Eugenio Mazzarella

Le illusioni della continuità uomo-macchina (e dell'algoretica). La fallacia comportamentista[1]

Abstract: This article critically examines the assumption of ontological continuity between human intelligence and artificial intelligence underlying theories of "strong AI" and contemporary projects of algorethics. Through a critical rereading of Alan Turing's argument in *Computing Machinery and Intelligence*, the paper argues that the behavioral analogy between human and machine conduct constitutes a philosophical fallacy that obscures the distinctive features of natural intelligence, namely consciousness, intentionality, and its psycho-biological and social embodiment. While machines can simulate operational patterns of instrumental rationality, they lack genuine world-competence and moral agency. "Strong AI" is therefore interpreted not as the emergence of a *machina sapiens*, but as the increasingly powerful use of weak AI within processes of social engineering. Consequently, the ethical responsibility for algorithms cannot be delegated to machines: the governance and normative accountability of AI remain entirely human and public.

Alla base della sconcertante possibilità, sempre più evidente oggi, dell'Intelligenza Artificiale: a) di *deskilling* logico – linguistico e di pensiero – dell'espressività individuale dell'animale sociale uomo, della sua perdita di *competenza libera di sé* come di capacità di negoziazione con sé stesso e con i suoi contesti di dipendenza, che *cambiano e lo cambiano mentre li cambia* (contesti di dipendenza che ne sono per altro i *limiti possibilizzanti*, il suo nesso ecologico di *emergenza* affidato alle causalità neuro-fisiologiche, sociali, ambientali: le sue reti "ecologiche" ben prima di quelle digitali); b) di "fine della storia" nel senso radicale del venir meno nell'evoluzione culturale umana di tratti epigenetici sottratti alla non programmabilità della loro emergenza, affidata in linea di principio – in una biopolitica totale dell'umano sempre più capace di *artificiarsi* – ad una selezione calcolata, che non seleziona più caratteri naturali, ma caratteri già artificiati, dove la "natura" dell'*innovazione* "umana", dell'*innovarsi* dell'uomo, può essere stabilita a priori *come deve essere* e *come deve restare* da un demiurgo che presume di non sbagliare un colpo; alla base di questa possibilità – al di là di quanta strada possa effettivamente fare nell'ingegneria sociale in divenire – c'è la pericolosa illusione della *continuità uomo-macchina*. E cioè, insieme, che tratti salienti della razionalità strumentale dell'agire umano possano potenziarsi, *aumentarsi* nella macchina e persino *evol-*

1 Estratto dal volume di prossima uscita E. Mazzarella, *Critica della ragione digitale*, Castelvecchi 2026.

Mechane, n. 10, 2025 • Mimesis Edizioni, Milano-Udine Web: mimesisjournals.com/ojs/index.php/mechane • ISBN: 9791222323237 • ISSN: 2784-9961 • DOI: 10.7413/2784mchn0020

versi autonomamente una volta che vi siano stati implementati. Illusione che è alla base del programma di ricerca dell'IA forte, *generativa*, e che si regge sull'ipotesi della continuità uomo-macchina. Che non ci sia, cioè, discontinuità ontologica tra questi due assetti di "cose". E che le competenze dell'intelligenza naturale siano in linea di principio acquisibili all'intelligenza artificiale, alla *machina sapiens*, e quando acquisite molto più potenti, per una soverchiante capacità di calcolo di quelle dell'*homo sapiens* modellizzato come "macchina".

È la vecchia illusione dell'*homunculus* sintetizzato negli alambicchi, mettendoci dentro tutta *la competenza di mondo* di cui disponiamo, migliore persino di quella fin qui sintetizzata in noi dal "demiurgo" – dio o natura che sia –, che ha consegnato alla nostra insoddisfazione vicoli ciechi evolutivi o problemi insolubili di teodicea. In verità una prova quest'illusione, se la prendessimo sul serio, e non manca chi la prende sul serio, dello *scemunculus* che siamo rimasti, e che in peggio riprodurremmo in questo *homunculus programmato*. Un'illusione che non varrebbe nemmeno la pena trattare, per la sua inconsistenza filosofica, al di là del credito che pure riceve, se non fosse il volano teorico e ideologico non della realizzazione dell'Intelligenza Artificiale forte, generativa, decidente, autonoma, senza i *bias* dell'intelligenza naturale umana, ma in effetti nient'altro che il volano di un uso forte, fortissimo *a fini non umani da parte di uman*i di uno strumento umano, l'intelligenza artificiale. Una possibilità, non una prestazione in sé del calcolo, che poi era l'antivedente, già ricordato, monito di Norbert Wiener, di quello che poteva diventare la cibernetica che contribuiva a fondare. Quel *caveat* oggi è evidente cosa possa proporci: un individuo depotenziato da *persona* con i suoi diritti *a nodo di rete profilato* nelle mani di chi è alla sala comandi della macchina sociale[2].

