

INTELLIGENZA ARTIFICIALE: SFIDE PER L'ETICA E PROSPETTIVE ANTROPOLOGICHE

- L. FLORIDI, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Raffaello Cortina Editore, Milano 2022;

- N. BOSTROM, *Superintelligenza. Tendenze, pericoli, strategie*, Bollati Boringhieri, Torino 2018;

F. FOSSA – V. SCHIAFFONATI – G. TAMBURRINI, *Automati e persone: introduzione all'etica dell'intelligenza artificiale e della robotica*, Carocci editore, Roma 2021.

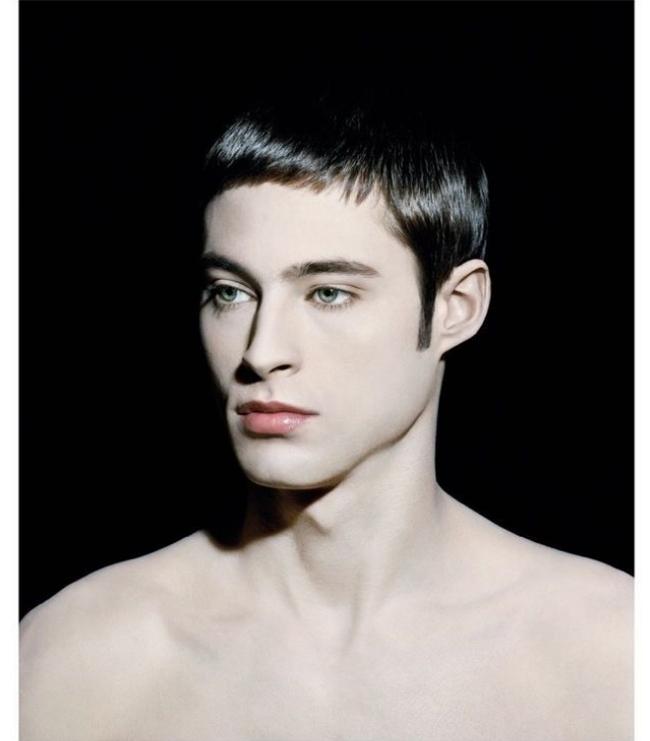
«Era anelito religioso corroborato dalla speranza, era il sacro Graal della scienza. Le nostre ambizioni in corsa su un ottovolante: un mito della creazione trasformato in realtà, un atto di mostruoso narcisismo. Non appena divenne fattibile non ci restò altra scelta che provarci, e al diavolo le conseguenze. A dirla nel più nobile dei modi, cercavamo di sottrarci alla nostra condizione mortale, di affrontare se non di sostituire la divinità con un io esemplare. In parole più povere, intendevamo ideare una versione migliore e più moderna di noi stessi e gioire del trionfo dell'estro, del brivido della nostra maestria.

Nell'autunno del ventesimo secolo finalmente accadde, il primo passo verso la conquista di un sogno antico, l'inizio di un lungo insegnamento in base al quale ci saremmo detti che, per quanto complicati fossimo, per quanto imprecisa e difficile risultasse la descrizione dei nostri gesti e comportamenti, anche i più banali, potevamo essere imitati e perfezionati. E io ero presente in quell'alba gelida: un giovane e smanioso pioniere dell'adozione.

Certo, le creature artificiali erano un cliché molto prima del loro arrivo, il che spiega come mai, al momento buono, per alcuni furono una delusione. Più agile della storia, come del progresso»

(Ian McEwan, *Macchine come me*, Einaudi)

IAN McEWAN
MACCHINE COME ME



EINAUDI

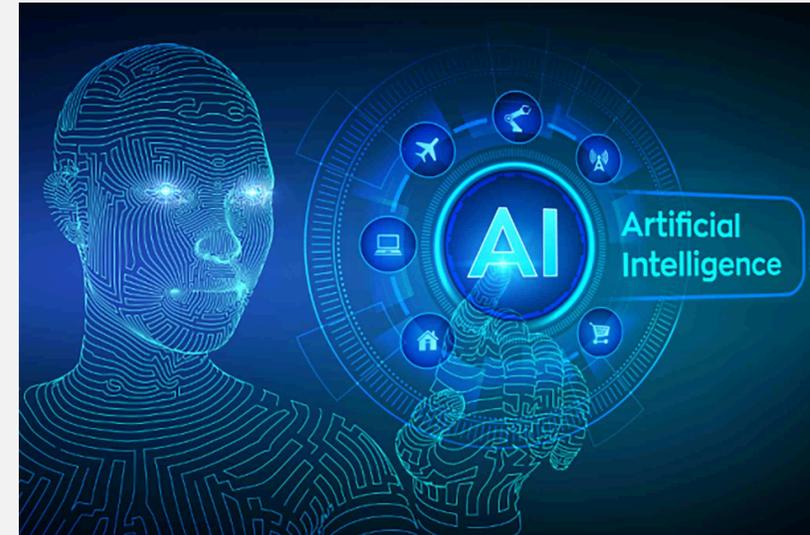
L'Intelligenza Artificiale sta sempre di più assumendo lo stato di questione centrale dei nostri tempi.

Lo è dal lato dei suoi

- **sviluppi scientifici,**
- delle **applicazioni tecnologiche,**
- delle **valutazioni** e delle **azioni** che nei suoi confronti vengono intraprese sia dalle istituzioni politiche sia dalle agenzie morali.

L'espressione **“Intelligenza Artificiale”** non ha una definizione sufficientemente rigorosa e **comprende cose molto diverse tra loro**.

- la ricerca fondamentale nell'ambito della computazione,
- i tentativi di simulazione/emulazione del pensiero umano,
- le applicazioni delle tecnologie derivanti dai nuovi strumenti di computazione e dalla razionalizzazione dei processi produttivi in tutti gli ambiti, da quelli dei servizi a quelli della manifattura, a quelli dell'agricoltura.
- le applicazioni delle tecnologie di computazione al controllo e all'organizzazione del comportamento umano, a livello sia individuale sia collettivo.



Si può definire l'«**Intelligenza artificiale**» come l'«insieme di studi e tecniche che tendono alla realizzazione di macchine, specialmente calcolatori elettronici, in grado di risolvere problemi e di riprodurre attività proprie dell'intelligenza umana» (Tullio De Mauro, *Grande dizionario italiano dell'uso*, Torino 2000).

La locuzione “intelligenza artificiale” è un evidente **ossimoro**, dal momento che attribuisce all'artificiale qualcosa che è essenzialmente “naturale”, in quanto l'intelligenza (riferimento alla ragione) è la prerogativa più propria della natura umana.

È obiettivamente difficile dare definizioni più precise,

- sia perché si tratta di una **materia in forte evoluzione**, sicché una definizione che delimitasse il territorio di sua competenza rischierebbe di escludere *a priori* sviluppi futuri che naturalmente le potrebbero appartenere,
- sia perché essa è **contemporaneamente una scienza e una tecnica**, ed è una **disciplina di frontiera**, una specie di affascinante “**punto multiplo**” in cui s'incontrano diversi domini del sapere: logica, informatica, psicologia, neuroscienze, filosofia.



IA è un'espressione utile per fare riferimento a una famiglia di scienze, metodi, paradigmi, tecnologie, prodotti e servizi.

- Un rapporto datato 2007 (Legg, Hutter, 2007) elencava già 53 definizioni di “intelligenza”. Il numero è in crescita (Russell, Norvig, 2016).

- *Wikipedia* risolve il problema optando per una tautologia:

«L'intelligenza artificiale (IA) è l'intelligenza mostrata dalle macchine, in contrasto con l'intelligenza naturale mostrata dagli esseri umani»

(“*Artificial Intelligence*”, 17 gennaio 2020).

Ciò dimostra che “probabilmente IA non è un termine scientifico ma un'espressione generica.

È una scorciatoia, usata per **riferirsi approssimativamente a diverse discipline, servizi, prodotti tecnoscientifici talora solo genericamente correlati.**



Ma c'è qualcosa che queste discipline, ambiti, paradigmi, metodi, tecniche o tecnologie riconducibili nell'alveo dell'IA **hanno in comune?** Sì.



L'interpretazione dell'IA **come nuova forma dell'agire** che non deve essere intelligente per avere successo si basa sulla tradizione ingegneristica: **può avere successo perché abbiamo trasformato il mondo (avvolgendolo) in un ambiente sempre più adattato al funzionamento dell'IA.**

Un po' di storia...

● L'idea di **delegare a congegni meccanici talune operazioni tipiche della mente** è molto antica. Basti pensare alle operazioni aritmetiche svolte con l'abaco, probabilmente inventato attorno al 5000 a.C. dai Cinesi.

● In epoca moderna, le prime macchine da calcolo sono dovute a **Pascal** (1623-1662) che nella metà del '600 costruì una addizionatrice meccanica (la *pascaline*), e a **Leibniz** (1646-1716) che, alla fine del secolo, la perfezionò per consentire le moltiplicazioni e le divisioni.

● Cresce l'interesse a trasferire alle macchine non solo il lavoro materiale — quello che comporta dispendio di energie fisiche — ma anche la fatica intellettuale richiesta.

