



Diritto e innovazione" class="voce">

Giudizio umano, giudizio artificiale

di [Marco Dell'Utri](#)

23 gennaio 2024

Sommario - 1. *Introduzione*. - 2. *Le raccolte informatiche di giurisprudenza, di norme e di dottrina. Il processo telematico*. - 3. *Che cos'è l'intelligenza artificiale?* - 4. *Che cos'è l'intelligenza?* - 5. *L'intelligenza rivelata dalla prestazione (il c.d. test di Turing)*. - 6. *Criteri sintattici e criteri semantici (la stanza cinese di J. Searle)*. - 7. *Il problema della fondazione simbolica (Wittgenstein e le 'forme di vita')*. - 8. *I tratti di una nuova rivoluzione copernicana*. - 9. *Sistemi esperti, machine learning, deep learning. Il rischio dell'opacità*. 10. *Del giudicare come rendere ragione (di sé) e come praxis*. - 11. *La coscienza e le radici del male*.

1. *Introduzione*.

L'incontro tra l'esperienza dell'intelligenza artificiale e l'attività del giudice costituisce uno dei temi di più rilevante interesse nell'ambito del dibattito giuridico-politico contemporaneo, e induce a considerarne gli effetti, destinati a ricadere, al di là dello stile o della qualità del lavoro del giudice, sull'essenza stessa della funzione giudiziaria.

La prospettiva su cui più frequentemente insiste lo sguardo che indugia sull'impiego dell'intelligenza artificiale (nel quadro delle diverse attività sociali) tende per lo più a soffermarsi sul tema della 'sostituibilità' dell'attività umana, nei suoi aspetti o manifestazioni più salienti, attraverso il contributo fornito dall'elaborazione delle macchine.

Della possibilità che le macchine siano (o, più concretamente, saranno) in grado assicurare con immediatezza lo svolgimento di funzioni giudiziarie, converrà, allo stato, limitarsi a rilevare, con

prudenza, il carattere meramente congetturale; e ciò, non tanto (o non solo) per la gravità delle implicazioni d'indole etico-sociale o culturale che appaiono connesse a una simile prospettiva, quanto per la natura ancora preliminare o ipotetica delle conoscenze o delle concrete strategie di elaborazione dei dati rilevanti ai fini di un accettabile o controllabile procedimento di giustizia predittiva.

Si tratta – rispetto a quanto già accade in relazione a molte occupazioni o professioni più agevolmente ‘surrogabili’ dal ricorso alle ‘macchine intelligenti’ – di anticipazioni che, con riguardo all’assunzione di funzioni giudiziarie, si ricollegano più spesso a *test*, o prove, condotte nel quadro di esperienze giuridiche diverse dalla nostra che appaiono ancora insufficienti a giustificare alcun tipo di realistica o concreta preoccupazione.[\[1\]](#)

Verosimilmente, converrà piuttosto insistere – con riguardo al tema più generale delle reazioni suscite dall’impiego dell’intelligenza artificiale - sulla necessità di tenersi distanti, tanto da un’ingenua e illusoria esaltazione entusiastica circa le ‘magnifiche sorti e progressive’ della potenza elaboratrice delle macchine, quanto da un’ingiustificata e insensata prospettiva ‘apocalittica’, destinata ad alimentare unicamente l’incubo della progressiva sostituzione dell’umano attraverso la realizzazione di organismi cibernetici dotati di nuove forme di intelligenza individuale o collettiva, secondo i termini della cosiddetta ‘singolarità’; un disegno capace di presagire, una volta resa autonoma l’elaborazione degli scopi delle singole intelligenze artificiali, il futuro dominio delle macchine sull’uomo e il relativo controllo nello spazio e nel tempo.

Entrambe quelle prospettive, pur largamente fruttuose sul piano dello sviluppo delle ideazioni cinematografiche o letterarie di carattere fantascientifico, devono ritenersi, viceversa, prive di alcun significato concreto sul piano della serietà del discorso culturale, se si eccettua il valore di laboratorio morale, in chiave di sperimentazione mentale, che rimane comunque riscontrabile nell’elaborazione e nella discussione pubblica suscitata dalla diffusione di simili forme d’arte.

2. Le raccolte informatiche di giurisprudenza, di norme e di dottrina. Il processo telematico.

Il riferimento alla nozione dell’intelligenza artificiale *tout court* deve ritenersi, intanto, fermo a una modalità ancora troppo generica di designazione, se si pone mente alle numerose e complesse modalità degli impieghi che accade di riscontrare nelle varie e diverse forme di applicazione dell’intelligenza artificiale.

In ambito giudiziario, varrà intanto sottolineare il carattere largamente sperimentato dell'esperienza informatica delle corti italiane degli ultimi decenni, con particolare riguardo alle applicazioni destinate alla raccolta, conservazione e sistematizzazione critica della giurisprudenza della Corte di cassazione: si tratta di un'attività che, avviata in forma embrionale tra gli anni '60 e '70 del secolo scorso presso il Centro elettronico di documentazione della Corte, è per lo più destinata a fornire a un contributo di informazione immediata sullo stato della giurisprudenza di legittimità in relazione alla risoluzione di singole questioni.[\[2\]](#)

Raccolte analoghe sono state efficacemente organizzate e realizzate, negli ultimi decenni, nell'ambito dell'editoria privata, anche allo scopo di associare, alla raccolta della giurisprudenza di legittimità e di merito, le pubblicazioni di carattere dottrinario già comparse sulle riviste 'cartacee' di maggiore diffusione in ambito nazionale.

Si tratta di iniziative che, quando trascendono le finalità commerciali degli editori privati, si estendono anche alla raccolta *online* della giurisprudenza di altre corti nazionali (di giurisdizione amministrativa) o degli organi di giustizia dell'Unione Europea.

Anche sul piano normativo, le istituzioni italiane hanno provveduto a realizzare efficaci sistemi di raccolta informatica di tutta la normativa primaria esistente, ossia di tutte le leggi e gli atti normativi di carattere legislativo nazionale e regionale, e di gran parte dell'attività regolamentare svolta dagli enti statali e territoriali.

Gran parte degli enti pubblici cui risale lo svolgimento di prerogative di carattere regolamentare gestiscono, peraltro, siti Internet dotati di attendibili raccolte informatiche di contenuto normativo.