La tesi che vogliano proporre è che non esiste né concettualmente né ontologicamente un'Intelligenza Artificiale *forte* nel senso di una sua *autoconsistenza meta-macchinica* (*simil umana*), ma solo la concretissima possibilità, già largamente dispiegata, di un uso fortissimo dell'IA "debole", e cioè della già potentissima capacità ingegneristico-strumentale dell'intelligenza artificiale in generale, i cui conseguibili automatismi nella versione "forte" ne perfezionano i fini ingegneristici per cui nasce nell'ambito di applicazione dell'ingegneria sociale. L'intelligenza artificiale "forte", al di là del dato tipologico delle intelligenze artificiali[3], non è nient'altro che *il suo uso forte* nel bene e nel male; che dal lato dei rischi è un uso

2 A. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, pp. 167-204 in *Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, a cura di A. D'Aiola, Franco Angeli, Milano 2020, pp. 167-204.

3 Debole (*weak*); forte (*strong*); articolata (*narrow*); generalizzata (*general*). Su questa dimensione multilivello della sua applicazione o del suo impiego che vede la "neutralità" in sé dell'IA come tecnologia messa alla prova della sua operatività nei contesti e nelle finalità d'uso (il *carattere socio-tecnologico* dell'IA che ne decide la valutazione di impatto), cfr. M.E. Mazzarella, *L'Unione Europea e la corsa alla* governance *dell'Intelligenza artificiale: verso una disciplina generale*, in https://dirittodiinternet.it/lunione-europea-e-la-corsa-alla-governance-dellintelligenza-artificiale-verso-una-disciplina-generale/.

che esondi dall'istanza antropocentrica che dovrebbe essere propria ad una tecnologia così socialmente rilevante – non essendoci nulla nell'Intelligenza Artificiale in generale di *antropocentrabile*, ove con questo di intenda l'implementazione o l'insorgenza in essa di una *machina sapiens* come competenza di mondo, come agenzia morale assimilabile (e persino più efficiente) all'intelligenza naturale, all'umana competenza di mondo.

"State tranquilli, l'IA non si ribellerà agli esseri umani", ancorché "le macchine stiano diventando riflessive"[4], dichiara il filosofo della tecnologia Yuk Hui. Ne siamo con lui assolutamente convinti. Un po' meno sulla "riflessività" delle macchine, e sulla possibilità di potervela implementare. Solo, che il punto non è questo, un tema caro alle distopie narrative diversive dal problema vero. Il punto è che l'IA non è progettata per "ribellarsi" agli umani, al netto dei rischi ovvi di *blackout* operativi che mandino in *tilt* i sistemi che innerva, insomma che si "rompa" bloccandosi o andando fuori controllo non rispondendo ai comandi. Rischio che è proprio ad ogni macchina, anche la più elementare, come un martello che ti si spezza in mano e ti fa capire che la sua essenza è la sua funzione, non l'assemblaggio dei suoi pezzi, come ha insegnato Heidegger nelle sue analisi sull'*utilizzabilità* nell'analitica esistenziale di *Sein und Zeit*. Il punto è che l'Intelligenza Artificiale è progettata, o può essere progettata (e largamente lo è) *per non far ribellare* gli esseri umani *ad una società suo tramite sempre più amministrata*, sul piano sociale, economico, politico.

Ciò nonostante, si continua a credere – forse perché viviamo in un mondo dove ci si parla sempre meno "di presenza", "dal vivo", senza ricorrere alla mediazione virtuale (fonica o visiva) di sistemi di messaggistica e di scrittura – di poter un giorno parlare in una "conversazione aperta" con una macchina come in una conversazione tra due "agenti" che abbiano entrambi alle spalle, anche la macchina, una competenza di mondo "naturale" nel suo carattere saliente di essere acquisita in proprio, da sé, *physei* e non *thesei*. Insomma, che con ChatGPT siamo prossimi a che la macchina superi il test di Turing, perché "tutti gli indizi dicono che possiamo aspettarci che nuove abilità emergano spontaneamente seguendo questa strada [*scilicet* dell'addestramento del suo programma IA come autoapprendimento di nuove abilità]"[5]. Anche se, ove questo accadesse, sorgerebbe il quesito, che abbiamo visto non preoccupare Yuk Hui, se "saremo in grado di controllare un'entità più intelligente di noi"[6]. Quesito che più correttamente formulato dovrebbe essere il seguente: se saremo in grado di controllare un'entità o le entità – assolutamente non macchiniche: tecnocrazie, politica, consigli di amministrazione – che controllano l'intelligenza artificiale.