● Gli sviluppi che più specificamente interessano la nascita dell'IA avvengono attorno alla metà del Novecento. Si devono ad **Alan Turing** (1912-1954) che nel 1936 propose un **modello ideale di calcolatore automatico "universale"**.

● Il primo lavoro che si usa ascrivere all'intelligenza artificiale risale al 1943, quando **Warren McCulloch** e **Walter Pitt** progettano una rete neurale artificiale.

● Nel 1956 **John McCarthy** riunì a Dartmouth i principali studiosi del tempo (fra i quali Marvin Minsky, Allen Newell, Claude Shannon e Herbert Simon) in un seminario in cui **propose il nome di «intelligenza artificiale»**.

● Il 1958 **McCarthy** produsse il *Lisp*, un linguaggio di programmazione ad alto livello dedicato specialmente all'IA (poi seguito, nel 1973, dal *Prolog*), e iniziò a sviluppare programmi generali per la soluzione di problemi.

● **S'incominciarono a studiare quelli che oggi sono chiamati «algoritmi genetici»**, ossia programmi capaci di **modificarsi automaticamente in modo da migliorare le proprie prestazioni**.

● Una grave difficoltà è l'«esplosione combinatoria», ossia l'aumento esplosivo del tempo di calcolo quando aumenta il numero di variabili del problema.

Stuart Russell e Peter Norvig propongono due distinzioni fondamentali.

1. La **prima** è fra macchine che “pensano” e macchine che si limitano a “operare” in maniera, in qualche misura, simile a quella degli esseri umani.
2. La **seconda** riguarda il termine di confronto per valutare le loro prestazioni; termine che può essere l’uomo reale o una sua idealizzata razionalità.

Le applicazioni tecniche riguardano principalmente l’operare razionale, mentre il dibattito filosofico insiste sulla possibilità di prestazioni “umane” e soprattutto di un “pensiero umano”.

Un’altra distinzione fondamentale, sulla quale il dibattito filosofico è acceso, è fra le cosiddette «**IA debole**» e «**IA forte**»:

1. Chi sostiene l’IA debole si accontenta di considerare macchine che agiscono “come se” fossero intelligenti;
2. Chi parla di IA forte, invece, asserisce la possibilità di macchine simili all’uomo fino a possedere un’autocoscienza.



La discussione riguarda due grandi temi:

1. **che cosa può fare l'IA,**
2. **che cosa è lecito fare con l'IA.**

- **L'IA debole** vuol costruire macchine che si comportino “come se” fossero intelligenti: ossia macchine capaci di risolvere “tutti” i problemi che l'intelligenza umana sa risolvere.

- **L'IA forte** vuole di più: afferma che la macchina che agisce in modo intelligente deve avere una “intelligenza cosciente”, una mente cosciente indistinguibile dalla mente umana.



L'IA debole si occupa della concreta costruzione o costruibilità di macchine “pensanti”,

L'IA forte vuol dare risposta al problema astratto di che cosa sia il loro “pensare”.



Si può credere all'IA forte ed essere scettici sull'IA debole (Russell e Norvig). Si può pensare che le macchine intelligenti, se fossero costruite, avrebbero un'intelligenza cosciente, ma ritenere che esse non possano essere costruite.

Assertazioni del tipo: «Il cervello è una macchina e perciò, in linea di principio, si può costruire una macchina che faccia tutto ciò che fa il cervello», se applicate alle proprietà della mente, assumono un chiaro sapore riduzionistico poiché assumono implicitamente che “mente” coincida con “cervello”.

Questa impostazione è chiaramente materialista ed estremamente discutibile.

Intelligenza artificiale forte.

L'intelligenza artificiale forte, il cui maggior esponente può essere considerato la professoressa Margaret Boden, ritiene che un computer correttamente programmato possa essere veramente dotato di una intelligenza pura, non distinguibile da quella umana. L'idea alla base di questa teoria risale al filosofo empirista inglese Thomas Hobbes, il quale aveva sostenuto che ragionare non era nient'altro che calcolare: la mente umana dunque non era altro che il prodotto di un complesso insieme di calcoli eseguiti dal cervello.

Intelligenza artificiale debole.

L'intelligenza artificiale debole del filosofo John Rogers Searle, invece, sostiene che un computer non potrà mai eguagliare la mente umana, ma potrà arrivare solo a simulare alcuni processi cognitivi umani senza mai riuscire a riprodurli nella loro totale complessità. L'obiezione si fonda sul fatto che i computer non possiedono una coscienza per cui si comportano come se fossero intelligenti, ma in realtà seguono procedure meccaniche, senza comprendere le loro azioni.

Sul piano della **realizzazione tecnica**, si può distinguere tra:

1. Un'impostazione “funzionale” o “comportamentistica”, per la quale non importa quale sia la struttura dotata dell'elaboratore sede dell'“intelligenza”,

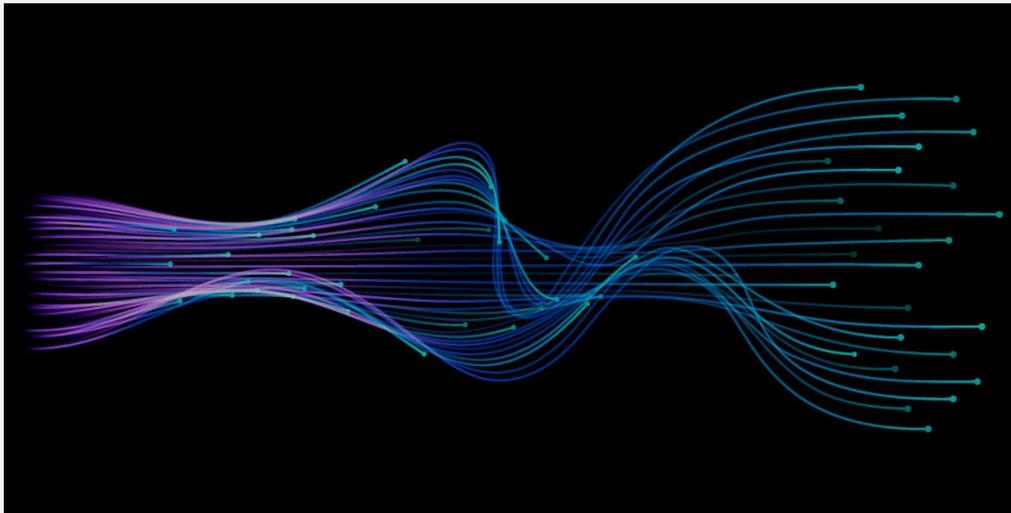


è stata chiamata anche “**emulazionistica**”: si è dell'idea che l'essenza del funzionamento del cervello non risieda nella sua struttura ma nelle sue prestazioni, e che queste possano essere ottenute, e con esiti forse addirittura migliori, anche da strutture completamente diverse.

2. Un'impostazione “strutturale” o “costruttivistica” o “connessionistica”, che vuole ottenere le stesse prestazioni del cervello umano riproducendo, in qualche modo, la sua struttura.



È stata chiamata anche “**simulazionistica**”: i sostenitori di quest'ultima ritengono che solo riproducendo il più fedelmente possibile il cervello si possono ottenere prestazioni ad esso paragonabili.



Il punto di vista privilegiato sembra essere quello “**emulazionistico**” (non “simulazionistico”): **l’idea che sta alla base è quella di costruire macchine che non necessariamente “simulino” riproducendo il comportamento del cervello umano, ma siano più semplicemente in grado di “emularlo”, selettivamente, nel risultato finale di certe operazioni.**

La macchina può essere concepita semplicemente come un trasformatore **sintattico** (codifica e combina i simboli) senza alcuna conoscenza **semantica** del problema (non intercetta il significato che sta dietro, espresso e contenuto nei simboli).

L’elaboratore, infatti, può trattare soltanto i legami “sintattici” e non i contenuti “semantici”, ossia il significato delle variabili sulle quali sta operando.

Il sistema esamina un largo numero di possibilità e **costruisce dinamicamente una soluzione.**



- Nella **memoria temporanea** (conoscenza dichiarativa su un particolare problema da risolvere) si ha una rappresentazione costituita da **(1) fatti veri introdotti all’inizio della consultazione o (2) dimostrati veri dal sistema nel corso della sessione di lavoro.**

- Nella **memoria a lungo termine** vengono invece mantenute regole che forniscono un insieme di **raccomandazioni, consigli, direttive strategiche** atti a costruire il patrimonio di conoscenza disponibile per risolvere il problema.

Il **ragionamento induttivo** procede da asserzioni singolari riguardanti **particolari** fatti o fenomeni (“esempi”) ad asserzioni **universali** esprimibili mediante ipotesi o teorie che spieghino i fatti dati e siano in grado di predirne di nuovi.

Mentre però **il ragionamento deduttivo** preserva la “**verità**” (nel senso di correttezza logica), quello **induttivo** non garantisce ciò, e quindi tali sistemi possono tendere ad un’eccessiva generalizzazione e a produrre **errori**.