Al di là dell'informatica applicata alla raccolta di dati riferiti ai c.d. 'formanti' del diritto, non può non sottolinearsi l'introduzione, da alcuni anni (e, nell'ultimo anno, dinanzi alla Corte di cassazione), del c.d. processo civile telematico, ossia di un sistema capace, non solo di governare lo scambio e l'archiviazione documentale degli atti dei soggetti che partecipano al processo, ma che si estende (talora attraverso il ricorso ad applicazioni fornite da terzi ed acquisite sul piano commerciale dall'amministrazione pubblica) alla celebrazione di udienze mediante collegamento a distanza: una tecnica di interazione tra giudice e parti ripresa e istituzionalizzata (sulla falsariga della legislazione emergenziale emanata durante l'epoca delle restrizioni imposte dalla diffusione del Covid-19), dalle più recenti riforme del processo civile.

Ciascuna di queste esperienze di modalità di impiego dell'informatica in ambito giudiziario deve ritenersi sostanzialmente limitata, così come obiettivamente concepita e concretamente

configurata, all'esecuzione di compiti di ausilio archivistico, di scambio documentale e di comunicazione a distanza.

Rimane invece ancora escluso (allo stato, anche secondo quanto emerge dalla lettura della *Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi* adottata, nel 2018, dalla *Commissione europea sull'efficienza della giustizia* - CEPEJ) che i giudici europei facciano uso quotidiano e regolare (o, quantomeno, ufficiale) di *software di giustizia predittiva in senso stretto*, ossia di *software* capaci - non solo, genericamente, di *ampliare* le informazioni riguardo al passato o al presente, ma anche - di generare una risposta in forma di proposta di decisione in relazione a casi particolari sottoposti alla computazione della macchina.

[\[3\]](#)

3. Che cos'è l'intelligenza artificiale?

Allo scopo di avviare utilmente la considerazione degli effetti o, in generale, dell'incidenza che l'intelligenza artificiale è in grado di assumere sul lavoro del giudice è in primo luogo indispensabile domandarsi *che cosa l'intelligenza artificiale sia e che cosa è possibile attendersi dal suo impiego*.

Nella Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (COM/2021/206 final, 21 aprile 2021) si legge che “*artificial intelligence system' (AI system) means a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments*” (Article 3, § 1(1); testo modificato dal Parlamento europeo il 14 giugno 2023).

In una recente sentenza, il Consiglio di Stato ha definito l'intelligenza artificiale «*un sistema che non si limita solo ad applicare le regole software e i parametri preimpostati (come fa invece l'algoritmo 'tradizionale') ma, al contrario, elabora costantemente nuovi criteri di inferenza tra dati e assume decisioni efficienti sulla base di tali elaborazioni, secondo un processo di apprendimento automatico*».[\[4\]](#)

Prima ancora di approfondire il senso o le modalità di struttura e di funzionamento di tali tecniche di elaborazione di dati, varrà muovere da una preliminare considerazione – se si vuole, ‘filosofica’ – del significato che è possibile ascrivere alla nozione di ‘intelligenza artificiale’.

Secondo una fortunata immagine – ripresa da uno dei nostri maggiori studiosi di filosofia dell'informazione [5] – l'intelligenza artificiale consisterebbe, non già nell'*incontro* tra l'applicazione della funzione intellettuale e la capacità operativa di un dispositivo meccanico (una sorta di unificazione tra funzioni mentali e movimento della materia), bensì nella singolare storia di un '*divorzio*' tra la capacità di agire (intesa come la capacità di organizzare le premesse operative per la risoluzione di un problema e di agire per ottenerla) e l'intelligenza che presiede a tale azione; in breve, le c.d. 'macchine intelligenti' realizzerebbero una capacità di agire (un'*agentività*', secondo l'inelegante termine invalso nel linguaggio specialistico) del tutto priva di intelligenza.

Il riferimento storico-culturale più spesso evocato, a proposito di tale prospettiva, è quello che allude al lavoro di elaborazione svolto attraverso i c.d. 'pensieri ciechi', ossia a quei pensieri che Leibniz, nel corso del XVII secolo, veniva definendo attraverso la scissione (già ritenuta impossibile da Cartesio) tra 'coscienza' e 'conoscenza,' ossia tramite la combinazione di calcolo e logica formale o simbolica.

L'esempio concreto, espressamente proposto dallo stesso Leibniz, è quello del 'chiliagono' (ossia del poligono con mille lati): se l'idea di un simile poligono – diversamente da un triangolo o da un quadrato – non è realisticamente pensabile, immaginabile o tantomeno figurabile, dello stesso poligono è tuttavia possibile calcolare agevolmente i dati e le misure attraverso il ricorso alla logica dei simboli.

In questo caso, secondo la prospettiva leibniziana, il calcolo e il ragionamento condotto sulle caratteristiche di quella figura geometrica prescindono totalmente dalla nostra capacità di pensarne o immaginarne figurativamente il significato.[6]

4. *Che cos'è l'intelligenza?*

La premessa, secondo cui la realizzazione di macchine capaci di trasformare la computazione di simboli in prestazioni operativamente adeguate non sarebbe caratterizzata dall'impiego di alcuna forma di intelligenza propria, induce a interrogarsi ulteriormente su cosa realmente sia, o possa ragionevolmente intendersi per 'intelligenza' o, più precisamente, cosa concretamente occorra affinché sia possibile predicare una simile qualità a un comportamento o a un'azione.

Al di là degli enormi progressi registrati negli ultimi anni dalle scienze umane (dalle discipline cognitive alle neuroscienze, dalle ricerche filosofiche a quelle psicologiche, ecc.), un accordo

sufficientemente ampio o diffuso sulla definizione dell'intelligenza deve ritenersi, allo stato, ancora escluso.

E tuttavia, malgrado l'impossibilità di definire in modo soddisfacente, o scientificamente univoco, la nozione di 'intelligenza', rimane tuttavia ferma la generalizzata capacità degli esseri umani (quantomeno) di 'riconoscere' l'intelligenza in termini pratici, ossia di riscontrarne il ricorso in quei caratteri comportamentali capaci di fornirne un esempio.

Se si immagina cosa possa rendere veramente intelligente un comportamento si è indotti a pensare alla capacità di intendere il senso comune, di ricavare insegnamenti dall'esperienza, alle abilità di apprendimento e di ragionamento, alle competenze comunicative, alla memoria, all'attitudine a vedere in qualcosa qualcos'altro in modo tale da destinarlo a un fine ulteriore, all'acume inferenziale, alla propensione a mettersi nei panni altrui.

Si tratta di forme rivelative dell'intelligenza in cui si esprime la capacità dell'essere umano di 'gestire informazioni' sotto forma di simboli o dati; elementi, questi ultimi, che non sono in alcun modo *privi di interpretazione*, quanto piuttosto *pattern* (schemi) informativi dotati di un significato loro proprio.