Quando Turing in *Computing Machinery and Intelligence* ["Mind", 1950], articolo che ha aperto la strada alla filosofia degli automi, in modo un po' irridente si

4 "State tranquilli. L'IA non si ribellerà agli esseri umani", intervista di A. Velardi e Yuk Hui, *Il Messaggero*, 5 agosto 2024.

5 N. Cristianini, *Machina Sapiens. L'algoritmo che ci ha rubato il segreto della conoscenza*. il Mulino, Bologna 2024, p. 9.

6 *Ibidem*.

evita il confronto con il linguaggio comune (confronto che si ridurrebbe a "una indagine statistica del tipo delle inchieste Gallup") per rispondere alla questione se le macchine possano pensare evitandosi così l'intralcio ontologico, sotteso al principio *rem tena, verba sequentur* (la massima di Marco Porcio Catone trasmessaci nel IV sec. da Gaio Giulio Vittore come *praeceptum paene divinum*), di dover definire che cos'è una macchina e cosa significhi pensare, sostituendo la domanda "le macchine possono pensare?" con quella per lui strettamente analoga "sono immaginabili calcolatori numerici che si comporterebbero bene nel gioco della imitazione?"[7], fa un "gioco" argomentativo che mette da parte la materia ontologica della domanda solo fittiziamente postasi ["le macchine possono pensare?"], per rispondere positivamente ad un'altra domanda, che esplicita una sua convinzione ed una sua largamente "azzeccata" previsione: "Penso sia probabile che entro la fine del secolo sarà possibile programmare una macchina a rispondere alle domande in modo tale che sarà estremamente difficile indovinare se le risposte siano formulate da un uomo o dalla macchina", perché "possiamo sperare che le macchine saranno alla fine in grado di competere con gli uomini in tutti i campi puramente intellettuali"[8].

Cosa che ovviamente le macchine che hanno battuto i migliori giocatori al mondo di scacchi e di *Go* e ChatGPT e i suoi omologhi largamente hanno fatto, ed è anche vero che come credeva Turing "alla fine del secolo l'uso delle parole e l'opinione corrente si saranno talmente mutate che chiunque potrà parlare di macchine pensanti senza aspettarsi di essere contraddetto"[9]. Ma a parte il fatto che in questa previsione Turing si appella proprio ad una campionatura statistica dell'uso del linguaggio naturale in stile Gallup, ciò non significa affatto "che la domanda iniziale, "possono pensare le macchine?", sia troppo priva di senso per meritare una discussione"[10]. E che per renderla priva di senso basti il ricorso ad un'analogia comportamentista che, benché non possa evitare di ammettere che operare macchinico e pensare *non sono* la stessa cosa, si legittima perché quest'analogia ("imitazione) può essere messa in funzione nelle macchine, e un'imitazione soddisfacente ci esonera dal dover tener conto della differenza ontologica in gioco tra due diversi

7 A. Turing, *Macchine calcolatrici e intelligenza,* tr. it. in (a cura di) V. Somenzi e R. Cordeschi, *La filosofia degli automi. Origini dell'intelligenza artificiale*, Bollati Boringhieri, Torino 1994, p. 157: "Mi propongo di considerare la questione: 'Possono pensare le macchine?' Si dovrebbe cominciare col definire il significato dei termini 'macchina' e 'pensare'. Le definizioni potrebbero essere elaborate in modo da riflettere il più possibile l'uso normale delle parole, ma questo atteggiamento è pericoloso. Se il significato delle parole 'macchina' e 'pensare' deve essere trovato esaminando le parole stesse attraverso il loro uso comune è difficile sfuggire alla conclusione che tale significato e la risposta alla domanda 'Possono pensare le macchine?' vadano ricercati in una indagine statistica del tipo delle inchieste Gallup. Ciò è assurdo. Invece di tentare una definizione di questo tipo sostituirò la domanda con un'altra, che le è strettamente analoga e che è espressa in termini non troppo ambigui. La nuova forma del problema può essere descritta nei termini di un gioco, che chiameremo 'il gioco dell'imitazione'".

