazione è assolutamente autoreferenziale (analitico, aprioristico), mentre quando l'agente è collocato in un ambiente potrà dirsi esecutivo sempre in virtù del contesto in cui opera. Questa relazione con l'esterno determina le modalità di azione/risposta.

L'idea è quella di *learners* procedurali che, tuttavia, non possono non basarsi sulla conoscenza: «Spesso le procedure sono semplici, ed è la conoscenza su cui si fondano che è complessa» (Domingos 2020, 30). Dare la possibilità al semplice *machine learning* di potersi auto-modificare sempre meglio, partendo dall'apprendimento (prova-errore) e quindi dal contatto con l'esterno, definisce il passaggio ulteriore dell'IA: « L'apprendimento conferisce un grande vantaggio evolutivo […]. L'apprendimento è un bene molto più che per sopravvivere e prosperare. Esso accelera l'evoluzione (Russell 2020, 18-19).

Ora, relazioni differenziate tra organismo e ambiente determinano uno stile cognitivo altrettanto differenziato. Se si considera l'idea del pluralismo cognitivo – partendo dal modello della *embodied cognition*[8] – allora questo comporta una considerazione che apre all'IA come un modello cognitivo a sé stante: «l'ipotesi della diversità dei sistemi cognitivi, insieme ai risultati dell'etologia artificiale, supporta la tesi che la mente animale sia incorporata, situata e irriducibilmente locale, invitandoci a rivisitare una teoria filosofica di vecchia data, inseparabile dalla prima formulazione moderna della filosofia della mente» (Dumouchel & Damiano 2019, 70). Da questa prima constatazione deriva il valore differenziato fra specie simbolica, sub-simbolica e varietà degli stili cognitivi:

> la mente umana, capace di comprendere il linguaggio e di dare significato alle cose, è diversa dagli altri sistemi cognitivi noti, siano essi naturali o artificiali. La nostra mente ha caratteristiche che non siamo (ancora) in grado di riprodurre e che non incontriamo altrove nel mondo naturale. Nondimeno un giorno potremo colmare il divario. La differenza riflette i limiti della nostra conoscenza, non rivela una separazione effettiva tra sistemi cognitivi di diverso tipo (Dumouchel & Damiano 2019, 75)[9].

Assumendo il valore della mente incorporata, situata, che tiene necessariamente conto del contesto ambientale, il pensiero si configura ulteriormente come dimensione della complessità che non può basarsi solo sull'accumulo statistico-quantitativo.

Il valore aggiunto dell'eterogeneità del cognitivo è dato, dunque, dall'idea che «vi siano diversi tipi di sistemi cognitivi, così profondamente diversi tra loro che la transizione dall'uno all'altro non risulta caratterizzabile nei termini dell'aggiunta di algoritmi più potenti o di capacità computazionali nuove e più veloci» (Dumouchel & Damiano 2019, 24). Ritornare sull'eterogeneità del cognitivo, forse, permette di abbandonare l'idea antropocentrica dell'intelligenza in riferimento alla complessità umana, pur non negando l'estendibilità del criterio del razionale basata sulla relazionalità soggetto-ambiente e che coinvolge insieme la sfera teoretica e la sfera pratica nella produzione di nuove conoscenze o nuove forme dell'azione.

[8] La cognizione incorporata qui si riferisce a un modello di intelligenza propriamente umana, in cui i processi cognitivi si combinano strettamente con la collocazione del soggetto in un determinato contesto. Questo significa assumere il corpo nella funzione che esercita dal punto di vista conoscitivo, in relazione alle risposte emotive che immediatamente offre rispetto all'ambiente esterno. Su questo tema soprattutto le neuroscienze hanno offerto un contributo fondamentale (Damasio 1995). L'intelligenza umana, qui riferibile al livello della complessità assoluta per come sottolineato nell'articolo, si avvale di un insieme di elementi per rendere i processi cognitivi e pratici completi, quindi consapevoli. Questo intreccio tra emozioni, corporeità, coscienza dei processi interni ed esterni rendono il valore dell'intelligenza in senso pieno, in quanto legata alla complessità assoluta. Per questo, qui si sostiene l'idea dell'IA come intelligenza altra o, meglio, come stile cognitivo differenziato rispetto a quello dell'essere umano.