L'essere umano vivente è immerso nel tempo, con le sue specifiche scansioni naturali e culturali da articolare: ha storia, memoria, immaginazione, senso delle possibilità e delle alternative, scarti di imprevedibile creatività, capacità di riflettere su sé stesso in base a una coscienza che non ha nulla a che vedere con i pensieri ciechi di leibniziana memoria.

In un testo d'indole divulgativa pubblicato alla metà degli anni '90 [\[7\]](#), Daniel Goleman prospettava la possibilità di guardare all'intelligenza e al suo significato attraverso la valorizzazione delle componenti emotive dell'affettività; momenti dell'esperienza individuale chiamati a rivestire una rilevanza non minore, per la vita sociale, di quella che siamo soliti attribuire all'intelligenza razionale o cognitiva.

Se ci soffermiamo ad osservare gli esseri viventi come esempi di chi 'sa stare al mondo', ci accorgiamo del carattere dirimente e decisivo del possesso, da parte di ciascun essere umano, di una propria 'storia': è nel corso (e in conseguenza di) questi irripetibili processi storici che ciascun essere umano forma la propria abilità di ragionare in modo brillante o meno, o anche di muoversi in maniera imprevedibile o acrobatica.

Ma i percorsi di quelle storie costituiscono, allo stesso modo (e contemporaneamente), anche le ragioni (o le cause) dei 'difetti' degli esseri umani, dei diversi temperamenti, delle irriducibili

sensibilità e delle credenze che giustificheranno le sensazioni del piacere e del dispiacere, dell'amore, della sofferenza e del dolore; in breve, le ragioni di tutte le differenze individuali destinate ad arricchire i contenuti delle relazioni umane e della convivenza.

Ed è proprio attraverso questa dimensione relazionale e la convivenza, come *strutture costitutive dell'umano*, che le comunità elaborano, collettivamente, nella reciproca capacità di dare e di ricevere, il senso dello stare al mondo e d'intenderne i significati in ambiti e contesti che appaiono fortemente connotati sul piano emozionale.

Se dietro ciascun individuo, così come dietro ciascuna comunità umana, è possibile ricostruire il senso di ‘una storia’, dietro ogni macchina si trovano ‘tante storie’ quanti sono i dati che ne alimentano l’attività di computazione: nessuna di tali singole storie è, tuttavia, ‘la storia’ di quella macchina e, dunque, nessuna linea coerente di significati e nessuno spazio di convivenza collettiva varranno a sostanziare il senso, temporalmente disteso, della sua esperienza di elaborazione.

Molti ricercatori vedono nel progresso della robotica una strada percorribile per il perfezionamento delle intelligenze artificiali: dotandosi di un corpo, le macchine potrebbero percepire e agire nel mondo, e in tal modo costruirsi un modello di causalità simile a quanto avviene nel mondo animale: si tratta di raggiungimenti che si prospettano concretamente attraverso una stretta collaborazione tra le neuroscienze sensoriomotorie, la robotica e l’intelligenza artificiale.

E tuttavia, anche in questo caso, il cammino che si prospetta non appare affatto così agevole: l’ostacolo più arduo, infatti, sembra rappresentato dalla condizione per cui l’implementazione delle capacità sensoriomotorie della macchina richiederebbe (diversamente da quanto avviene per l’esercizio delle nostre attività mentali) una quantità di calcolo incredibilmente spropositata.

Secondo l’informatico e ricercatore canadese Hans Moravec, ciò è dovuto alla circostanza secondo cui le porzioni di cervello animale dedicate alle abilità sensoriali e motorie sono costruite su centinaia di milioni di anni di interazioni con il mondo, là dove le abilità mentali di alto livello costituiscono un’innovazione relativamente recente, molto più semplici e quindi replicabili in una macchina.[\[8\]](#)

5. L’intelligenza rivelata dalla prestazione (il c.d. *test di Turing*).

In assenza di un criterio oggettivo capace di fornire una definizione ‘sostanziale’ di intelligenza sufficientemente condivisa, a partire dagli anni ‘50 si è arrivati, in ambito informatico, a distinguere l’intelligenza del comportamento della macchina attraverso la relativa sottoposizione al c.d. *test di Turing* e alla verifica della relativa capacità di ‘comportarsi come un uomo’, ossia di ‘simulare’ in termini pratici i risultati dell’intelligenza umana.

In termini elementari, il *test* di Turing è organizzato attraverso la collocazione, in uno spazio isolato, di un uomo a cui è attribuito il compito di porre domande (ad es., attraverso un *computer*) e, in un diverso spazio separato, di un altro uomo e di una macchina dotata di intelligenza artificiale: ciascuno di questi due ultimi soggetti (il secondo uomo e la macchina) è tenuto a rispondere alle domande del primo uomo; in tal caso, se nel giro di breve tempo l’uomo che pone le domande non è in grado di comprendere quale dei due soggetti che risponde sia la macchina, e quale l’essere umano, si potrà affermare che la macchina ha superato il *test* di Turing.

Di recente, il laboratorio di ricerca della società *OpenAI Inc.* ha diffuso in rete (al libero e gratuito accesso della collettività) uno strumento di conversazione dotato di intelligenza artificiale (*ChatGpt*), attraverso il quale a ciascun utente è aperta la possibilità di procedere autonomamente al tentativo di un (simulacro di) *test* di Turing.[\[9\]](#)

Digitando nell’ottobre del 2023 [\[10\]](#) la domanda: “*Il veicolo non entra nel parcheggio perché è troppo piccolo. Che cosa è troppo piccolo?*”, la risposta del programma è stata: “*Il veicolo è troppo piccolo. Potrebbe essere che il veicolo sia troppo corto o stretto per il parcheggio*”.

Alla successiva digitazione (sempre nel medesimo periodo) della domanda: “*Il piede non entra nello stivale perché è troppo piccolo. Che cosa è troppo piccolo?*”, la risposta è stata: “*Il piede è troppo piccolo: probabilmente il piede è di dimensioni tali da non riempire tutto lo stivale*”.

Appare evidente come nel fornire risposta a quelle domande – comprensibili unicamente attraverso la ricerca del significato della proposizione premessa alla frase interrogativa (che allude al decisivo valore semantico del verbo ‘*entrare*’ rispetto al rapporto tra le dimensioni del veicolo o del piede e quelle del parcheggio o dello stivale) – la macchina non comprende il significato delle informazioni fornite, limitandosi a computarle secondo modalità puramente sintattiche (ossia attraverso l’applicazione delle regole di cui la stessa macchina dispone) formulando, di conseguenza, la propria risposta.