● Se è dunque vero che

...i calcolatori possono memorizzare con facilità grandi quantità di informazioni, operano in nanosecondi e possono svolgere enormi moli di calcoli aritmetici senza errore, mentre gli uomini non sono in grado di avvicinarsi a tali prestazioni,

● è soprattutto vero che

gli uomini, normalmente, svolgono “semplici” compiti come camminare, parlare, interpretare una scena visiva o comprendere una frase, ragionare su eventi di senso comune, trattare situazioni incerte, in modo molto più brillante ed efficiente dei più raffinati e costosi programmi di IA risultanti dall’approccio simbolico e funzionale.



Gli esseri umani, intelligenti, combinano continuamente ragionamento induttivo e inferenza deduttiva,
e per ciò stesso possono raggiungere e «preservare la verità».

Al **motore inferenziale** (in informatica = algoritmo che simula le modalità con cui la mente umana trae delle conclusioni logiche attraverso il ragionamento) è delegata la scelta di quale conoscenza sia opportuno utilizzare, istante dopo istante, nel processo risolutivo.

Un **sistema esperto**, ossia un «**sistema basato sulla conoscenza**», è uno strumento in grado di risolvere problemi in un dominio limitato, ma con prestazioni simili a quelle di un esperto umano del dominio stesso.



Il compito fondamentale di un sistema esperto è quello di **coadiuvare l'attività di utenze professionali**, laddove è usualmente richiesta la consulenza di uno specialista umano dotato di competenza e capacità di giudizio.

Le ricerche di IA hanno fatto luce sui **problemi realizzativi di tali strumenti**, affermando la necessità di restringere, per quanto possibile, il campo di applicazione.

Rispetto ad un esperto umano, questi applicativi si rivelano più **limitati e superficiali**, visto che non dispongono di quella completezza che costituisce la **conoscenza culturale** della persona competente.

Inoltre

Non è possibile sperare che un sistema esperto possa giungere a conclusioni in maniera intuitiva o saltando alcuni passaggi logici, affidandosi al “**buon senso**” o al meccanismo della **analogia**, com'è invece prerogativa dell'uomo.

A questa serie di questioni si può rispondere ad esempio grazie agli algoritmi genetici.

Ma se è vero che le molte delle applicazioni descritte mostrano che la macchina può allargare di molto e modificare la gamma delle sue possibilità,

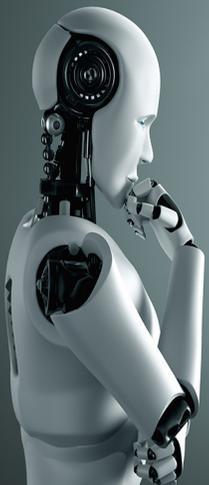
è anche vero che anche queste modificazioni sono, in qualche modo, previste e perciò potenzialmente incluse nella programmazione originaria.

Questo dato mette in luce l'essenziale **dipendenza della macchina dall'uomo e dalla sua progettualità:**

- a) **come costruire una “conoscenza di fondo” a partire dalla quale impostare l'apprendimento;**
- b) **come organizzare il processo di apprendimento in modo da ottimizzare il suo rendimento;**
- c) **come realizzare i procedimenti induttivi che dall'esperienza generano conoscenza;**
- d) **come controllare l'acquisizione dei dati sensoriali.**

Di fronte alle domande fondamentali dal punto di vista etico sull'Intelligenza Artificiale, si mostra ancora una volta tutta la profondità del **principio metodologico tradizionale**: *“Numquam nega, raro adfirma, distingue frequenter”* (NON NEGARE MAI, AFFERMA RARAMENTE, DISTINGUI FREQUENTEMENTE)

Più si va avanti nel conoscere, più le categorie si moltiplicano; e più si va avanti ancora, più queste ultime si raffinano. Il sapere più evoluto diffida delle negazioni, non si accontenta delle affermazioni e si alimenta di distinzioni frequenti.



È noto che nel maggio 1997, a New York, una macchina (*Deep Blue*) ha battuto in un *match* di sei partite il campione del mondo di scacchi **Garry Kasparov**.

Questo episodio ha acceso i riflettori su un'altra questione:

l'utilizzo della **logica** e l'**automazione delle dimostrazioni matematiche**.

Si tratta di un altro campo applicativo in cui l'IA ha raggiunto notevoli risultati.

La nozione di «**programmazione logica**» nasce agli inizi degli anni '70.

I **linguaggi di programmazione** più diffusi, dal *Fortran* al *Pascal* al *C*, sono basati sul paradigma imperativo, secondo il quale un programma consiste in una sequenza di comandi che specificano in modo estremamente dettagliato le operazioni che dovranno essere eseguite dall'elaboratore per risolvere il problema dato.

Le macchine non potranno dirsi intelligenti fino a quando non saranno in grado non solo di accrescere (quantitativamente) le proprie conoscenze ma anche di migliorare (qualitativamente) le proprie abilità.

L'apprendimento consiste in cambiamenti del sistema che siano adattativi, nel senso che mettono in grado il sistema di svolgere la volta successiva lo stesso compito in modo più efficiente.



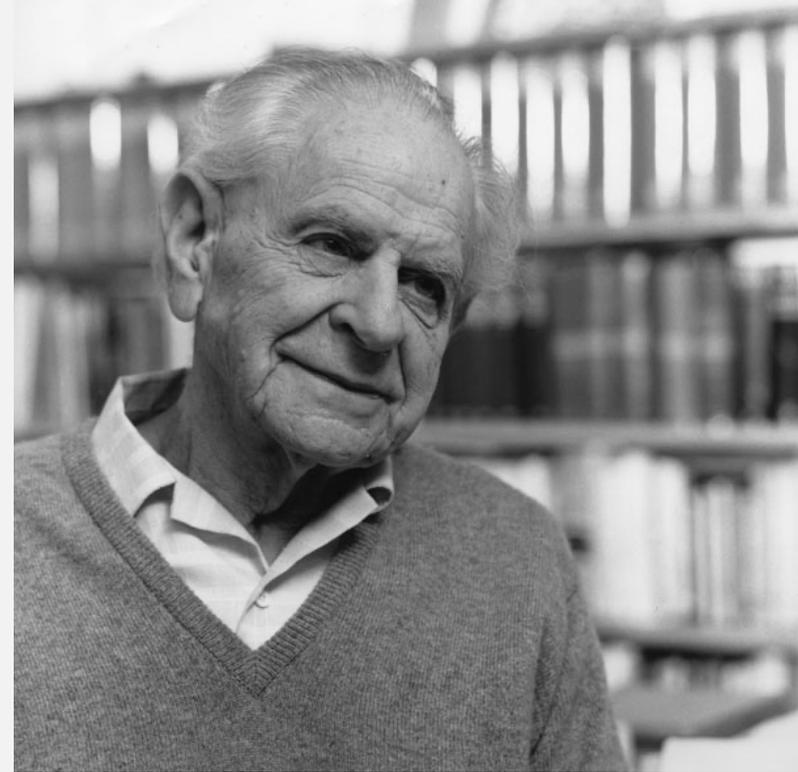
Karl Popper contestò fortemente che avesse un contenuto epistemologicamente valido dare alla ricerca sulla computazione e sulla simulazione del pensiero umano la qualificazione di “intelligenza artificiale”.

Gli algoritmi di computazione, come processi e come prodotti, possono infatti rappresentare soltanto una parte, e non quella più significativa e importante, dell’intelligenza umana.

La disponibilità oggi di **strumenti di calcolo sempre più potenti** non fa venir meno questa fondamentale verità.

Le **questioni etiche** che si pongono nei confronti dell’Intelligenza Artificiale difficilmente possono riguardare i progressi che vengono fatti sia nell’ambito della computazione sia in quello della simulazione del pensiero umano, in quanto

si tratta di **progressi conoscitivi**, non diversi da quelli che avvengono negli ambiti della fisica o della chimica o della biologia.



Un apporto che sottopone a **scrutinio etico** l'Intelligenza Artificiale, sia nell'**ideazione** come anche nell'**applicazione** tecnica, riguarda

- sia la conformità delle procedure ai canoni etici fondamentali e alla deontologia,
- sia le potenziali conseguenze negative che possono derivare dall'utilizzo dei suoi risultati.

Non vi è infatti nulla di assoluto nell'attività umana, nulla che non debba essere ricondotto al rispetto dei principi etici fondamentali.

L'Intelligenza Artificiale

- Non nasce da una volontà negativa.
- Non nasce da una volontà di dominio.
- Nasce da una volontà conoscitiva.

Tuttavia la volontà conoscitiva può però condurre a conseguenze lontane dai principi dell'etica.



Come disse nel 1947 J. **Robert Oppenheimer**, tra gli inventori della bomba atomica: «**In un senso crudo che non potrebbe essere cancellato da nessuna accezione volgare o umoristica, i fisici hanno conosciuto il peccato**».