[9] La differenziazione cognitiva non implica una gerarchia che si regga sul valore aggiunto della razionalità, il che significherebbe un'inversione di marcia verso la tendenza antropocentrica. A questo proposito si rinvia a Nee 2005.

In qualche modo, l'accesso a questo livello di complessità *totale* chiama in causa il principio del *redress by design* e la *Strong-AI*, in quanto la macchina è capace di apprendimento profondo (*deep learning*), potendo riconfigurare l'azione in vista degli obiettivi da raggiungere (Pacchioni 2019). Quello che si trova a essere amplificato è proprio l'accuratezza e la complessità del *feedback* fornito da una macchina potenzialmente situata in un contesto.

La complessità totale, dunque, seppure prenda in considerazione la forma combinata teoretico-pratica che caratterizza l'umano, rimane un livello in cui la mente incorporata non interviene fino a che non si dia una percezione totale dell'ambiente. Proprio qui intervengono una serie di linee di ricerca oggi essenziali, quali le neuroscienze e gli studi di Antonio Damasio sul marcatore somatico [10] e la cibernetica (Marchesini 2002; Manzocco 2014). Inserire nel quadro complessivo dell'individualizzazione dimensioni come quella corporea (anche nella forma delle reti neurali artificiali) e quella emotiva significa raggiungere il livello della complessità *assoluta* – questa, per sua stessa definizione, introduce un numero di variabili eccessivo. Un tale approccio permette un ulteriore spostamento dall'intelligenza e dalla tendenza antropomorfa verso la teoria della diversificazione cognitiva. Quello che si definisce come "intelligenza", nel senso intimamente umano, è determinato da due ragioni fra loro interconnesse – per come evidenziato da Sadin:

> La prima è che queste architetture computazionali sono prive di corpo; esse non sono altro che delle macchine calcolatrici la cui funzione si limita alla semplice elaborazione di flussi informativi astratti. E nel caso in cui esse si trovino collegate a dei sensori (cibernetica), non fanno altro che ridurre certi elementi del reale a dei codici binari, trovandosi escluse da un'infinità di dimensioni che invece la nostra sensibilità coglie e che sfuggono ai principi di una modellizzazione matematica […]. La seconda ragione è che non esiste intelligenza che viva isolata, chiusa nelle proprie logiche; e penso al principio di progressione che consiste nell'esercitarsi da solo "contro sé stessi" come in una bolla, conformemente alla logica detta per "rafforzamento" […]. Perché l'intelligenza è indissociabile da rapporti aperti e indeterminati con gli esseri e le cose, da un contesto epigenetico, da un ambiente, cioè, composito nel quale evolvere e distinguersi (Sadin 2019, 23).

I primi passi della cibernetica erano infatti orientati a fornire questo alla macchina. Gli enti non possono essere concepiti come sistemi isolati se vogliono crescere, evolvere; per questo l'importanza nevralgica degli organi effettori e degli organi di senso. La sempre maggiore configurazione antropomorfa, la creazione di corpi per le intelligenze artificiali mira ad avvicinare la macchina a una forma di complessità assoluta – che come si è visto costituisce un livello di analisi che include il valore delle emozioni e del corpo, in riferimento all'intreccio conoscitivo-pratico dell'essere umano. Da qui, l'idea dell'*organismic embodiement*: «ricreare nei sistemi robotici la complessa interrelazione tra cognizione e dinamiche fisiologiche di regolazione emozionale tipica di agenti umani e animali» (Dumouchel & Damiano 2019, 119).

Un ulteriore livello di articolazione è dunque basato sul sistema stimolo-risposta, quindi su una forma di agentività riflessa (Russell 2020) che risponde allo stimolo ma ancora non comprende il valore dell'obiettivo. Si prenda ad esempio un veicolo

[10] Scrive Antonio Damasio: «quando viene alla mente, sia pure a lampi, l'esito negativo connesso con una determinata opzione di risposta, si avverte una sensazione spiacevole alla bocca dello stomaco. Dato che ciò riguarda il corpo, ho definito il fenomeno con il termine tecnico di stato *somatico*; […]. Esso forza l'attenzione sull'esito negativo al quale può condurre una data azione, e agisce come un segnale automatico di allarme […]; vi permette di *scegliere entro un numero minore di alternative*» (Damasio 1995, 245).

autonomo: sa che deve fermarsi se passa un pedone, ma non comprende qualitativamente il valore di una vita umana e dunque il perché sia necessario fermarsi. Nonostante l'aumento della complessità del sistema, non si arriva a un livello di interfaccia empatica con l'essere umano, essendo una semplice percezione di pericolo programmata (dato-quantità) a fermare l'attività della macchina.