Nel caso di specie, non riuscendo a individuare il soggetto sottinteso della seconda proposizione (“*perché è troppo piccolo*”) attraverso un’attendibile criterio capace di discriminare le ricorrenze di *veicolo*, *parcheggio*, *piede* o *stivale* “*troppo piccolo*” (rispetto al verbo ‘*entrare*’), la macchina ha

fatto ricorso all'unico criterio disponibile, ossia al criterio logico-grammaticale costituito dal soggetto (espresso) della prima proposizione (il *veicolo*, il *piede*), giungendo tuttavia, in tal modo, a un esito (da considerare) ragionevolmente paradossale (il *veicolo* non entra nel parcheggio perché *lo stesso veicolo* è troppo piccolo; il *piede* non entra nello stivale perché *lo stesso piede* è troppo piccolo).

Se solo la macchina avesse *compreso il significato* del verbo 'entrare', posto in rapporto al tema delle dimensioni del veicolo e del parcheggio o del piede e dello stivale, non avrebbe avuto difficoltà, rispetto alla dimensione della 'piccolezza', nel discriminare tra il *veicolo* e il *parcheggio* o tra il *piede* e lo *stivale*.

Infatti, ponendo (sempre nello stesso periodo) le domande: "La macchina non entra nel parcheggio perché è troppo piccolo. Che cosa è troppo piccolo?" e la domanda "Il piede non entra nella scarpa perché è troppo piccola. Che cosa è troppo piccolo?", le risposte sono state: "Il parcheggio è troppo piccolo. Probabilmente perché lo spazio disponibile nel parcheggio è limitato" e "La scarpa è troppo piccola perché, evidentemente, di dimensioni inferiori rispetto a quelle del piede".

In tali ultimi esempi, l'elaboratore non ha avuto dubbi nel rispondere che il parcheggio e la scarpa sono 'troppo piccoli': facilitata, nel primo caso, dalla distinzione di genere (femminile-maschile) tra il soggetto espresso della prima proposizione (*la macchina-femminile*) e l'aggettivo della seconda (*piccolo-maschile*) e, nel secondo caso, attraverso il dato della declinazione al femminile dell'aggettivo 'piccola' (femminile) rispetto al genere del soggetto della proposizione principale, *il piede* (maschile), la computazione non poteva che riannodare i legami tra gli aggettivi e i sostantivi valendosi del criterio del genere, con la conseguente affermazione della 'piccolezza' del *parcheggio* (e non già della *macchina*) e della *scarpa* (e non già del *piede*).

In ogni caso, la comprensione *semantica* del verbo 'entrare', posto in rapporto al tema delle dimensioni della macchina e del parcheggio o del piede e della scarpa, deve ritenersi rimasta, ai fini delle risposte fornite dall'elaboratore, del tutto esclusa.

6. – *Criteri sintattici e criteri semantici (la stanza cinese di J. Searle).*

Le argomentazioni sin qui articolate – unitamente all'oggettivo rilievo della sperimentata incapacità delle intelligenze artificiali attualmente in circolazione di superare stabilmente il *test* di Turing – consentono di ribadire come l'attività della c.d. 'intelligenza artificiale' proceda

secondo criteri che sono propriamente *sintattici* (ossia che prevedono la mera applicazione di regole predeterminate nella gestione dei simboli utilizzati) e non già *semantici* (ossia legati alla comprensione dei significati di detti simboli).

In breve, mentre le macchine appaiono pienamente capaci di processare molto velocemente impressionanti quantità di dati, le stesse non sono in grado di comprendere il significato dei dati che processano; e ciò, almeno fino a quando non sarà possibile realizzare il progetto di una futuribile ‘*macchina semantica*’.

Sul versante di tale interpretazione dell’intelligenza artificiale, uno dei più autorevoli e qualificati filosofi della mente e del linguaggio contemporanei, John Searle, ha proposto il noto esperimento mentale della c.d. *stanza cinese*.

Secondo tale esperimento, si immagina la collocazione di un uomo di lingua inglese (che conosce solo tale lingua) in una stanza diversa da quella occupata da un uomo cinese che comunica con lui, in lingua cinese, attraverso un *computer*.

Si assume, inoltre, che l’uomo di lingua inglese, non comprendendo una sola parola di cinese, utilizzi uno speciale (e prodigioso) libro di regole, capace di fornirgli, in lingua inglese, le istruzioni migliori sulle risposte (con caratteri cinesi) più adeguate alle domande ricevute, senza peraltro rivelare il significato dei caratteri cinesi della domanda e della risposta.

Si immagina, infine, che la conversazione tra i due proceda, via *computer*, per diversi minuti con piena soddisfazione dell’interlocutore di lingua cinese: in tal caso potrà dirsi che l’uomo di lingua inglese ha saputo sostenere, per un tempo significativo, una produttiva conversazione in lingua cinese, pur non capendo assolutamente nulla di ciò su cui si era conversato.

Secondo Searle, l’uomo di lingua inglese si è comportato, in tale vicenda, esattamente come fosse un *computer* chiamato a calcolare una formula sulla base di un programma e di simboli formali. Non era dunque necessario che comprendesse ciò che doveva fare, o il significato dei simboli utilizzati, poiché, ai fini della buona riuscita della conversazione, gli era unicamente richiesto di seguire le istruzioni precedentemente fornitegli.

7. – Il problema della fondazione simbolica (Wittgenstein e le ‘forme di vita’).

Lo scoglio su cui è destinata a infrangersi ogni pretesa dell’intelligenza artificiale di adeguare l’intelligenza umana è, dunque, uno scoglio d’indole *semantica*: l’intelligenza artificiale non ha (o,

se si vuole, non ha ancora) alcuna capacità di attribuire significati ai dati che le vengono sottoposti per la computazione.

Va da sé che una simile incapacità non è nient'altro che la corrispettiva incapacità dell'uomo di attribuire alla macchina la disponibilità di una simile competenza.

Tale incapacità, d'altro canto, discende, a sua volta, dall'impossibilità, allo stato, di rispondere, in modo oggettivamente univoco, all'interrogativo che attiene al modo attraverso il quale i dati o, in generale, i simboli linguistici elaborati dall'uomo nell'ambito delle proprie relazioni (e poi trasmessi alla macchina), acquistano il proprio significato; un interrogativo noto, nel campo dell'intelligenza artificiale, come il problema della '*fondazione simbolica*'.

Anche con riguardo a tale problema occorre riconoscere, in capo alle scienze umane, un non (ancora) estinto debito conoscitivo: non esiste, infatti, (ancora) alcuna chiara o univoca comprensione (al di là delle pur rilevantissime e promettenti ipotesi di studio prospettate al dibattito scientifico) del modo in cui il problema della *fondazione simbolica* (ossia la risposta alla domanda sul 'come' i dati, o talune informazioni, acquistino significato) è risolto dagli stessi esseri umani.