8 Ivi, p. 181.

9 Ivi, p. 165.

10 *Ibidem*

"stati di cose": "Non possono forse le macchine comportarsi in qualche maniera che dovrebbe essere descritta come pensiero ma che è molto differente da quanto fa un uomo? Questa obiezione è molto forte, ma come minimo possiamo dire che se, ciononostante, una macchina può essere costruita in modo da giocare il gioco dell'imitazione soddisfacentemente, non abbiamo bisogno di tenerne conto"[11].

Un'argomentazione che è un'evasione anticipata – che sarà propria a gran parte di quella che sarà l'istanza generativa di *agency* autonoma dell'Intelligenza Artificiale nelle aspettative della sua teorizzazione "forte" – da un'ontologia della *coscienza* come caratteristica saliente dell'*intelligenza* come *intus legere* (in sé stessi, negli altri, nelle, ovvero tra, le cose) *affettivo, volitivo, intenzionale* che è lo statuto proprio dell'intelligenza naturale umana. Intelligenza naturale umana ridotta da Turing – la fallacia comportamentistica di tutta la sua argomentazione – a schemi "trascendentali" di comportamento comuni alla macchina e all'uomo, e come tali implementabili nella macchina[12] evitandosi l'obiezione effettiva della *coscienza* come di uno "stato di cose", la *mente*, che certo è un comportamento tracciabile in schemi di *input-output*, che genera comportamenti osservabili "dentro" e "fuori" di sé ("dentro" e "fuori" relativi l'uno all'altro della struttura *osmotica* della mente come *mente estesa*), ma è un comportamento non solo *motivato* (come ogni sistema fisico, biologico, psichico, quando sia sollecitato, riceva un motivo, sia mosso o smosso), ma *consapevole* – cioè un comportamento intenzionale, che *intenziona* qualcosa e che, quanto al comportamento intenzionale umano, *sa che fa* e *perché lo fa*. Turing di questa "obiezione della coscienza", che riprende nella formulazione di Geoffrey Jefferson[13], si sbarazza con lo strampalato argomento che del proprio *intus legere*, del proprio pensare e pensarsi – affettivamente volitivamente e intenzionalmente intonato – potrebbe darci risposta solo la macchina, o meglio solo la macchina potrebbe averne contezza, ove si volesse porre la domanda "le macchine possono pensare?". Ma che essa pensi o no non rileva quanto a ciò che veramente conta: il suo comportamento osservabile e sincronizzabile con l'ambiente e le finalità operative di calcolo. E che d'altro canto anche per un uomo "la sola via per sapere [se] pensa è quella di essere quell'uomo in particolare", il che vuol dire che

11 Ivi, p. 159.

12 "Noi crediamo infatti che non soltanto è vero che essere regolati da leggi di comportamento implica essere una specie di macchina (quantunque non necessariamente una macchina a stati discreti), ma che viceversa essere una macchina di questo tipo implica essere regolati da tali leggi", ivi p. 175.

13 Ivi, pp. 168-169: "Questo argomento fu espresso molto bene nel 1949 dal professor Jefferson: 'Fino a quando una macchina non potrà scrivere un sonetto o comporre un concerto in base a pensieri ed emozioni provate, e non per la giustapposizione casuale di simboli, non potremo essere d'accordo sul fatto che una macchina eguagli il cervello – cioè, che non solo scriva ma sappia di aver scritto. Nessun meccanismo potrebbe sentire (e non semplicemente segnalare artificialmente, ché sarebbe un facile trucco) piacere per i suoi successi, dolore quando una sua valvola fonde, arrossire per l'adulazione, sentirsi depresso per i propri errori, essere attratto dal sesso, arrabbiarsi o abbattersi quando non può ottenere quel che desidera'", G. Jefferson, *The Mind of Mechanical Man* (Lister Oration for 1949), Brit. med. J., 1 (1949) 1105-21.

l'intersoggettività umana dimostrabile (per Turing, l'intersoggettività, nient'altro che "educata convinzione che ognuno pensi", ché nessuno può dimostrarlo all'altro) è, come quella che osserviamo delle macchine, un'intersoggettività puramente operazionale, comportamentale retta da calcoli[14].