Qui si presenta uno dei problemi primari, per cui la riproduzione della sfera emotiva passa non solo da meccanismi interni alla soggettività, ma anche legati all'interazione fra individui. L'affettività ha un'origine sociale, inter-individuale e si sviluppa in base al rapporto individuo-ambiente: «le nostre emozioni e le nostre reazioni empatiche non sono imprese solitarie – private. Sono opere comuni, a cui molti partecipano. Le emozioni e l'affetto rinviano al corpo ma non al corpo dell'organismo individuale. La loro incorporazione ha come sede quanto possiamo definire un "corpo sociale"» (Dumouchel & Damiano 2019, 137).

Il riconoscimento del valore delle emozioni è anche un'indagine su uno stile cognitivo incarnato che, intercettato dagli studi sull'IA, spingono verso l'imitazione di una complessità assoluta in cui le «Emozioni non [sono assunte] come ostacoli alla razionalità umana, ma come processi essenziali all'espressione di strategie cognitive adattive» (Dumouchel & Damiano 2019, 25). Questa riflessione rimane in linea con la ricerca neuroscientifica di Damasio in virtù dell'intreccio fra emozioni primarie come risposta all'ambiente e quindi strategia istintiva di adattamento.

Mentre la macchina tenta di arrivare a un'imitazione totale della complessità dell'essere umano, questa stessa complessità è un sistema in espansione con un numero di variabili crescente. Metaforicamente, si è di fronte a una corsa tra Achille e la tartaruga che mostra come la macchina difficilmente riesca ad arrivare all'obiettivo di un'imitazione assoluta dell'essere umano: «non siamo cyborg intellettuali e metafisici. Siamo attori epistemici all'interno di una società cognitiva complessa, costituita da sistemi cognitivi di tipo diverso – alcuni naturali, altri artificiali» (Dumouchel & Damiano 2019.88). Quindi, l'IA entra nel sistema delle relazioni inter-soggettive, ma ponendosi nei termini di "intelligenza" diversificata, non necessariamente come impronta di quella umana. Se quest'ultima può rappresentare un modello di partenza più articolato, la macchina segue una propria strada nel percorso della propria evoluzione tecnica; essa rappresenta, dunque, un altro agente che non è chiamato a fare dipendere il sistema-essere umano dal sistema-IA. In tal caso, il primo verrebbe appiattito sul secondo (come si è visto nel precedente paragrafo), mentre il secondo rimarrebbe sempre intrappolato nelle maglie di un'imitazione insensata, laddove invece la diversità del cognitivo gli permette di svolgere una funzione diversa e, forse, complementare rispetto al sistema delle relazioni sociali.

4. *Dissonanze perturbanti e incontri faccia-a-faccia*

Ritornando brevemente sul livello di complessità totale, si può dire che una macchina può dunque percepire in forme diverse e non necessariamente sensoriali, come nel caso di agenti analitico-esecutivi complessi (i veicoli autonomi), ma l'epistemologia della percezione determina un'immediata diversità fra lo stile percettivo della macchina e quello delle specie naturali. L'IA simula la percezione in un processo di raccolta dati, interfacciandosi con un sistema ben più articolato che non contempla solo un pensiero di tipo calcolante: quello umano.