Ciò che è unicamente noto è che la capacità di processare informazioni dotate di significato (e di comprenderne pienamente il senso) è precisamente quello in cui agenti intelligenti come gli esseri umani sono in grado di primeggiare in assoluto.

Tutti gli esseri umani completamente e normalmente sviluppati sembrano avvolti nel 'bozzolo' del proprio spazio semantico.

In senso stretto, gli uomini non apprendono mai consapevolmente 'puri dati' privi di significato. La percezione genuina di dati totalmente non interpretati, benché possibile in particolari circostanze, non costituisce la norma, e non può essere parte di un'esperienza cosciente a lungo sostenibile.

In breve, l'uomo non percepisce mai puri dati colti *in isolamento*, ma sempre in quanto *parte di un contesto semantico* che, inevitabilmente, li dota di un significato determinato.

Quelli che vanno sotto il nome di 'dati grezzi' sono i dati che mancano di una *specifica e rilevante* interpretazione, ma non di *qualsiasi* interpretazione.

Sappiamo molto poco del modo in cui costruiamo con successo, attraverso il linguaggio, le storie dense di informazioni che ci tengono insieme e di cui siamo parte.

Su tale punto, Ludwig Wittgenstein ci ha insegnato, attraverso la revisione, nelle *Ricerche filosofiche*, delle tesi contenute nel *Tractatus logico-philosophicus*, che il significato delle espressioni del linguaggio che normalmente utilizziamo dipende dal contesto storico specifico in cui quelle espressioni vengono utilizzate.

Se non esiste ‘il’ significato del linguaggio è perché il significato del linguaggio è una concreta *forma di vita*.

Il senso delle nostre azioni e della nostra vita risiede in uno specifico ambito storico-relazionale, ossia nelle particolari *forme di vita* entro le quali siamo immersi: “Si può imparare a conoscere gli uomini?” si domanda Wittgenstein: “Sì, qualcuno può farlo. Non però attraverso un corso d’insegnamento, ma attraverso l’*“esperienza”*”.[\[11\]](#) Il significato delle parole è dato dall’uso che se ne fa nell’esperienza e non da una pretesa corrispondenza tra segni e cose. L’esperienza si frammenta in regioni di esperienza (‘forme di vita’), all’interno delle quali hanno luogo i corrispondenti giochi linguistici: parlare un certo linguaggio fa parte di una certa *forma di vita*.

[\[12\]](#)

Se le cose stanno così, allora, agenti umani e agenti artificiali appartengono a mondi differenti e ci si può attendere che questi ultimi, non solo abbiano capacità dissimili dalle nostre, ma che commettano anche errori di altro genere.

Il comportamento intelligente si basa sulla comprensione dei significati, più che sulla manipolazione sintattica (ossia secondo regole) di simboli, e gli agenti che operano in termini semantici o in termini sintattici, mentre possono entrambi raggiungere facilmente i medesimi obiettivi in modo efficiente e con successo, sono inclini a incorrere in differenti tipologie di possibili errori.

Nonostante i tentativi che ancora vengono svolti da molti ricercatori e studiosi, l’intelligenza artificiale non dovrebbe tentare di ‘simulare’ il comportamento umano intelligente: l’intelligenza artificiale dovrebbe piuttosto *cercare di emulare i suoi risultati*: è noto, al riguardo, il commento dell’informatico olandese Edsger W. Dijkstra, secondo cui “chiedersi se un computer possa pensare è tanto interessante quanto chiedersi se un sottomarino possa nuotare”.[\[13\]](#)

Seppure, infatti, non disponiamo di tecnologie esperte dal punto di vista semantico, la memoria sopravanza l’intelligenza sul piano della prestazione, sì che tecnologie puramente sintattiche possono eludere il problema del significato e della comprensione, offrendoci ciò di cui abbiamo bisogno con grande efficacia (una traduzione, un ristorante preferito, un libro interessante, un biglietto a un prezzo migliore, e così via).

8. I tratti di una nuova rivoluzione copernicana.

Si è in precedenza accennato al *test* di Turing e si è notata la stretta inerenza tra l'appropriatezza delle modalità di formulazione dei quesiti posti alle ‘macchine intelligenti’ e la qualità delle risposte acquisibili.

La necessità di conformare le domande in modo coerente alla struttura e alle modalità di funzionamento delle macchine, al fine di ottenere risposte adeguate (si è visto come la domanda sul parcheggio e sul veicolo debba essere posta in modo tale da non trarre in inganno l’elaboratore chiamato a conversare con un umano), esprime, in termini elementari, la stessa necessità che, a livello politico, sociale od economico, governa la costruzione degli ambienti più adatti al corretto ed efficiente funzionamento delle macchine dotate di intelligenza artificiale alle quali sempre più frequentemente chiediamo prestazioni.

Si pensi agli ambienti industriali creati attorno alle macchine che realizzano prodotti in serie, o si pensi, più semplicemente, alla conformazione degli elettrodomestici che circondano la vita quotidiana delle persone; macchine che richiedono, per il loro funzionamento, l’esecuzione di azioni che, benché elementari, appaiono ben diverse da quelle eventualmente richieste per la realizzazione diretta, da parte dell’uomo, delle medesime prestazioni.[\[14\]](#)

Si tratta di esempi che rendono evidenti le necessità imposte dalla convivenza o dalla collaborazione degli esseri umani con le macchine dotate di intelligenza artificiale: nel cooperare con simili macchine è indispensabile che l’uomo adegui, alle esigenze operative di quelle, le variazioni dei propri comportamenti, stilizzando le movenze della propria esperienza di relazione con le macchine secondo la logica di queste.

Appare del tutto intuitivo – presa coscienza della crescente pervasività dei sistemi di intelligenza artificiale che circondano i numerosi e diversi ambiti della nostra esperienza personale e collettiva – come l’impegno ripetuto e costante all’adeguamento dei comportamenti alle esigenze operative delle macchine comporti il rischio del c.d. ‘*effetto gregge*’, ossia dell’induzione, con il passare del tempo, di stili di vita e di abitudini di ragionamento confinati alla riflessione sulle sole operazioni di ausilio alle principali prestazioni della macchina e, dunque, il rischio di un crollo della consapevolezza individuale e collettiva circa il carattere autonomo o indotto (e quindi eterodiretto) dei nostri stessi pensieri.

Si tratta della squilibrata prospettiva secondo cui il *General Intellect* (di marxiana memoria) incorporato nelle macchine finisce per conferire forma alle relazioni e di distorcere i comportamenti umani, in un contesto in cui il funzionamento effettivo dei sistemi (commerciali, finanziari, economici, amministrativi, politici, ecc.) in tanto è possibile, in quanto tutto (pensieri e comportamenti umani inclusi) è conformato attorno alle macchine.