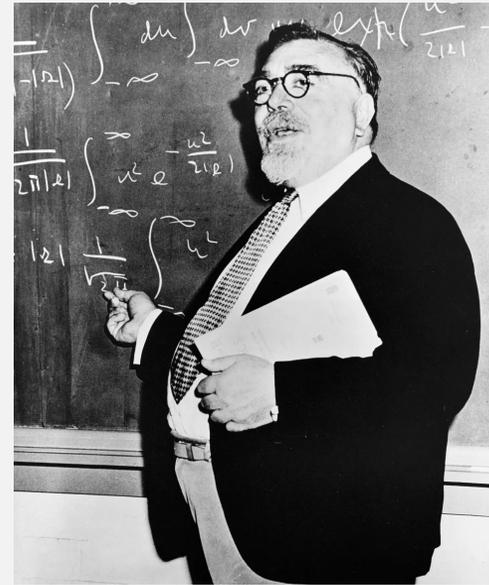
Già nel 1950 **Norbert Wiener**, uno dei padri della cibernetica e dell'Intelligenza Artificiale, aveva ammonito circa l'**utilizzo inumano degli esseri umani** ove si cedesse alle macchine il **potere di scelta e di controllo**.

Il **“peccato”** può non nascere da una volontà individuale. **Esso può essere il risultato oggettivo di un concatenarsi di azioni individuali** ognuna delle quali, singolarmente presa, è conforme alla morale, ma che combinate producono effetti moralmente inaccettabili.



Sappiamo che un'**irrazionalità collettiva** può derivare – e molto spesso deriva, come nel caso dell'inquinamento – dalla somma di **innumerevoli azioni umane razionali**.

Soltanto l'azione **di regole morali e giuridiche** può evitare che questo si produca.



L'essere umano rimane il soggetto etico di riferimento.

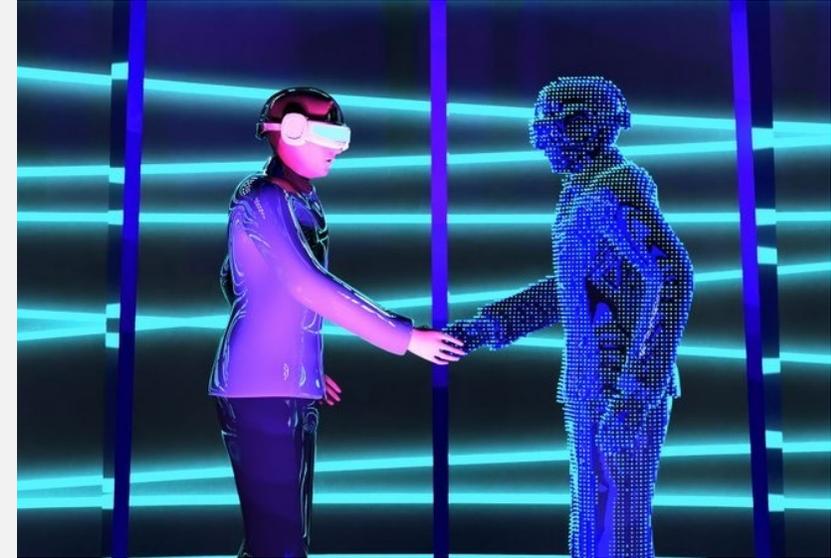
Se pensiamo che la definizione di Intelligenza Artificiale sia questa: «Una macchina intelligente in grado di prendere decisioni al posto dell'uomo», **sbagliamo**. Anche se sempre più sofisticate, infatti, le macchine sono pur sempre dei **meri esecutori di istruzioni impartite dall'uomo sotto forma di numeri e formule, che poco hanno a che vedere con facoltà tipicamente umane come le emozioni, il coraggio, la responsabilità o l'immaginazione**.

Vi sono aspetti dell'**applicazione** dell'Intelligenza Artificiale che sono chiaramente **contrari ai principi etici**:

- **controllare i comportamenti e, ancor più, le idee individuali sino a ridurre di fatto a finzione la libera scelta delle persone;**
- **trasformare la società in un mero aggregato di monadi, restringendo fortemente le comunità e la socialità naturale, per poi paradossalmente riproporle ma in mondi virtuali come il Metaverso (realtà virtuale condivisa tramite internet, dove si è rappresentati in tre dimensioni attraverso il proprio avatar);**
- **spingere la razionalizzazione dei processi produttivi sino al punto di togliere dignità alle lavoratrici ed ai lavoratori;**
- **imporre un modello algoritmico alle decisioni dei giudici e delle autorità di vigilanza.**



Opporsi a questo significa volere regole giuridiche vincolanti che impediscano la reificazione della persona umana e l'artificializzazione dell'uomo.



Il **diritto positivo** può dispiegare la sua azione secondo un modello che è conosciuto da sempre, riguardo sia i **diritti individuali** sia riguardo i **diritti sociali**.

Necessita l'elaborazione di un modello da adeguare alle nuove circostanze, che si ispiri eticamente alla **visione umano-centrica nel rapporto uomo-macchina**.

Una visione che riconosca alcune **dimensioni antropologiche insostituibili e non simulabili** quali

- l'autoconsapevolezza,
- la libertà,
- la relazionalità interpersonale,
- la capacità immaginativa e raffigurativa,
- il senso del sacro, del numinoso, del divino, del mistero.

Una visione che affermi l'esigenza di **mantenere un controllo umano significativo e che faccia propria l'esigenza di evitare eccessive deleghe tecnologiche, costruendo tecnologie dell'Intelligenza Artificiale affidabili, spiegabili, tracciabili, trasparenti e inclusive**.

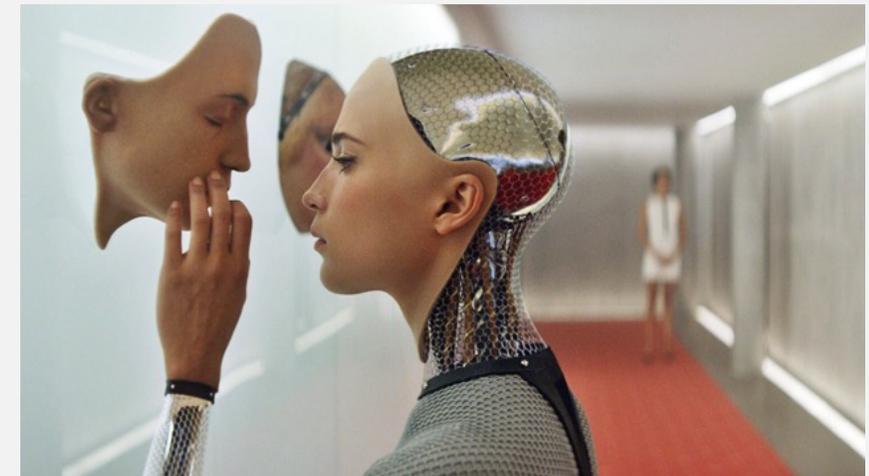
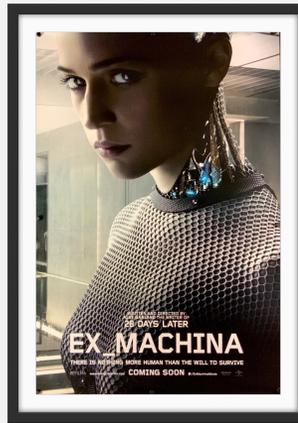
La sfida maggiore della regolazione delle applicazioni dell'Intelligenza Artificiale dal punto di vista dei valori etico-costituzionali è quella di **contrastare il modello di stampo utilitaristico** che esse stanno imprimendo a tutto il mondo della produzione economica.



Oltre che sui **rischi** di un **annullamento dell'umano**, o di una sua **radicale trasformazione** in altro da sé (sfide del transumanesimo, postumanesimo), le scienze sociali richiamano l'attenzione sugli strumenti di governo, controllo e orientamento dello sviluppo tecnologico, e sulla riflessione etica rispetto all'armonizzazione dei percorsi di ricerca tecnologica con i principi della dignità umana e della giustizia sociale.



Una questione importante che si pone è come far sì che le regole giuridiche non blocchino o limitino fortemente un giusto progresso tecnologico, ovvero un progresso che migliori le prospettive di vita dei cittadini.



Distingue frequenter

Le regole e gli obblighi giuridici relativi alle applicazioni sono distinti da quelli che devono riguardare la corretta direzione della ricerca sull'Intelligenza Artificiale, specialmente di quella parte di essa che mira a riprodurre attraverso gli artifici il pensiero umano e, secondo certe linee di ricerca, anche le emozioni e i sentimenti umani.



Non solo conoscere, ma riprodurre.

Qui la dimensione esclusivamente, o anche prevalentemente, giuridica, è largamente insufficiente. Lo è per la natura stessa dello strumento giuridico, che elabora norme volte a evitare danni certi o probabili che riguardino singole persone o l'intera società.

Ci si trova di fronte agli interessi delle generazioni future: difficilmente le norme giuridiche possono essere di guida o di contenimento di comportamenti le cui conseguenze negative non sono né certe né determinabili al presente e tantomeno al futuro, e che non consentono, quindi, quel giudizio di valore o disvalore proprio della norma giuridica.



Qui si dispiega lo spazio proprio e specifico dell'etica. Spetta alla coscienza dei ricercatori, come singole persone e come comunità, svolgere il ruolo di guida nel decidere su cosa fare ricerca, e come farla.

È affar di coscienza **valutare i rischi per le generazioni presenti e per quelle future** che possano derivare dallo spingere la frontiera delle conoscenze oltre un limite tale che le ricadute tecnologiche potrebbero entrare in contrasto con i valori fondamentali della persona umana.