La macchina è dunque chiamata a tenere conto della complessità totale in riferimento ad alcuni criteri perché possa eseguire l'azione: (i) l'osservabilità parziale o totale

dell'ambiente; (ii) se l'ambiente e la catena delle azioni da eseguire sono discrete o continue; (iii) se l'ambiente comprende o meno altri agenti; (iv) se i risultati delle azioni sono prevedibili o imprevedibili, in base alle leggi della fisica che governano l'ambiente e se quindi queste leggi sono note o meno all'agente; (v) se l'ambiente è in continuo cambiamento, il che richiede necessità di decisioni veloci da prendere; (vi) la finestra temporale delle conseguenze delle azioni che misura la qualità delle decisioni, in questo caso maggiore è l'estensione in termini temporali e anche spaziali, maggiore sarà il numero di variabili in gioco (Russell 2020, 44). Quanto appena elencato serve a dare un'idea di cosa la macchina debba tenere in conto e i relativi criteri di programmazione. Costruire delle IA simili è l'attuale risposta dell'avanzamento della tecnica di fronte all'aumento del fattore complessità che

> ha molte teste, come l'Idra. Una di queste è la complessità spaziale, cioè il numero di bit di informazione occupati dall'algoritmo nella memoria del computer […]. Poi c'è la sorella malvagia, la complessità temporale, cioè il tempo necessario per eseguire l'algoritmo […]. Il volto più spaventoso del mostro della complessità, tuttavia, è la complessità umana. Quando gli algoritmi si fanno troppo complicati perché il nostro povero cervello di esseri umani possa capirli, quando le interazioni tra le varie componenti dell'algoritmo diventano troppe e troppo involute, al loro interno si insinuano errori che non riusciamo a trovare e a correggere, e l'algoritmo non fa ciò che vorremmo. E anche se, in un modo o nell'altro, riusciamo a farlo funzionare, finisce per essere troppo complicato per chi deve usarlo, non si integra con gli altri algoritmi e rischia di essere fonte di guai in futuro (Domingos 2020, 27).

Di fronte a questa complessità assoluta è la macchina che teme l'essere umano e la sua inimitabilità, non tanto la paura contraria che circola sulle apocalissi dell'umanità causate dalle super-intelligenze. Per arrivare a quella complessità è necessario che la macchina apprenda – e che lo faccia in autonomia: «un computer che non sa imparare non riuscirà a competere a lungo con un essere umano» (Domingos 2020, 31). L'apprendimento amplifica la complessità del sistema. Ma un maggiore numero di dati, se slegati da un'incapacità pratica della loro gestione, non determina un livello di intelligenza più alto; anzi, una più alta possibilità di stasi o di errore.

Ora, mentre nei primi due livelli si sono considerate la complessità relativa e la complessità totale, il terzo livello apre alla dimensione della complessità assoluta, già vista nel precedente paragrafo, in riferimento a un sistema robotico di interfaccia uomo-IA. Si parla, dunque, di veri e proprio agenti interattivi, di doppi (Rank 1971), di sostituti (Dumouchel & Damiano 2019). Su questo livello, si consideri sempre la parzialità dell'approccio. Al momento esistono determinate intelligenze artificiali che, poggiando sempre su costruzioni algoritmiche, vengono create funzionalmente per essere introdotte in ambito sociale (per esempio, *robot-sitter*). In questo caso, però, seppure il livello di complessità appaia come assoluto, lo è solo, e letteralmente, a un livello epidermico: sia perché non esiste un algoritmo definitivo (Domingos 2020) in grado di rispondere totalmente all'ambiente nelle modalità cognitivo-pratiche con cui lo farebbe un essere umano, sia perché ogni agente interattivo presenta un livello di programmazione circoscritto allo scopo per cui è programmato – da cui l'idea di intelligenza tronca, non totale ma sempre relativa all'obiettivo che si pone. Se l'IA esce fuori da quello stimolo ambientale, allora decade la sua funzione, in quanto non si mostra capace di rispondere adeguatamente allo stimolo posto. Il robot *Geminoid*, sosia di Hiroshi Ishiguro, è una presenza sociale che serve al suo originale umano per poter tenere delle conferenze a distanza. Nonostante questo abbia un certo livello di interazione, è comunque un robot-burattino che non ha una propria autonomia, è una maschera attraverso cui Ishiguro

interagisce. L'unica evidente dimensione che lo contraddistingue da altre IA più deboli è quella corporea. La presenza del corpo è qui determinante e si pone momentaneamente a compimento di quel percorso di imitazione della complessità, seppure in maniera incompleta. Questo vuol dire che *Geminoid* risponde ai comandi del padrone, è roboticamente evoluto in quanto presenta un corpo, ma non è autonomo: né cognitivamente né praticamente.