Avvolgendo il mondo, le nostre tecnologie conformano i nostri ambienti fisici e concettuali, inducendo ad adattare a loro i nostri comportamenti e i nostri pensieri, poiché questa diviene la via migliore e più facile (o talora l'unica) per far funzionare le cose.

A lungo andare, tuttavia, la mancata sorveglianza sui pensieri e sui comportamenti o, più in generale, la carente consapevolizzazione delle logiche di funzionamento delle macchine con cui dividiamo il nostro tempo (con la conseguente necessaria calibrazione delle aspettative di prestazione) non potrà che favorire, pericolosamente, la generalizzazione di gravi e rovinosi processi di alienazione intellettuale.

9. Sistemi esperti, machine learning, deep learning. Il rischio dell'opacità.

Alla luce delle analisi che precedono è possibile tornare, con un grado di minore ingenuità, alla considerazione del rapporto tra l'esperienza professionale del giudice e l'impiego dell'intelligenza artificiale in ambito giudiziario.

Si è detto come l'uso dei sistemi informatici nei contesti giudiziari sia attualmente limitato alla richiesta di prestazioni contenute alle funzioni di archiviazione documentale o alla ragionata selezione, attraverso la mediazione di operatori logici, dei dati elementari utili alla ricostruzione di indirizzi giurisprudenziali funzionali alla risoluzione di questioni giuridiche astrattamente considerate.

Si è altresì accennato alla perdurante esclusione, allo stato, (anche secondo quanto emerge dalle registrazioni della *Commissione europea sull'efficienza della giustizia* - CEPEJ) di un uso quotidiano e regolare (o, quantomeno, ufficiale), da parte dei giudici europei, di *software* di *giustizia predittiva in senso stretto*, ossia di *software* capaci (non solo, genericamente, di *ampliare* le informazioni riguardo al passato o al presente, bensì anche) di generare una risposta in forma di proposta di decisione in relazione a casi particolari sottoposti alla computazione della macchina.

Le ragioni più spesso associate alle valutazioni positive sulla giustizia predittiva (quando non all'incoraggiamento volto alla relativa introduzione nella pratica quotidiana) appaiono per lo più argomentate con riguardo agli innegabili vantaggi che l'uso di un simile strumento sarebbe in grado di assicurare sul versante della prevedibilità delle decisioni e della certezza del diritto, oltre che dell'incomparabile rapidità dei processi di computazione e di proposizione delle soluzioni che l'uso delle macchine intelligenti verosimilmente varrebbe a garantire.

Gli studiosi dei temi della giustizia predittiva sono soliti sottolineare, in termini elementari, la fondamentale distinzione tra forme di intelligenza artificiale strutturata sulla base di *sistemi esperti* e forme di intelligenza artificiale strutturata sulla base di sistemi realizzati per *l'apprendimento automatico*.[\[15\]](#)

Nei primi casi, la conoscenza del sistema è fornita direttamente dall'uomo: si tratta di sistemi caratterizzati dalla capacità di fornire una rapida cognizione dello stato delle conoscenze già acquisite, eventualmente aggregate e riaggregate secondo criteri predeterminati, ma sostanzialmente incapaci di generare risposte nuove, dotate di informazioni di per sé non contenute in quelle immesse originariamente nel sistema attraverso i *dataset* di input.

La contenuta efficienza mostrata dall'impiego dei *sistemi esperti* nell'ambito dell'attività predittiva può trovare la sua spiegazione nell'enorme difficoltà di formalizzare *a priori* il ragionamento giuridico in schemi passibili di essere omogenizzati nelle regole di una sintassi, sì da riconoscere e trattare in modo soddisfacente il contenuto di casi nuovi mai precedentemente esaminati.

Nelle forme dell'intelligenza artificiale strutturata sulla base di *sistemi realizzati per l'apprendimento automatico*, viceversa, la conoscenza del sistema è il frutto di un'autonoma capacità della macchina di estrarre le informazioni rilevanti dai dati di *input* e così imparare a svolgere in modo adeguato le funzioni a esse affidate.

A lora volta, i sistemi per l'apprendimento automatico possono differenziarsi, in termini generalissimi, nelle forme dell'*addestramento supervisionato, per rinforzo* e dell'*addestramento non supervisionato*.

Nell'ipotesi dell'*addestramento supervisionato*, il sistema può essere, ad es., nutrito di dati tratti da casi che si ritengono risolti correttamente: adottando questa forma il sistema finirà col 'ragionare' e rispondere secondo la logica dei casi forniti, ed è intuibile il rischio di ottenere risultati che tendono a risolversi in una sostanziale 'pietrificazione' della giurisprudenza formata secondo logiche inevitabilmente conservative.

Nella forma dell'*addestramento per rinforzo* il sistema è viceversa addestrato per rispondere secondo una *propria* elaborazione dei dati forniti e riceverà, dall'uomo che se ne prende cura, una risposta positiva o negativa al suo *output*; tali risposte saranno in seguito utilizzate dal sistema per regolare il proprio comportamento futuro in base ai *feedback* della propria stessa attività. In tal caso, il sistema risponderà in coerenza ai valori e alle logiche di coloro ai quali è stata attribuita la responsabilità di fornire i dati di *feedback*.

L'*addestramento non supervisionato*, infine, è il sistema nutrito attraverso il ricorso ad un'ampia base di dati (ricorre talora l'uso della controversa espressione '*big data*') dalla cui elaborazione saranno ricavati (come primo *output*) schemi formalizzati di ragionamento che costituiranno, a loro volta, la base per la formulazione delle risposte ai casi nuovi.

La struttura di *machine learning* da ultimo sommariamente descritta è quella verosimilmente tenuta presente in quei tentativi di definizione dell'intelligenza artificiale, cui si è inizialmente accennato, da parte del Parlamento e del Consiglio Europeo, e del Consiglio di Stato italiano, in cui si evidenziano i caratteri più significativi delle forme attualmente più diffuse di intelligenza artificiale: i vari livelli di autonomia (*varying levels of autonomy*), la costante elaborazione di nuovi criteri di inferenza tra dati e l'automaticità dei processi di apprendimento.

Sono definizioni che se pure evidenziano, dell'intelligenza artificiale di più recente generazione, la facoltà di compensare l'incompletezza delle informazioni ricevute in partenza attraverso l'apprendimento, o la capacità di assumere decisioni in situazioni di irriducibile incertezza, allo stesso tempo sottolineano come l'assolvimento di tali funzioni avvenga in un ambito di perdurante imperscrutabilità dei meccanismi causali interni.