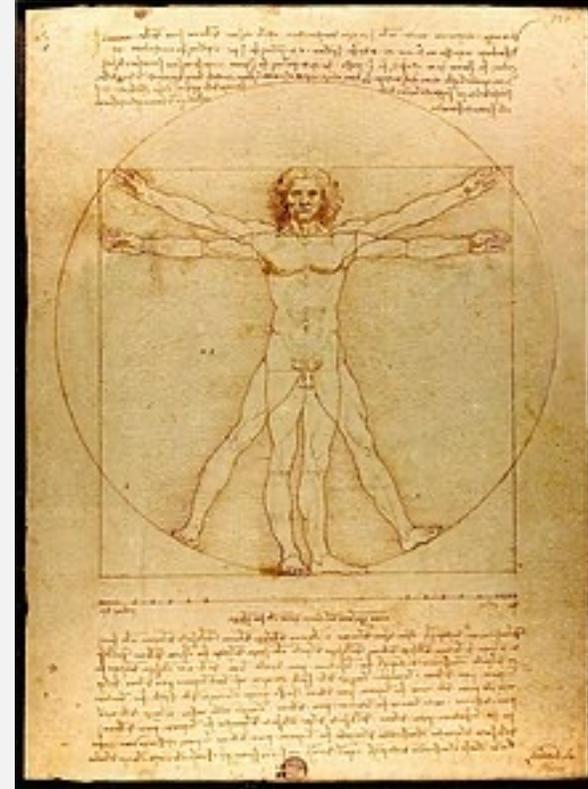
In questa attenzione costante la **coscienza dei ricercatori si nutre della riflessione comune** che deriva dalle concezioni che vengono elaborate e mantenute dalle **comunità morali** nelle quali (1) **si elaborano** e (2) **si rivedono continuamente le visioni antropologiche** che stanno alla base della determinazione dei fini propri della persona e della società.



Senza una **antropologia di spessore**, le regole morali sono affermazioni di un mero “dover essere” privo di un fondamento universale, e quindi soggetto al flusso del relativismo storico.

Questa determinazione deve unire gli sforzi tanto di chi condivide (1) una **morale derivante da una fede religiosa** e quindi da valori trascendenti, quanto di chi condivide (2) una **morale basata su valori immanenti** ma che non per questo destinati a dissolversi nel relativismo.

Il problema serio, ancora in attesa di soluzione, è dunque quello di **come giungere a una comunanza etica nella società del pluralismo**.



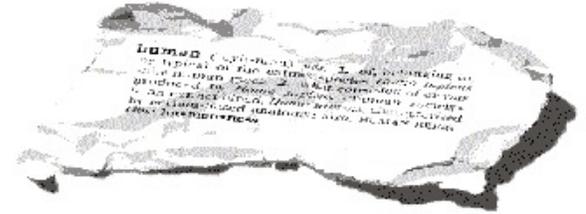
Nel 2000 **Bill Joy**, uno dei protagonisti dell'informatica della Silicon Valley, in un articolo sulla rivista *Wired*, che suscitò molte discussioni, affrontò il tema: **“Perché il futuro non ha bisogno di noi”**. Sostenne che genetica, nanotecnologie e robotica, sono in grado di costruire sistemi tecnologici che possono mettere a rischio il futuro dell'umanità più di quanto abbiano fatto, un secolo fa, le tecnologie nucleari, biologiche e chimiche nel settore degli armamenti.

Le attuali tecnologie sono più facili da realizzare, alla portata anche di singole persone, ma meno controllabili, perché particolarmente complesse.

Per cui:

nel settore dell'informatica mantengono importanza e attualità gli interrogativi etici e i principi di responsabilità e di precauzione.

W I R E D



**Why the Future
Doesn't Need Us**
By Bill Joy

Rimangono ineludibili le **domande sul rapporto tra conoscenza scientifica e applicazioni tecnologiche**. Esse riguardano, in particolare:

- il significato autentico del progresso,
- la relazione tra natura e cultura,
- l'esercizio della libertà e della responsabilità nelle applicazioni tecnico-scientifiche,
- il senso del limite,
- il rapporto fra razionalità e sfera emotiva nella determinazione del comportamento umano,
- il governo delle tecnologie e il loro autonomo sviluppo.

Come in parte si è detto...

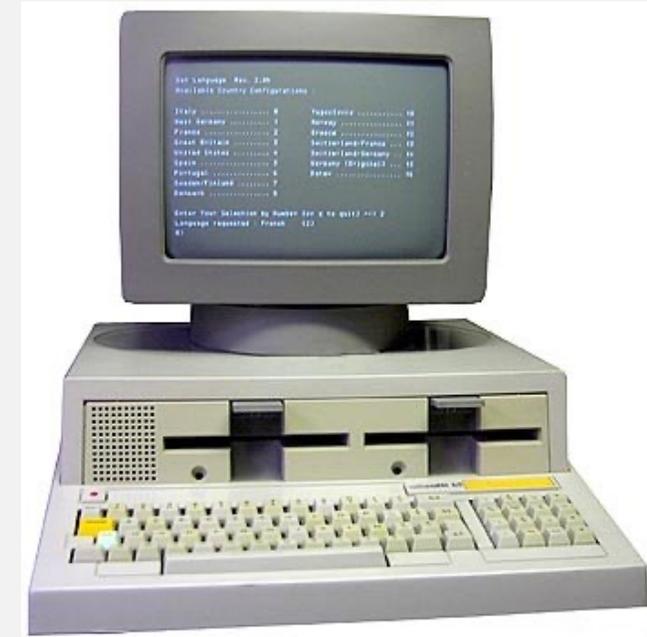
Tra la fine dell'Ottocento e la prima metà del Novecento gli studi di logica (Boole), di matematica (Turing, Church, Gödel), di elettronica delle valvole e dei transistor, hanno posto le basi per la costruzione di computer e di apparati digitali, permettendo di realizzare le “**macchine da calcolo**” immaginate da studiosi dei secoli precedenti (Pascal, Leibniz, Babbage).

Nel giro di tre decenni, tra il 1930 e il 1960, furono costruiti i **calcolatori elettromeccanici** (Zuse), quelli **elettronici** (Atanasoff e Berry) e i **computer** (von Neumann).

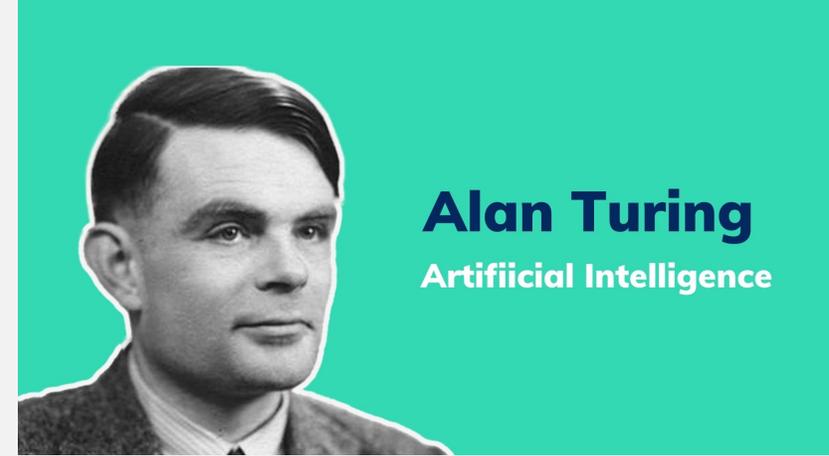
Negli anni '50 l'Italia guadagnò una posizione di rilievo, con la C.E.P. (Calcolatrice Elettronica Pisana) e con l'ELEA realizzato da Mario Tchou all'**Olivetti**.

Negli anni '60...ecco il microchip con **Enrico Faggin**...

Ma è soprattutto per effetto delle applicazioni delle tecnologie digitali che il progresso tecnologico ha conosciuto uno sviluppo rapido e pervasivo, fino a diventare oggi la cornice ineliminabile della nostra vita quotidiana.



L'universalità della macchina, ottenuta per mezzo di programmi codificati e memorizzati, e l'estensione delle sue funzioni dal calcolo alla gestione delle informazioni, per una certa similitudine con le attività intellettive dell'uomo, ha indotto a parlare di Intelligenza Artificiale (Turing, Minsky, McCarthy).



Turing presenta e discute anche un fantasioso elenco di operazioni che, secondo i suoi avversari, “la macchina non potrà mai fare”.

Alcuni punti dell'elenco, come «**apprendere dall'esperienza**», sono stati, almeno ad un certo livello, realizzati; altri, semplicemente, non riguardano la macchina in quanto soggetto dell'IA (es. «essere bello», «far innamorare qualcuno di sé», oppure «gustare delle fragole con panna»: si può immaginare un robot dotato di sensori del gusto e dell'olfatto, e di un programma che discrimini i gusti gradevoli).

Anche in questo caso la questione riguarderebbe soprattutto la robotica, ma con un problema in più: “gustare” implica già una facoltà tipica della mente umana e quindi chiama in causa il problema di fondo dell'IA forte.