Il corpo, qui, è assunto nella sua funzione comunicativa di semplice presenza, non perché possieda una propria autonomia complessiva. Tuttavia, questo livello di costruzione dell'IA tenta di avvicinarsi all'essere umano, fornendo l'intelligenza artificiale di una struttura corporea; questo perché, come si è già visto «gli approcci emergenti della mente incorporata (*embodied mind*) riconoscono un ruolo fondamentale, nella conoscenza umana, al corpo e alle emozioni» (Dumouchel & Damiano 2019, 57). Rimane comunque un problema di ordine ontologico: la complessità è ancora ridotta a un *aut-aut* (o il corpo autonomo o la mente autonoma) che, momentaneamente, non è colmabile. Massima prestazione cognitiva e corporeità della macchina ancora non si congiungono; le emozioni sono solo riprodotte con una mimica esterna che pare produrre un effetto di straniamento nell'essere umano che vi si interfaccia, da cui la *uncanny valley* di Mori.

Il tema del perturbante – qui rievocato partendo dal testo di Otto Rank *Der Doppelgänger* del 1914 – ritorna sul rapporto alterità-paura. Proprio dall'interpretazione di Rank, infatti, Freud recupera il valore di una regressione verso stati psichici animistico-primitivi. Tali stati si riferiscono a una condizione di confusione, dovuta a un'identità ancora sfumata fra l'io e ciò che gli è estraneo/esterno. Qui il valore del perturbante è riconducibile a ciò che genera paura, terrore e anche disgusto; questo implica un distanziamento fra l'io e ciò che questi percepisce come perturbante.

Contestualizzando il tema a quanto qui in oggetto: maggiore è la somiglianza dell'IA all'umano, maggiore sarà la sua inefficienza. Rispetto al semplice assistente vocale, il robot diventa sempre più perturbante e tende a causare, stando a Mori, un certo livello di paura derivante da un eccessivo senso di familiarità. Fino a che si tratta di un robot semovente dalle sembianze animali (come la foca Paro), la percezione rimane sul livello del non-familiare. L'implicazione è che l'IA non-somigliante all'essere umano svolge la propria funzione di tacito e servile supporto. Tuttavia, con l'aumento della complessità dell'interfaccia (simulazioni corporee, vocali, espressive, di movimento) c'è un aumento proporzionale del senso del perturbante che produce straniamento, rendendo dunque la macchina socialmente e interattivamente inefficiente. Anzi, questa rievocherebbe una paura tale da produrre allontanamento essendo percepita come un'individualità piena o, addirittura, come minaccia perché potenzialmente sostitutiva del suo creatore.

Il salto della robotica sociale è, in qualche modo, quello di introdurre dei sostituti ma riconoscendone uno stile cognitivo differenziato, più o meno imitativo, e che può tornare utile – anche solo per simulazione esterna – a comprendere la complessità cognitivo-affettiva-relazionale degli esseri umani. La conoscenza permette attività di *coding*, ma il pensiero richiede un'intersezione pluri-dimensionale. La dimensione corporeo-percettiva svolge un ruolo fondamentale per la rappresentazione e l'orientamento nel mondo; questo rende il livello di compatibilità con l'umano significativamente alto, ma richiede un processo di elaborazione di conoscenze, azioni ed emozioni sia interno che esterno per la macchina.

Questo permette di uscire fuori dal paradigma teorico dell'antropocentrismo cognitivo, guardando all'IA come stile cognitivo *altro* da quello umano, distinguendolo per tempi e scopi propri di funzionamento. Il percorso evolutivo che segue l'IA, dunque,

è quello dal robot-automa a vero e proprio attore sociale, chiamando in causa l'empatia e arrivando a toccare il limite del perturbante.

Il robot non è immediatamente morale se rimane sul livello imitativo della complessità totale, esso è conforme alla legge imposta, ai principi per cui è programmato, dunque agirà legalmente: questo è un confinamento della morale sulla semplice conoscenza dei principi di programmazione. Non c'è etica, ma solo aderenza e imitazione senza simulazione. L'IA rimane su un piano eteronomo, di dipendenza dall'algoritmo di programmazione – ossia ancora su un livello di conoscenza aprioristica ma non autonoma o flessibile. Al contrario, l'azione morale è una vocazione autonoma, non imposta, che in questo caso non si rende possibile, se non facendo del robot un'individualità piena.