In breve, mentre è possibile individuare agevolmente gli *input* e gli *output* del sistema, rimane del tutto occulto 'cosa' accade all'interno della 'scatola nera', ossia i termini e le forme della catena causale che determina il passaggio da quegli *input* a quegli *output*.

Le forme di intelligenza artificiale alle quali si attribuiscono le maggiori potenzialità consistono in applicazioni dotate di capacità di previsione, come i *robot* di ultima generazione e come tutti i dispositivi capaci di *machine learning*, ossia di apprendimento continuo, grazie alle sofisticate tecniche di *deep learning*, alla raccolta massiva di dati (*big data*) e al collegamento al *cloud*, ossia a quell'insieme di *server* remoti che offrono servizi per l'archiviazione e la gestione delle informazioni: un potente *hub* computazionale in grado di conservare, elaborare ed erogare enormi masse di dati, da cui attingere continuamente per gli aggiornamenti (gli *upgrading*).

Da questa prospettiva, “i *learner*, come sono definiti gli algoritmi di apprendimento, sono degli ottimi studenti. Essi consistono generalmente in linee di codice che vanno dalle centinaia ai milioni, e sono ricavati induttivamente, grazie ai *big data*, da imponenti flussi di dati che generano poi istruzioni precise e univoche. A loro volta, tali dati sono il materiale cui attingono gli algoritmi di apprendimento”.[\[16\]](#)

E tuttavia, “la possibilità di cambiare la struttura del sistema apre, tuttavia, virtualmente la strada a un tipo di IA in grado di trascendere i fini che le assegnano gli umani e di agire, di conseguenza, secondo una logica propria, per noi diventata insondabile”.[\[17\]](#)

Questa dimensione di imperscrutabilità o di insondabilità dei meccanismi di elaborazione autonoma delle macchine, che vanno dal processo dei *big data* alla costruzione dell’*output*, costituisce, in particolare, una delle più rilevanti criticità del *deep learning* strutturato attraverso le c.d. *reti neurali*, ossia attraverso quell’insieme di ‘nodi’ variamente legati tra loro che veicolano informazioni numeriche, a loro volta rappresentative di una certa valutazione delle informazioni (dati) che vengono volta a volta sottoposte alla macchina.

Una delle caratteristiche più insidiose di questi sistemi è appunto rappresentata dall’opacità delle operazioni che vengono svolte nella computazione delle informazioni, poiché spesso taluni passaggi tra i nodi non risultano spiegabili o adeguatamente comprensibili per il ragionamento umano, quando non offuscati dalle privative industriali legate alla collocazione del prodotto informatico sul mercato.

In termini banali, all’interno di un sistema del genere la descrizione dei fatti di una controversia diventa un insieme di vettori numerici, che debbono essere collegati al possibile esito della stessa, secondo le correlazioni statistiche elaborate dall’algoritmo di apprendimento automatico sulla base delle combinazioni tra descrizioni di fatti ed esiti presenti negli insiemi (*dataset*) di addestramento.

Trattandosi di sistemi opachi rispetto all’utilizzatore, questi sistemi assicurano una risposta, ma non forniscono spiegazioni.

Nei casi in cui a un giudice fosse affidato il compito di interagire e cooperare, nell’esercizio delle proprie funzioni, con macchine strutturate secondo schemi così definiti, parrebbe assolutamente indeclinabile la necessità che a quel giudice sia stata preventivamente assicurata la possibilità di comprendere con chiarezza la logica di funzionamento e lo stile di computazione del sistema di cui si avvale.

In breve, occorrerà che il giudice disponga di una piena consapevolezza della natura dello specifico *dataset* di *input* con il quale sistema è stato alimentato; dei criteri che hanno governato la ricerca e l'analisi di tale *dataset*; dei caratteri dei processi di successiva selezione algoritmica attraverso cui i dati sono destinati a transitare; degli obiettivi diretti e delle finalità cui sono ordinati i diversi passaggi, o nodi, del sistema, etc.: si tratta di elementi conoscitivi indispensabili affinché il giudice possa esercitare la propria *autonoma capacità di valutazione critica* del significato dell'*output* di sistema.

A mero titolo di esempio, se un determinato sistema risulta progettato (in ragione delle modalità specifiche dell'analisi dei dati che lo alimentano e delle strutture algoritmiche che presiedono alla formulazione della decisione di *output*) allo scopo di prevedere l'esito possibile di certi casi sulla base del passato, un simile sistema potrà soddisfare l'obiettivo di anticipare la sentenza del 'giudice medio': un obiettivo che, mentre avvicina l'interesse dell'utilizzatore avvocato o cittadino (là dove interessati a prevedere il probabile comportamento futuro del giudice, sì da impostare di conseguenza il proprio comportamento processuale), assume un'importanza assai più limitata per il giudice, il cui compito è, viceversa, quello di adottare una decisione giuridicamente corretta.

Da qui l'esigenza di comprendere in profondità, nel quadro di un'analisi dei rapporti tra il giudice e l'intelligenza artificiale, il senso stesso di quella particolare attività in cui consiste "*il giudicare del giudice*".

10. *Del giudicare come rendere ragione (di sé) e come praxis.*

Si è detto del carattere indispensabile, per il giudice, dell'acquisizione delle nozioni elementari per la comprensione delle modalità di funzionamento del modello di intelligenza artificiale con cui si trova ad interagire; e si è sottolineata l'indeclinabilità del compito di padroneggiare le forme di funzionamento dei sistemi di intelligenza artificiale e il significato dei relativi risultati, al fine di farne materia di una successiva elaborazione in sede decisionale.

Si tratta di questioni che assumono un rilievo centrale ai fini della meditazione sull'intelligenza artificiale e sui suoi riflessi sul 'giudicare'.

Sarà bene evidenziare come, a differenza di gran parte delle attività professionali o, in genere, delle attività lavorative o delle funzioni sociali in cui assume valore preponderante il risultato dell'attività (il prodotto) rispetto al procedimento produttivo, con riguardo alla funzione

giudiziaria l'esito del lavoro, in tanto ha valore (e conta), in quanto il procedimento per realizzarlo abbia rispettato una o più condizioni o *standard* di natura qualitativa; nel senso che, più che il risultato, occorre aver riguardo, preliminarmente, all'itinerario seguito allo scopo di raggiungerlo o di definirlo.

In breve, benché sia del tutto ovvio che l'efficacia concreta del *dictum* del giudice si ricolleghi con immediatezza al contenuto del 'dispositivo' di un provvedimento (la condanna a una pena o al pagamento di una somma di denaro; l'annullamento di un contratto; l'assoluzione da un'imputazione di reato), tale dispositivo, in tanto è destinato a sussistere (e a produrre effetti), in quanto costituisca l'esito di un percorso motivazionale logicamente congruo e comprensibile che lo sostiene e lo giustifica.