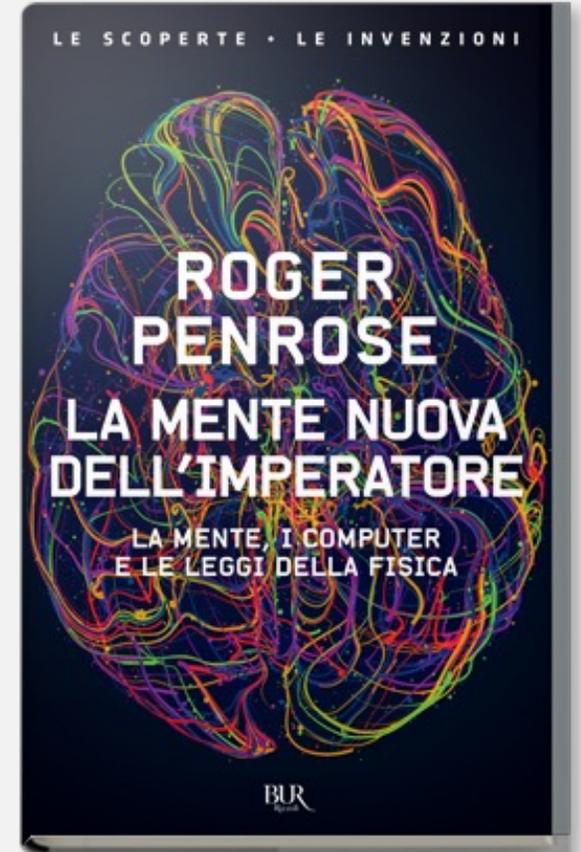
Altrettanto può dirsi dell'operazione più inquietante: «essere l'oggetto del proprio pensiero», che chiama in causa l'autocoscienza.

Sono questi sviluppi velocissimi che hanno imposto sempre più una riflessione etica...perché ci si è trovati di fronte a qualcosa di più rispetto a «macchine» più ponti ed efficienti...

1. Un problema legato a tutto ciò è quello della «**terminazione**»: l'esecuzione di un certo programma avrà un termine, o potrebbe proseguire, in linea teorica, all'infinito?

2. Inoltre, **a differenza delle macchine l'uomo sa «porsi fuori dal sistema» per cui l'esperienza comune mostra come egli sappia effettivamente superare il limite della pura logica e sia in grado di guardare i problemi logici “dall'esterno”.**

3. In più, come ha scritto **Roger Penrose**,
«L'intelligenza e la coscienza reali, non artificiali, hanno la capacità di connettere livelli logici, sintattici e semantici diversi, e di metterli continuamente in discussione, come nessun computer pensabile su basi fisiche (o fisico-chimiche, o biologiche, comunque artificiali) appare concepibilmente in grado di fare».



Si giunge ad un punto centrale, vale a dire al “**problema della coscienza**”.

John Searle lo propone mediante l’esperimento concettuale della “**protesi cerebrale**”, ipotizzando che, se con un intervento di raffinatissima microchirurgia, si riuscisse a sostituire uno per uno tutti i neuroni di un cervello con altrettanti microcircuiti elettronici che funzionino esattamente allo stesso modo dei neuroni, e **se di conseguenza** si riuscisse a riprodurre tutte le connessioni sinaptiche...che succederebbe della coscienza di quell’uomo?

- Qualsiasi risposta data a questo interrogativo ci fa muovere nello spazio di una “**prospettiva riduzionistica**”: l’identificazione della mente con il suo “supporto” materiale, il cervello, visto come una “macchina” pienamente riproducibile con dispositivi artificiali.

- Da questo punto di vista se si dà differenza fra mente e macchina, essa è da attribuire o a una **temporanea** insufficienza della macchina, a cui porre rimedio in futuro, o a un limite che la mente non sa ancora d’avere.



Più in generale, s'intende qui per **riduzionismo** l'idea che la mente umana possa essere simulata, almeno in via di principio, da sistemi artificiali capaci di riprodurre le prestazioni in maniera così perfetta da rendere indistinguibile l'una dagli altri.

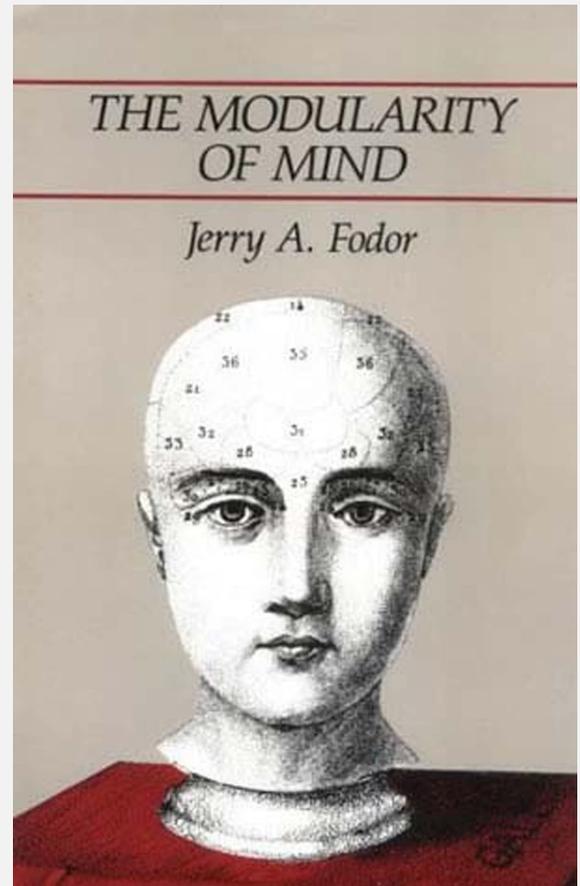
Jerry A. Fodor (filosofo, esperto di scienze cognitive) preferisce chiamare questa posizione «materialismo», e la contrappone al «dualismo» cartesiano che è rifiutato per la sua incapacità di rendere conto adeguatamente della causazione mentale.

Infatti, se la mente è qualcosa di non fisico, non ha di conseguenza una posizione nello spazio fisico.

- Come è possibile, allora, che una causa mentale dia luogo a un effetto comportamentale che invece ha una posizione nello spazio?

- Come può il non fisico dar luogo al fisico senza violare le leggi di conservazione della massa, dell'energia e della quantità di moto?

Per Fodor le simulazioni al calcolatore dei processi cerebrali forniscono modelli degli aspetti formali di questi processi, ma la simulazione non va confusa con la riproduzione.



L'IA è soltanto un «modello di simulazione/emulazione» dell'intelligenza naturale (e, per ora, soltanto di alcuni suoi aspetti), utilissimo ai fini pratici come tutti i modelli di simulazione/emulazione, ma nulla di più; e, se pur dovesse dar segno di qualcosa di simile alla coscienza o all'intenzionalità, dovremmo dire che si tratta soltanto di una simulazione/emulazione della coscienza e dell'intenzionalità.

- Se valesse il paradigma riduzionistico per cui la macchina sarebbe dotata anche di intenzionalità, allora la delega sarebbe profondamente inquietante.

- Ma se invece, più ragionevolmente, si ricorda che la macchina è programmata dall'uomo e dipende da lui anche quando è stata dotata di algoritmi "genetici" che sviluppano la programmazione in maniera non preordinata, allora il problema si riduce a stabilire in qual misura affidarsi alla "protesi intellettuale" per progettare una terapia, o un qualche intervento di rilevante peso economico o sociale.





Le **tecnologie digitali** hanno prodotto un vero e proprio “**mutamento antropologico**”, in almeno due settori fondamentali:

- **l'ingegneria genetica**: dopo le acquisizioni di microbiologia, biochimica, biologia molecolare, il contributo dell'informatica ha permesso di sequenziare il DNA e ha aperto la strada ad applicazioni in grado di manipolare i geni, con la prospettiva di riprodurre artificialmente la vita umana (la sfera riproduttiva viene separata da quella sessuale e si rendono equivalenti la vita umana “naturale” e quella “artificiale”).

- la **robotica**: la realizzazione di robot “intelligenti” tende a far scomparire, almeno nella mentalità comune, la differenza tra persona umana ed essere artificiale. In questa direzione spingono anche iniziative come quelle che vorrebbero definire la “personalità giuridica” del robot per giungere a stabilirne anche una qualche forma di “responsabilità civile”.



Accanto a potenzialità immense, che hanno portato a definire il nostro tempo “**Rinascimento digitale**” (De Kerchove), molti autori additano **i rischi di delegare alle macchine attività tipiche dell'uomo**, rendendolo meno capace di esercitare le caratteristiche peculiari della propria umanità: pensare, conoscere, decidere, scegliere, agire con responsabilità, entrare in relazione con gli altri.



La disponibilità di **tecnologie potenti e complesse produce alcuni effetti rilevanti:**

a) Una concezione utilitarista e de-responsabilizzante, che si esprime con l'affermazione: “Se è fattibile, perché non dovrebbe essere fatto?”.

b) L'idea secondo cui le macchine e gli algoritmi possano compiere le attività umane meglio di quanto l'uomo stesso riesca a fare.

c) Le gerarchie sociali sono determinate da chi produce e possiede le tecnologie: non si tratta di confini esclusivamente legati ad aspetti economici (potere che la tecnica attribuisce a chi può “creare la vita” a suo piacimento).