Se a questa tipologia di agenti si aggiunge un corpo e la possibilità di un'interazione, si passa allora ai sostituti. Alcune delle loro caratteristiche – che li distinguono dagli altri oggetti tecnici – sono, appunto, l'autorità, la presenza sociale e fisica e la capacità indefinita di coordinarsi e ri-coordinarsi tra loro e con i loro partner umani (Dumouchel & Damiano 2019). Di fronte a questa presenza totale, la familiarità si approssima verso la zona del perturbante – che è invece pressoché assente o minima nel caso di agenti analitici o solo esecutivi, in quanto il livello di interfaccia con l'essere umano è abbastanza debole; la stessa curva di Mori ha infatti influenzato la produzione della robotica sociale, mostrando il limite massimo di utilità della macchina:

> l'obiettivo è determinare il punto esatto in cui la sensazione di familiarità si trasforma nell'inquietante percezione di una presenza pericolosa. Una delle ipotesi esplorate in questo contesto è che il disagio sia prodotto non dalle differenze residuali, ma dall'eccessiva similarità […] l'aspetto inquietante risiederebbe nella perdita della differenza: l'impossibilità di continuare a pensare che i robot siano effettivamente diversi da noi – non siano (come) noi […]. Ci inquieta l'idea che gli agenti artificiali iniziano ad agire come esseri umano, con tutto quello che questo comporta quanto a incertezza, imprevedibilità e pericolosità. Se ciò accadesse i robot […] diventerebbero per noi sconosciuti quanto i nostri simili (Dumouchel & Damiano 2019, 33-34).

L'interazione, in sostanza, è determinata da un'interfaccia che inquieta l'individuo/interlocutore del robot; l'altro offre il proprio sguardo, per cui anche la semplice presenza determina un senso di spaesamento. Su questo tema torna anche Sadin, in un'ottica più catastrofista. L'autore, infatti, descrive «una voce che si rivolge a noi, […] dotata di un livello di conoscenza e competenza senza eguali e che ci dispensa continuamente buoni consigli, si vedrà investita di un'autorità e di un'aura tali per cui sarà sempre più difficile, visti i continui perfezionamenti, non percepire questi dialoghi come "naturali" e non prender per "oro colato" qualsiasi sua parola» (Sadin 2019, 53).

Rispetto alla semplice presenza dell'automa, anche se non si è veramente osservati con l'intenzione di un'altra individualità, l'inquietudine non si modifica, ma rimane tale perché «[io] sono bersaglio dello sguardo dell'altro» (Dumouchel & Damiano 2019, 139). A questo, infatti, si abbinano gli studi di bio-mimetica che servono a ottenere il massimo della riproducibilità dei comportamenti umani – per lo meno quelli esteriorizzabili. Allo scopo di specificare meglio: «per essere riconosciuto come un robot sociale, un agente robotico deve essere in grado di provocare nei propri interlocutori umani quel tipo di percezione dell'alterità caratteristica delle relazioni sociali di base – le interazioni dirette, "faccia a faccia", dalle quali, in ultima analisi, derivano tutte le altre relazioni sociali» (Dumouchel & Damiano 2019, 104).

Questa tendenza, in realtà, rappresenta un'«inquietante passione: generare doppi artificiali di sé stesso» (Sadin 2019, 37). Il grafico di Mori rievoca nella sua ultima

sezione un *Cyborg-Buddha*, piena e consapevole espressione che i robot, nel loro impiego come agenti interattivi e sociali, funzionano nel momento in cui non presentano una somiglianza eccessiva rispetto all'essere umano.