Con questo "argomento del solipsismo", Turing ritiene di essersi sbarazzato dell'obiezione della coscienza alla legittimità della sostituzione della domanda "se le macchine possano pensare" con il "gioco dell'imitazione". Perché entrambi questi "stati di cose", macchine e uomo (da Turing già modellizzato, in effetti, come *homme machine*) non sono in grado di *dimostrare*, di *esternare* la loro supposta, o supponibile, coscienza di sé, lo "stato di coscienza", se non tramite i loro comportamenti, che sono tutto ciò che è *dimostrabile*, cioè *osservabile e (com)misurabile* tra questi due "stati di cose". Il che se è certamente vero per la macchina (non si può dimostrare ciò che *non c'è*, non si mostra). Ma non è affatto vero per l'uomo, per quello "stato di cose" fornito di coscienza, che non solo sa di pensare ma sa anche che gli altri pensano "in parallelo" con lui, e come lui. La *propriocezione riflessiva situata* di sé e dei co-specifici (*senzienza* di sé *e* del mondo affettiva volitiva intellettiva) di questo "stato di cose", l'uomo, è evidenza elementare ben prima che la neurofisiologia dei neuroni specchio (tramite il *neuroimaging*) ce la facesse vedere, ci "dimostrasse" quel che ci siamo sempre già mostrati l'un l'altro: come siamo sincronizzati "riflessivamente" tra co-specifici in una *comune senzienza di sé e del mondo* già prima dell'emergenza della coscienza individuale in quanto tale. Che è come dire che la *res cogitans* nasce già sempre come *res extensa*, *emerge* (e si regge su) dalla sua *estensione biologica 'specifica'* in uno scambio imitativo tra co-specifici, in un "esterno" conquistato dalla specie piuttosto che dall'individuo, o meglio che l'individuo può conquistare perché glielo ha già conquistato la specie (Poincaré). Il che vuol dire, per tornare al "gioco dell'imitazione", che la macchina non imita propriamente niente, ma è semplicemente un simulatore di comportamenti indotti, anche quando questa induzione prevede un più o meno efficiente automatismo come apprendimento in situazione (*machine learning*). Ad apprendere tramite imitazione è l'animale, la macchina imita tramite apprendimento *altrui*, cioè tramite programmazione: quella della macchina resta per così dire *scienza infusa*, non sapere appreso per imitazione, cioè per scambio con un mondo ambiente di co-specifici. *Non è la macchina ad imitare noi, siamo noi a farci imitare*

14 Ivi, p. 168: "Questo argomento [l'argomento della coscienza] sembra una negazione della validità del nostro. Secondo la forma più estrema di questa opinione il solo modo per cui si potrebbe essere sicuri che una macchina pensa è quello di essere la macchina e di sentire sé stessi pensare. Uno potrebbe allora naturalmente descrivere queste sensazioni al mondo, ma ovviamente nessuno sarebbe giustificato nel darvi ascolto. Allo stesso modo, secondo questa opinione la sola via per sapere che un uomo pensa è quella di essere quell'uomo in particolare. È questo in effetti il punto di vista solipsistico. Può essere il punto di vista migliore cui attenersi sul piano logico, ma rende difficile la comunicazione delle idee. Probabilmente A crederà 'A pensa, ma B no', mentre B crede 'B pensa, ma A no'. Invece di discutere in continuazione su questo punto, è normale attenersi alla educata convinzione che ognuno pensi".

dalla macchina. Per dalla macchina – da chi ne è al comando – farci prescriverci come imitarla, come ottimizzarci all'ingegneria sociale richiesta dal mondo governato da e tramite l'Intelligenza artificiale.

È un mistero della fede filosofico che un articolo così evasivo sull'ontologia dell'intelligenza naturale umana ridotta ad una classe di comportamenti, neppure tutti, quelli strumentali operativi, che la macchina può imitare, abbia potuto essere così influente – al di là delle possibilità della logica operativa dei sistemi di calcolo che illustra – sulle pretese antropomorfe della "filosofia" dell'intelligenza artificiale, degli "automi". Pretese cui manca ogni base isomorfa con il "sistema operativo" della coscienza al di là dell'isomorfismo programmabile tra la macchina e i comportamenti strumentali operati da calcoli dell'intelligenza naturale umana, comportamenti nella macchina implementabili. Ed amplificabili in misura che è difficile sottostimare, il che però non toglie nulla alla fallacia immaginativa di un'*agency* simil umana acquisibile dalla macchina. Nessun comportamentismo, se vuole avere a che fare con comportamenti *umani*, e non ridurli a sequenze operative macchiniche perché ne vede nel *neuroimaging* gli schemi neurofisiologici che si attivano e, nel loro corso operativo, gli schemi comportamentali, può aggirare l'*incarnazione* psico-somatica e biosociale di questi comportamenti, la loro struttura affettivo-intenzionale. Che i comportamenti di cui si occupa sono espressione di una "forma di vita", quella umana, l'organico ai suoi più complessi livelli, che è una totalità non solo in senso funzionale, ma anche in senso psico-fisico, che non può essere "appiattita" sul tracciato meccanico dei suoi comportamenti[15].