Sempre più frequentemente gli sviluppi tecnologici si orientano all'organizzazione di “**smart city**”, in cui la mobilità, il consumo di energia, il funzionamento degli apparati domestici, l'economia della condivisione e il lavoro d'ufficio a domicilio saranno agevolati dalla disponibilità di (1) **reti di comunicazione potenti**, da (2) **reti di sensori e sistemi di controllo diffusi ovunque** in maniera pervasiva, di (3) **mezzi di trasporto** che si muovono senza la necessità di un guidatore umano.



Per supportare gli apparati intelligenti, i robot intelligenti, le città intelligenti, servono **algoritmi sempre più sofisticati**, in grado di aggiungere autonomamente alle proprie funzioni di base anche nuove funzioni, poggianti su schemi predeterminati di “**apprendimento**” e a partire dall’analisi di informazioni, per loro natura incomplete, approssimative e incerte.

Gli algoritmi decisionali, apparentemente neutri e oggettivi, seguono la logica determinata dal progettista o le varianti previste, **separando** e **spezzettando** però **la responsabilità** tra molteplici attori:

- **committente,**
- **progettista,**
- **costruttore,**
- **utilizzatore.**

A questo si aggiunga...

Nella società degli algoritmi rimangono sullo sfondo due problemi in grado di condizionare l’evoluzione della società e la stessa qualità di vita delle persone:

1. **L’impatto sull’ambiente delle tecnologie digitali**, che avviene non tanto in maniera diretta e visibile, ma attraverso il consumo di energia, necessaria in quantità sempre crescenti.
2. **Il controllo dei flussi di informazioni personali che circolano sulla rete.** L’analisi di grandi quantità di dati e la loro correlazione consente di disegnare scenari e previsioni utili per prevenire o contrastare fenomeni dannosi, ma anche per fotografare con estrema accuratezza i comportamenti e gli orientamenti personali, mettendo le basi per condizionarli secondo gli interessi dei detentori delle informazioni, in maniera inavvertita dagli interessati.

Circa il **legame tra scienza e applicazioni tecnologiche**, vengono chiamati in causa temi di natura etica e filosofica sul rapporto tra conoscenza e suo utilizzo, ma anche aspetti economici e sociali riguardanti il sistema industriale e produttivo necessario per la realizzazione, e le aspettative del mercato che orientano la diffusione.



Il concetto di “neutralità” della scienza e della tecnologia sono ormai superati dall’impatto che esse hanno sulla cultura e sulla società, mentre la logica del profitto trova nelle tecniche di marketing metodi sofisticati per creare bisogni e alimentare una logica di sostituzione con sempre nuove versioni degli apparati tecnologici.

«La vita stessa per come la conosciamo oggi è diventata inconcepibile senza la presenza di pratiche, prodotti, servizi e tecnologie digitali. Stiamo parlando di un nuovo capitolo della storia umana. Naturalmente, molti altri capitoli l’hanno preceduto. Erano tutti ugualmente significativi. L’umanità ha sperimentato un mondo prima e dopo la ruota, la lavorazione del ferro, l’alfabeto, la stampa, il motore, l’elettricità, la televisione o il telefono. Ogni trasformazione è unica. Alcune di queste hanno cambiato in maniera irreversibile il modo in cui comprendiamo noi stessi, la nostra realtà e l’esperienza che ne facciamo, con implicazioni complesse e di lungo periodo».



Le generazioni future non sapranno mai com'era una realtà esclusivamente analogica, offline, predigitale. Siamo l'ultima generazione che l'avrà vissuta.

Proprio perché la rivoluzione digitale è appena iniziata, abbiamo la possibilità di plasmarla in modi positivi che possono fare progredire sia l'umanità sia il nostro pianeta.

Comprendere le trasformazioni tecnologiche in atto è cruciale se vogliamo guidare la rivoluzione digitale in una direzione che sia equa dal punto di vista sociale e sostenibile da quello ambientale.

● Si dà un'interpretazione dell'IA contemporanea in termini di riserva di capacità di agire, che è stata resa possibile da due fattori:

1) il divorzio tra la capacità di risolvere problemi e di portare a termine con successo compiti al fine di raggiungere un obiettivo, e la necessità di essere intelligenti nel farlo;

2) la progressiva trasformazione del nostro ambiente in un'infosfera* adattata all'IA, che rende tale divorzio non solo possibile ma anche efficace.



“Prima siamo noi a dare forma agli edifici; poi sono questi a dare forma a noi”
(Winston Churchill)

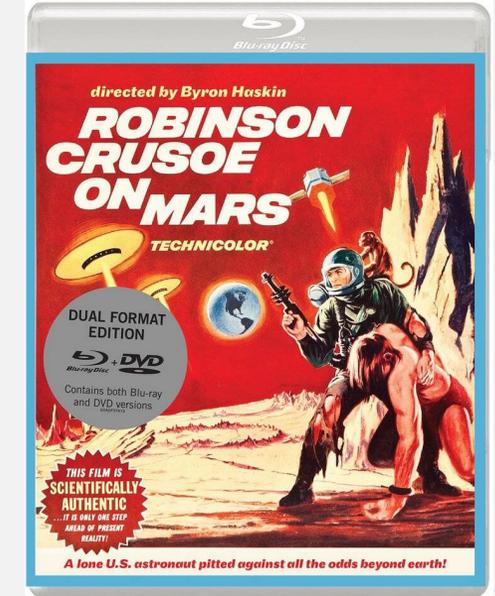
*insieme dei mezzi di comunicazione e delle informazioni da essi diffuse

Nel 1964, la Paramount Pictures distribuì **Robinson Crusoe su Marte**. Il film descriveva le avventure del comandante Christopher “Kit” Draper, un astronauta statunitense naufragato su Marte. Se lo guardiamo su YouTube anche solo per pochi minuti, ci rendiamo conto di quanto radicalmente sia cambiato il mondo in pochi decenni. In particolare, il computer che compare all’inizio del film sembra un motore di epoca vittoriana, con leve, ingranaggi e quadranti. Robinson Crusoe su Marte apparteneva a un’epoca diversa, tecnologicamente e culturalmente più vicina al secolo scorso che al nostro.

Descrive una realtà moderna, non contemporanea, basata sull’hardware e non sul software. Computer portatili, Internet, servizi web, touch screen, smartphone, orologi intelligenti, social media, shopping online, video e musica in streaming, automobili a guida autonoma, tosaerba robotizzati o assistenti virtuali non esistono ancora.

Il film mostra una tecnologia fatta di dadi, bulloni e meccanismi che seguono le leggi della fisica newtoniana. È una realtà del tutto analogica, basata sugli atomi piuttosto che sui byte.

L’intelligenza artificiale è più un progetto che una realtà.



Tutte le generazioni passate che vivevano in un mondo esclusivamente analogico hanno concepito e sperimentato posizione e presenza come due lati inseparabili della stessa situazione umana: l’essere situati nello spazio e nel tempo, qui e ora.

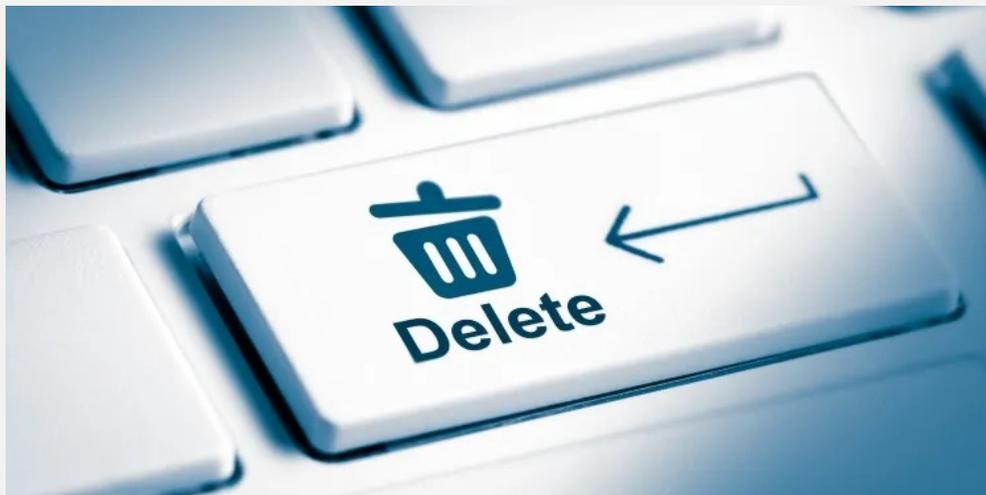
In un mondo digitale, invece, è «naturale» che uno possa trovarsi fisicamente in un posto (casa, bar, lavoro, ecc.) ed essere presente interattivamente in un altro (pagina Facebook, Instagram, ecc.).



Più memoria, più velocità e più ambienti e interazioni digitali hanno generato **più dati**.