Ora, anche nel caso del robot che è capace di mimare somaticamente le emozioni umane, si è di fronte a un'imitazione della sfera emotiva che nasce da una conoscenza della dimensione affettiva esterna, ma non da un'immediata reazione primaria come può esserlo l'emozione reale (per esempio, paura, gioia, rabbia). Si rimane su un livello di programmazione perturbante, ma pur sempre fasulla e simulativa; per cui i robot agiscono o provano emozioni "come se". Quello che le emozioni dicono negli esseri umani, in combinazione con la dimensione cognitiva, è il modo in cui permettono di costruire una narrazione. La storia personale (risultante ultima della complessità assoluta) ha un carattere qualitativo-simbolico.

Nel contesto sociale, il robot rimane presenza perturbante, imitazione ed esteriorizzazione dell'emotività umana; è una complessità apparente. Il robot – anche nel caso di esatte simulazioni interne ed esterne – rimane lontano dall'essere umano, in quanto incapace di accedere alla sfera del significato; esso è una metà dell'umano, un'ombra (Rank 1971).

Se, dunque, gli stili cognitivi sono plurali e differenziati, lo stesso dicasi di due termini come "complessità" e "intelligenza". Scrive Sadin:

> niente a che vedere con ciò che costituisce noi esseri umani, sempre proiettati verso una gran quantità di aspirazioni diverse. C'è un'irriducibilità della vita, così come c'è un'irriducibilità dell'intelligenza umana, entrambe refrattarie a qualsiasi definizione circoscritta e a qualsiasi categorizzazione rigida. Come del resto c'è un'irriducibilità dei nostri sentimenti che resiste a ogni iniziativa di schematizzazione integrale. L'intelligenza artificiale non è in alcun modo una replica della nostra intelligenza nemmeno parziale: è l'abuso del linguaggio che ci fa credere che essa potrebbe essere in grado di sostituirsi con naturalezza alla nostra intelligenza al fine di garantire una migliore gestione delle cose che ci riguardano. In realtà si tratta più esattamente di una *metodologia della razionalità*, fondata su schemi restrittivi e volta a soddisfare qualsiasi tipo di interesse (Sadin 2019, 23).

Il crollo del modello dell'intelligenza artificiale è dunque dato dalla sua sostanziale incompiutezza, rispetto a ciò che l'essere umano rappresenta come specie a sé stante, né migliore degli animali, né peggiore dell'IA più evoluta. L'eterogeneità del cognitivo è una linea teorica che riesce a giustificare e a garantire il giusto spazio alla diversità dei sistemi organismo-ambiente. La ricostruzione qui proposta rileva (a) i livelli di complessità differenti (relativo, totale, assoluto) nell'essere umano e (b) la natura olistica di ogni sistema cognitivo data l' interazione con l'ambiente esterno, che produce risposte diverse e quindi modifiche nel sistema stesso. Lo stesso vale per ogni singolo livello di complessità, compreso quello relativo, in cui l'intreccio fra abilità mentali (analitiche o sintetiche, simboliche o sub-simboliche) determina sistemi differenti, fra loro comunicanti e irriducibili.

La riproducibilità o la replicabilità di un sistema determina la creazione di un surrogato che è "altro" rispetto a quello di partenza. Nei casi di IA più evoluti, quando un tale livello di riproduzione determina interfacce complesse, questo non significa fare della macchina un supporto ma sfiorare i limiti del perturbante, delegittimando il valore funzionale, più o meno complesso, per cui è programmata.

Data di pubblicazione: 15.09.2021

Riferimenti bibliografici

Allport, G., et al. (a cura di) (1970). *Psicologia esistenziale*. Roma: Astrolabio.

Anders, G. (2007). *L'uomo è antiquato*, 2 voll. Torino: Bollati Boringhieri.

Arendt, H. (2005). *Teoria del giudizio politico*. Genova: Il Nuovo Melangolo.

Arendt, H. (2009). *La vita della mente*. Bologna: Il mulino.

Arendt, H. (2010). *Responsabilità e giudizio.* Torino: Einaudi.

Arendt, H. (2014a). *La banalità del male. Eichmann a Gerusalemme*. Milano: Feltrinelli.

Arendt, H. (2014b). *Vita Activa. La condizione umana*. Milano: Bompiani.

Aristotele (2001). *L'anima*. Milano: Bompiani.

Aristotele (2013). *Etica nicomachea*. Roma-Bari: Laterza.

Benveniste, E. (2001). *Il vocabolario delle istituzioni indoeuropee*. *Economia, parentela, società*, vol. 1. Torino: Einaudi.