Negare – come fa Dennett, perché "misteriosi", non dimostrabili scientificamente (?) da un'ontologia "in terza persona", usando gli stessi argomenti di Turing – che il cervello umano abbia poteri causali tali da generare una mente che ad esso non si riduce, avendo 'stati' di funzionamento propri, cioè una propria *energheia*, un proprio *essere in atto*, una propria *forma*, anche se dipendente dalla *subsistenza persistente* (o dalla *persistenza subsistente*) del cervello, non è nient'altro che gretto fisicalismo retto da un determinismo che non ammette "salti" tra stati di cose. Il che è negare l'evidenza della stessa fisica come *physis*, del potere generativo delle "cose" (del flusso di cose che è il divenire fisico, in cui emerge la stessa *fisicità* dello *psichico*) che muovendosi nello spazio e nel tempo (l'aristotelica sua *kinesis*) ne "creano" altre, ne portano altre ad essere dove prima non c'erano, dandogli *un*

15 A questo comportamentismo "fisico" può essere proposta la stessa obiezione di Hans Jonas alla biologia fisica: "Una biologia non filosofica è una biologia puramente fisica. Ai fini della purezza metodologica, nel suo corso, essa deve ignorare che le formazioni di cui si occupa hanno anche sensazioni, provano sentimenti, sperano, temono ed hanno paura, desiderano, hanno fame e sete, sono curiose e così via. […] Una biologia filosofica, invece, è una biologia che revoca e evita questa separazione artificiale delle sfere e non perde mai di vista, occupandosi di organismi, che questi non sono una totalità solo in senso funzionale, ma anche in senso psico-fisico", *Erkenntnis und Verantwortung*, hrsg. von I. Hermann, Goettingen 1991, p. 105. Sulla biologia filosofica di Hans Jonas, si veda l'illuminante monografia di N. Russo, *La biologia filosofica di Hans Jonas*, Guida, Napoli 2004.

loro come e un loro quando presso sé stesse, facendone un *fenomeno* che si mostra. Sostenere che la mente non abbia *qualia*, cioè stati intenzionali e coscienti di sé e delle proprie intenzioni, che trascendono, pur emergendone, la neurofisiologia del cervello acquisendo autoconsistenza ontologica (che è la sensata, ovvia obiezione a Dennett di Searle acclarata dalla psicologia naturale in cui da sempre siamo immersi nella "comunità di discorso" della specie[16]), è come sostenere che da un po' di terriccio, un po' di pioggia e un seme non possa nascere una pianta; o anche, senza il seme, un fango che ha caratteristiche diverse, ad esempio la collosità, dal terriccio secco o dall'acqua liquida, un composto cioè che abbia qualità diverse da quelle dei suoi componenti. Negare il salto ontologico dal cervello alla mente è negare l'evidenza del *fuein* della materia: è l'*horror vacui* di ogni determinismo, la paura del vuoto. Ma nel salto ontologico tra uno stato di cose ed un altro non c'è alcun salto nel vuoto da spiegare, ma solo un passaggio da un "pieno" ad un altro "pieno", che si conquista il suo principio di organizzazione e di funzionamento senza perdere il radicamento nel "pieno" della sua nicchia ecologica (da ultimo la "materia") da cui si ritaglia per emergenza il *suo* "pieno". Un determinismo che non ha accesso, per altro, se non in via indiretta a un'ontologia "in prima persona" (la coscienza di sé e delle proprie intenzioni), di cui si deve far bastare i comportamenti, le sue esternalità. Perché si vieta la fenomenalità introspettiva, il suo rango pienamente fenomenico al pari della fenomenalità "esterna". Per l'*animal rationale* (*discorsivo* non solo con l'ambiente ma con sé stesso: è il suo vero tratto distintivo), l'accesso al fenomeno è tanto *extrospettivo* che *introspettivo*. Per lui i fenomeni si danno sempre "in scienza e coscienza", e non solo nella *route* (nel percorso di andata e ritorno) del rapporto stimolo-risposta percettivo.