Non aiuta, in questo senso, una certa (diffusa) inclinazione 'produttivistica' nella considerazione dei temi del processo e del giudizio. L'accostamento del processo all'idea della 'produzione' induce a guardarvi come a un'attività 'servente' o 'strumentale' alla realizzazione di fini ad essa estranei; un impegno volto, attraverso la tecnica giudiziaria, al compimento del prodotto, e dunque del giudizio (la decisione, la sentenza), di regola chiamato a tradurre, in termini matematici, uno dei principali indici di misurazione della produttività del giudice.

Un'antica tradizione di origine aristotelica[18] – la cui più recente riscoperta ha costituito un tratto essenziale di una parte significativa del pensiero etico-politico del Seconde Novecento[19] – invita a distinguere, della vita pratica dell'uomo, l'attività produttiva (la *poiesis*, governata dalla *techne*) dalla prassi (la *praxis*), avente se stessa quale propria finalità: l'attività per cui l'uomo pone la propria stessa azione come oggetto di un percorso di graduale educazione e perfezionamento, attraverso il governo (non già della 'tecnica', bensì) della 'saggezza' (*phronesis*).

Produzione e prassi (*poiesis* e *praxis*) valgono a distinguersi dunque in ciò, che il sapere teorico generale di cui l'uomo dispone è destinato, nella produzione che si avvale della tecnica (*techné*), a trasferirsi sulle cose allo scopo di trasformarle in conformità ad esso; per cui il prodotto, come fine di per sé estraneo all'attività produttiva, diviene lo specchio (o, meglio *uno* specchio) concretizzato del sapere teorico.

Nel caso della prassi, governata dai canoni della saggezza (*phronesis*), il sapere teorico generale è viceversa chiamato a combinarsi o a 'contaminarsi' con la realtà, con le circostanze e le vicende del mondo, affinché sappia modificarle, ma insieme anche lasciarsene modificare, sì da dar luogo a una nuova forma di sapere capace, con saggezza, di coniugare, e tenere insieme, il generale e il particolare.

Seguendo il filo di queste linee argomentative, alla descrizione dell'attività processuale sembra dunque convenire la qualificazione nei termini di una *prassi* (di una *praxis*), ossia di una specifica attività pratica, governata dalla saggezza, che ha fine in sé stessa.

Ogni atto del processo è il giudizio stesso (in taluni contesti, ‘processo’ e ‘giudizio’ sono usati come sinonimi): in realtà, il processo è il giudizio che si va facendo in un tempo e in un luogo determinati, in una dimensione spazio-temporale specificamente qualificata.

Da questo punto di vista assume un preciso significato la regola costituzionale per cui “tutti i provvedimenti giurisdizionali devono essere motivati” (art. 111 Cost.): si tratta di un principio in cui si riflette la necessità che, in un sistema democratico e liberale, ogni decisione costituisca l'esito (ossia solo il *punto terminale*) di un percorso logico attraverso il quale il giudice rende *concretamente ragione* della scelta delle premesse su cui ha ritenuto di fondare i propri ragionamenti e dei modi attraverso i quali da tali premesse ha provveduto ad argomentare criticamente le conseguenze che ha inteso trarne.

In breve, l'atto del giudicare (quantomeno sul piano giudiziario) potrà ragionevolmente sintetizzarsi in questo elementare assolvimento del dovere del giudicante di dare sinteticamente ragione di sé, ossia delle scelte che ha ritenuto di privilegiare intorno ai fatti naturali e a quelli normativi dei quali ha inteso informare i percorsi logico-argomentativi posti a fondamento del giudizio.

Là dove il giudice intenda avvalersi dell'ausilio dell'intelligenza artificiale che si esprime nelle forme della giustizia predittiva sarà pertanto indispensabile (proprio allo scopo di assolvere a quel dovere di dare compiuta ragione di sé e delle proprie scelte) entrare, con il proprio interlocutore automatizzato, in una relazione di comprensione critica fondata sull'acquisizione di una piena coscienza della specifica articolazione dei percorsi computazionali che presiedono all'*output* del sistema; delle articolate potenzialità di questo, ma insieme anche dei limiti o, in generale, della qualità specifica dell'interlocuzione che intrattiene con esso.[\[20\]](#)

11. La coscienza e le radici del male.

Il lungo discorso sin qui condotto si è posto l'obiettivo di individuare, sia pure in termini elementari e non esaurienti, il profilo dei rischi che si annidano nell'incontro tra l'esperienza dell'intelligenza artificiale e l'attività del giudice.

Si tratta degli stessi rischi che da sempre insidiano il corso dell'esperienza umana e che, con riguardo al tema che viene coinvolgendo la considerazione dell'intelligenza artificiale, si traducono nei pericoli del 'funzionalismo tecnocratico' o della 'burocratizzazione'.

Nell'accennare a tale fenomeno, sarà ancora utile riferirsi alle pagine preziose dedicate da Hannah Arendt – oltre '*Le origini del totalitarismo*'[\[21\]](#) e la '*Banalità del male*'[\[22\]](#) – alla figura tradizionale, se non proprio germinale del pensiero occidentale, di Socrate.[\[23\]](#)

Nell'affrontare la densità di quelle brevi pagine, Adriana Cavarero[\[24\]](#) sottolinea come la figura di Socrate abbia finito con l'attraversare l'intera opera di Hannah Arendt, apparendole come il modello di un pensare critico alla base della capacità umana di giudicare, l'invenzione della coscienza come tribunale interno dell'io che, facendosi 'due-in-uno', si interroga e dà conto di sé a sé stesso.

Quando Arendt ripensa il nesso tra condizione umana e politica, ne ricostruisce le vicende concettuali fin dalle origini greche, a partire dal dato reale della de-umanizzazione prodotta dalla macchina infernale dei campi di sterminio.

Nella terrificante originalità del sistema totalitario, di cui Auschwitz è l'atroce emblema, è in gioco la cancellazione dell'unicità della persona e perciò la distruzione della spontaneità come componente costitutiva del comportamento umano, al fine di produrre un esemplare completamente indistinguibile e indefinibile della specie *homo sapiens*.

Questo post-umano fabbricato dal campo di sterminio non solo coincide con l'evento del collasso delle definizioni di 'uomo' e di 'umanità' elaborate dalla tradizione metafisica, ma si accompagna al sospetto che proprio questa tradizione, nata con Platone, sia in qualche modo implicata nel verificarsi di tale collasso.

Arendt pone in evidenza come il male radicale abbia a che fare con l'eliminazione dell'imprevedibilità della n