Bloch, E. (2019). *Il principio speranza*, 3 voll. Milano-Udine: Mimesis.

Borghini, A., & Casetta, E. (2013). *Filosofia della biologia*. Roma: Carocci.

Bruner, J. (1992). *La ricerca del significato*. Torino: Bollati Boringhieri.

Capra, F., & Luisi, P. L. (2020). *Vita e natura. Una visione sistemica*. Milano: Aboca Edizioni.

Carotenuto, A. (2002). *Freud. Il perturbante*. Milano: Bompiani.

Ceruti, M. (2007). *La danza che crea*. Milano: Feltrinelli.

Ciafardone, R. (2007). *La critica della ragion pura di Kant. Introduzione alla lettura*. Roma: Carocci.

Damasio, A. R. (1995). *L'errore di Cartesio. Emozione, ragione e cervello umano*. Milano: Adelphi.

Damiano, L. (2009). *Unità in dialogo*. Milano: Bruno Mondadori.

Dicker, G. (2004). *Kant's Theory of Knowledge. An Analytical Introduction.* Oxford: Oxford University Press.

Dilthey, W. (2007). *Introduzione alle scienze dello spirito*. Milano: Bompiani.

Domingos, P. (2020). *L'algoritmo definitivo. La macchina che impara da sola e il futuro del nostro mondo*. Torino: Bollati Boringhieri.

Forti, S. (2006). *Hannah Arendt tra filosofia e politica*. Milano: Mondadori.

Freud, S. (1989). *Il perturbante*. In S. Freud, *Opere*, vol. 9. Torino: Bollati Boringhieri.

Fuchs, T. (2021). *Ecologia del cervello. Fenomenologia e biologia della mente incarnata*. Roma: Astrolabio.

Kant, I. (2007). *Critica della ragion pura*. Roma-Bari: Laterza.

Lindsay, P. H., & Norman, D. A. (1987). *L'uomo elaboratore di informazioni.* Firenze: Giunti.

Manzocco, R. (2014). *Esseri umani 2.0. Transumanesimo. Il pensiero dopo l'uomo*. Milano: Springer.

Marchesini, R. (2002). *Post-human. Verso nuovi modelli di esistenza*. Torino: Bollati Boringhieri.

Marcucci, S. (1997). *Guida alla lettura della "Critica della ragion pura" di Kant*. Roma-Bari: Laterza.

Maturana, H. R., & Valera, F. J. (1999). *L'albero della conoscenza.* Milano: Garzanti.

Mori, M. (1970). The Uncanny Valley. *Energy*, *7*(4), 33-35.

Nee, S. (2005). The Great Chain of Being. *Nature*, 435, 429-435.

Nida-Rümelin, J., & Weidenfeld, N. (2019). *Umanesimo digitale. Un'etica per l'epoca delle intelligenze artificiali*. Milano: FrancoAngeli.

Pacchioni, G. (2019) *L'ultimo sapiens. Viaggio al termine della nostra specie*. Bologna: Il Mulino.

Rank, O. (1971). *The Double. A Psychoanalytic Study*. Chapel Hill: University of North Carolina Press.

Rossi, F. (2019). *Il confine futuro*. *Possiamo fidarci dell'intelligenza artificiale?*. Milano: Feltrinelli.

Data di pubblicazione: 15.09.2021

Russell, S. (2020). *Human Compatible*. *Artificial Intelligence and the Problem of Control*. London: Penguin.

Russell, S., & Norvig, P. (2010) *Intelligenza artificiale. Un approccio moderno*, vol. 1. Milano: Pearson.

Sacchetto, M. (1989). *La critica della ragion pura di Kant e il problema della fondazione della conoscenza scientifica nel pensiero contemporaneo*. Milano: Paravia.

Sadin, É. (2019). *Critica della ragione artificiale. Una difesa dell'umanità*. Milano: LUISS.

Tamburrini, G. (2020). *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*. Roma: Carocci.

Vaccarezza, M. S. (2012). *Razionalità pratica e attenzione alla realtà. Prospettive contemporanee*. Roma: Orthotes.

Data di pubblicazione: 15.09.2021