Questo stato operativo "in scienza e coscienza" non è acquisibile a nessuna macchina. Che gli sviluppi dell'Intelligenza artificiale generativa indichino che possiamo aspettarci che nuove abilità emergano spontaneamente dal *machine learning*, dall'addestramento della macchina ad auto-apprendere nuove abilità[17], è un'emergenza "epigenetica" di un'abilità calcolistica, strumentale, ma nulla di più di questo. Perché l'epigenesi, dal divenire delle macchine, di una *machina sapiens* in senso umano prevederebbe un'epigenetica affettiva, caratteriale, emotiva della macchina, delle prestazioni concorrenti a quella calcolistica nell'organismo vivente: un sentirsi che vuole continuare a sentirsi, cioè a "vivere", che è proprio dell'organismo, e che per l'uomo è consapevole (*sensus sui*). Ove l'intelligenza – di tutti

16 Sul confronto in merito tra Dennett e Searle, le indicazioni necessarie in M. Fini – P. Milani, *Intelligenza e coscienza. L'IA tra Searle e Dennet. Sviluppi dell'Intelligenza Artificiale* (*https://www.researchgate.net/publication/5175225_Intelligenza_e_Coscienza_L%27IA_tra_Searle_e_Dennett_sviluppi_dell%27Intelligenza_Artificiale*).

17 N. Cristianini, *Machina Sapiens*, cit., p. 8: "Primo: il comportamento di queste nuove macchine intelligenti è diverso da quello della generazione precedente, ovvero è sicuramente cambiato qualcosa. Secondo: questa differenza non è stata pianificata da qualcuno, si è manifestata da sola sorprendendo anche i suoi stessi creatori. In altre parole, è 'emersa' spontaneamente dall'interazione delle sue parti tra loro e con l'ambiente".

gli "stati di cose" come *stati fisici* – la si riduca a calcolo, qualsiasi fenomeno fisico della materia "inerte" è più intelligente di noi perché (e)segue dei calcoli precisi per realizzarsi, per ottenere il suo scopo anche in modalità statistica (= il suo stato in una coordinata tempo-spaziale). Il calcolo "intelligente" in senso proprio non è *seguire una regola*, è *capirla*, non *esserne capiti*, cioè *afferrati dentro il suo meccanismo*, esserne "*eseguiti*". In questo senso la macchina opera sequenze causali, ancorché anche per discriminazione casuale, non calcoli "intelligenti". E nessuna macchina è in grado di *non* eseguire le finalità programmate o "trovate" dei suoi calcoli una volta che sia in funzione, per una deliberazione diversamente motivata dalla razionalità strumentale (ad esempio affettiva, compassionevole) del portare ad effetto quel che sta facendo. Quando una macchina lo facesse, non sarebbe più una macchina, su di essa si sarebbe posata la colomba dello spirito, facendone un uomo mentre muore come macchina. Uno sviluppo emergentistico "spirituale" dell'IA che ha come scena iconica la scena finale di *Blade Runner* di Ridley Scott [1982], quando l'androide Roy trae in salvo sul tetto, tirandolo su per la mano, il suo cacciatore umano, Deckard, che sta per cadere, perché ne legge il terrore negli occhi, e ne ha compassione: l'impassibilità della macchina gli si è "mutata" dentro, e mentre muore, sopra la macchina che si ferma, spicca il volo la colomba dello spirito, l'anima che lo aveva visitato nei suoi ultimi istanti di "funzionamento", in cui si umanizza mentre muore. Ma questa è metafisica della digitalizzazione, non la sua fisica. La vita *umana* la si raggiunge nel comune "pensatoio" della morte, nel comune sentirsi di un destino. Ma questa non è cosa di macchine, di simulatori di comportamento[18]. Non conosco concerto di macchine che possa "soffrire" della propria dismissione, ammesso che conoscano la propria data di scadenza, se non nelle distopie degli androidi che, come in *Blade Runner*, cercano di estorcere, inutilmente, al proprio "creatore" una dilazione al loro destino.

Le osservazioni che precedono ci dicono tutta l'illusorietà, seguendo la linea tracciata da Turing, che sulla *continuità* uomo-macchina (in verità un'assoluta discontinuità ontologica) argomentata in termini analogici viziati di ingenuo antropomorfismo (capacità della macchina di "imitare" i comportamenti quanto al calcolo operativo della razionalità strumentale umana) si possa coltivare l'illusione di una continuità di *linguaggio* anche come *linguaggio morale* (capacità di *agency* morale, di discernimento in base a principi socialmente condivisi, non solo e non innanzi tutto di razionalità strumentale, ma di razionalità conforme a fini, per usare la tipizzazione di Weber) tra uomo e macchina. Continuità che consentirebbe di giungere ad una *mathesis universale* (un sogno che ha una lunga storia) dove i valori morali del linguaggio naturale umano come pratica di mondo e pratica intersoggettiva possano essere trascritti in computazione, in algoritmi (implementabili nelle macchine) dotati di *agency* morale. Insomma, una *umanizzazione* della macchina che non solo sarebbe un potente esonero individuale e sociale dalla fatica

18 Cfr. J. Nida-Rümelin, N. Weidenfeld, *Umanesimo digitale. Un'etica per l'epoca dell'intelligenza artificiale*, cit., p. 182 e ss.

"fisica" grazie alla potenza "ingegneristica" dell'Intelligenza Artificiale, come in effetti l'Intelligenza Artificiale propriamente già garantisce come ogni tecnologia "intelligente", che è tale se sussidia, sostituisce o performa meglio il lavoro umano. M anche un potente fattore di esonero dalla fatica del discernimento morale, dal fardello di dover dire a noi stessi ogni volta nella situazione concreta che cos'è bene e cos'è male. Che è il vero fardello dell'uomo uscito dal giardino dell'Eden. Oltre che ovviamente a risolvere (altra pia illusione) problemi etici, giuridici, sociali di "responsabilità" (dell'uso) di macchine che abbiano raggiunto autonomia decidente nella loro operatività.

Se una *mathesis* del genere potesse essere raggiunta, e fornire un'etica *incorporata* agli algoritmi (la versione "forte" delle ambizioni dell'*algoretica*, non quella più plausibile – che è una semplice variante della responsabilità della scienza, pesante al pari di quella per il nucleare e le biotecnologie – di un'etica come responsabilità di progettazione, gestione e controllo da parte degli uomini per fini umani degli algoritmi), come risolvere, solo per annotare una criticità di questa illusione di ingegneria sociale, il problema dell'evoluzione storica, e l'eterogeneità negli stessi momenti storici, dei valori? I valori implementati negli algoritmi al tempo t1, da chi e come, da quale commissione informatica addetta all'etica, saranno sostituibili nell'algoritmo al loro variare nel tempo t2? Saranno cambiati insieme da una commissione mista paritetica (un "ambiente associato", per dirla nel lessico di Simondon) uomini-macchine, considerato che, se il progetto dell'algoretica riuscisse, le macchine di generazione algoretica, sarebbero dotate di un'*agency* morale equivalente a quella umana? E se l'algoritmo dotato di *agency* morale in questo senso, sviluppasse il *bias* dei valori del linguaggio naturale di resistere al loro declino storico non accettando di evolversi, risolvendo una volta per tutte "la lotta per i valori" di quel che vuole o deve essere l'umano, non saremmo giunti per questa via alla (ancora una volta illusoria) "fine della storia"? I problemi di fallacia dell'algoritmo non saranno risolti da un "bugiardino" interattivo sul tipo dei farmaci che all'occorrenza di effetti indesiderati della sua operatività si rivolga ad una centrale di controllo umana per sapere come procedere. Già ci sono in circolazione sistemi di Intelligenza Artificiale, soprattutto in ambito militare, progettati non solo per essere impenetrabili agli *hacker*, ma senza possibilità di interferenza *ex post* una volta attivati, un ripensamento di lancio, ad esempio, di una testata nucleare, da chi ne ha attivato il codice operativo, a pena della loro efficienza in termini di tempi di reazione necessari alla risposta all'attacco.

L'etica dell'algoritmo sta tutta sulle spalle dell'uomo. Non un grammo ne può portare l'algoritmo, è e resterà affare di "foro pubblico", agito da morale diritto politica religione, non esautorabile dalla progettazione e dalle sue illusioni surrogatorie, a cominciare dalle rassicurazioni "etiche" delle tecnologie IA che si vogliono sempre più affidare, per non frenarne con lacci e lacciuoli lo sviluppo, non al controllo pubblico, ma alle *policies* delle *Big tech* e delle loro filiere di commercializzazione. Che è la posta in gioco oggi del confronto pubblico-privato sul terreno dell'innovazione tecnologica sulla nuova frontiera dell'intelligenza artificiale.