Ne consegue che stiamo rapidamente passando **da una cultura della registrazione a una della cancellazione**: la questione non è più che cosa salvare ma che cosa eliminare per fare spazio ai nuovi dati.

La vecchia IA era per lo più simbolica e poteva essere interpretata come una branca della logica matematica, ma **la nuova IA è principalmente connessionista** e potrebbe essere interpretata come una branca della statistica.



Queste tendenze hanno modificato il modo in cui impariamo, giochiamo, lavoriamo, amiamo, odiamo, scegliamo, decidiamo, produciamo, vendiamo, ci compriamo, consumiamo, pubblicizziamo, ci divertiamo, ci preoccupiamo di qualcosa e ce ne prendiamo cura, socializziamo, comunichiamo, e così via.

Il problema dell'intelligenza artificiale è quello di far sì che una macchina agisca con modalità che sarebbero definite intelligenti se un essere umano si comportasse allo stesso modo.

Questo non ha nulla a che vedere con il pensiero ma esclusivamente con il comportamento: se un essere umano si comportasse in quel modo, quel comportamento sarebbe definito intelligente. Ciò però non significa che la macchina sia intelligente o che addirittura stia pensando.

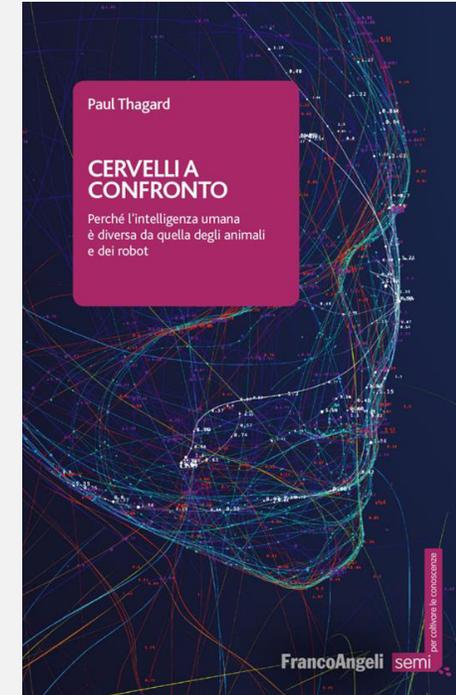
Se dunque ci poniamo la domanda: **“Possono le macchine pensare?”**, dovremmo iniziare **definendo** il significato dei termini **“macchina”** e **“pensare/pensiero”**.



Solo se si accetta l'idea che il vero **cervello**, che si trova negli animali – in particolare negli esseri umani – sia una sorta di **macchina**, ne consegue che il nostro computer digitale, opportunamente programmato, si comporterà come un cervello.

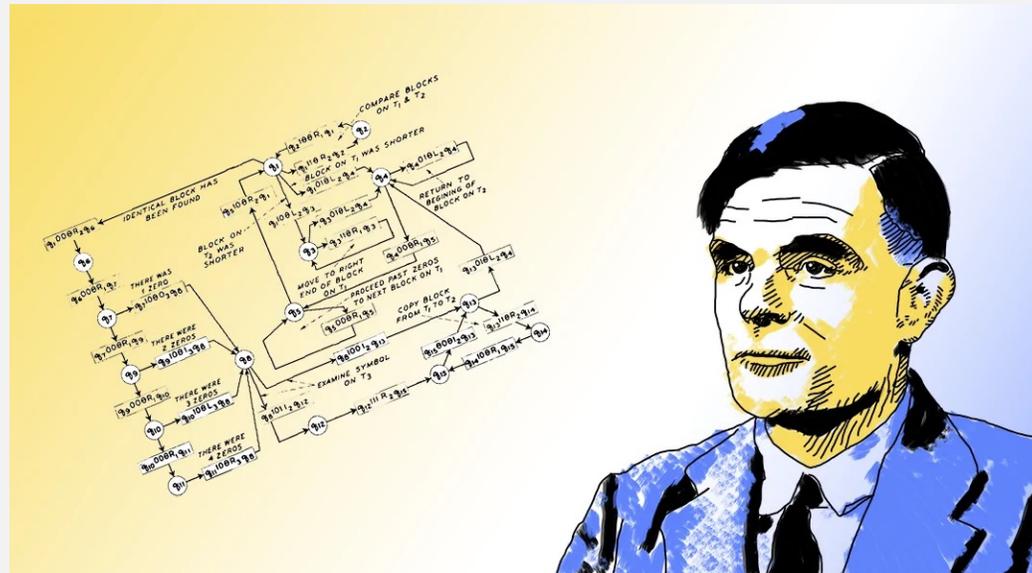


Quindi alla base di una risposta piuttosto che di un'altra sta l'idea – filosofica – che abbiamo dei termini implicati.



Abbiamo una vaga idea di ciò che possiamo considerare macchina. Anche una burocrazia può essere descritta non del tutto irragionevolmente come una macchina (F. **Kafka**). Un bollitore, un frigorifero, un treno, un computer, un tostapane, una lavastoviglie...sono tutte macchine, in un senso o nell'altro. Anche il nostro corpo è una macchina.

Così il nostro cuore. Perché non il cervello?



Posizione che potrebbe essere legittima, ma **un'operazione del genere implica un'opzione metafisica** (non strettamente scientifica) che chiama in causa categorie ontologiche, filosofiche, perfino religiose.

Altrimenti si potrebbe dire che la questione se un computer possa pensare non è poi così diversa della questione se un sottomarino possa nuotare (Edsger Wybe Dijkstra): **eseguire operazioni, avere certe caratteristiche, svolgere delle funzioni, ecc., non vuol dire essere intelligenti** (nel senso in cui si predica dell'essere umano) ecc.



L'IA di successo non riguarda la produzione ma la sostituzione dell'intelligenza umana.

Una lavastoviglie non pulisce i piatti come lo facciamo noi, ma alla fine del processo i suoi piatti puliti sono indistinguibili dai nostri, anzi possono essere anche più puliti (efficacia), utilizzando meno risorse (efficienza). Lo stesso vale per l'IA.

AlphaGo non ha giocato come il grande maestro cinese di Go, **Ke Jie**, ma ha vinto comunque.

Parimenti, le **automobili autonome** non sono auto guidate da robot umanoidi seduti al volante al nostro posto, ma sono **modi per reinventare completamente l'auto e il suo ambiente**.

Nell'IA è il risultato che conta, non se l'agente o il suo comportamento sono intelligenti. Per questo, **l'IA non concerne la capacità di riprodurre l'intelligenza umana, ma in realtà la capacità di farne a meno**.

Il digitale sta re-ontologizzando la natura stessa (e quindi il significato) del nostro ambiente, l'infosfera, la quale al contempo sta progressivamente diventando il mondo in cui viviamo.

Quando parliamo di città smart, facciamo riferimento anche al fatto che stiamo trasformando gli habitat sociali in luoghi in cui i robot possono operare con successo. Da decenni avvolgiamo il mondo intorno alle tecnologie digitali in modo invisibile e senza rendercene interamente conto.

Il futuro dell'IA risiede anche in un maggiore avvolgimento, per esempio, in termini di 5G, e in una crescente dimensione *onlife*, vale a dire nel fatto che siamo tutti costantemente connessi e trascorriamo sempre più tempo nell'infosfera**, mentre tutte le nuove informazioni nascono sempre più digitali.**

1) Negli anni '40 e '50, **il computer era una stanza e vi camminavamo dentro per lavorare con esso e al suo interno.** Programmare significava usare un cacciavite. L'interazione uomo-computer era una relazione somatica o fisica (cfr. computer mostrato in *Robinson Crusoe su Marte*).

2) Negli anni '70 **siamo usciti dal computer, per sederci di fronte a esso.** L'interazione uomo-computer divenne una relazione semantica, resa in seguito più facile dal sistema operativo per dischi, dalle righe di testo.



3) Adesso **il mondo è “avvolto” e viene trasformato in un luogo adattato all'IA e gli esseri umani possono diventare inavvertitamente parte del meccanismo.** **Avvolgere vuol dire adattare l'ambiente e le attività alle capacità dell'IA.** Più sofisticate sono queste capacità, meno necessario è l'avvolgimento.

Il **rischio** che corriamo è che, avvolgendo il mondo, le nostre tecnologie e in particolare l'IA possano plasmare i nostri ambienti fisici e concettuali e costringerci a adattarci a essi dal momento che questo è il modo più semplice o migliore, e talvolta l'unico, per far funzionare le cose.



Dal momento che l'IA è il coniuge stupido ma laborioso e l'umanità quello intelligente ma pigro, chi si adatterà a chi?

L'IA può svolgere un ruolo importante perché abbiamo bisogno di modalità sempre più intelligenti per elaborare immense quantità di dati, in maniera efficiente, efficace, sostenibile ed equa.

Ma l'IA dovrebbe essere trattata come una normale tecnologia, come una delle tante soluzioni che l'ingegno umano è riuscito a escogitare.

Questo potrebbe non essere del tutto esatto, considerata la particolare «natura» dell'IA rispetto a quanto abbiamo già visto e sperimentato in ambito tecnologico.

Allora dobbiamo chiederci se le soluzioni di IA

- **rimpiazzeranno davvero le soluzioni precedenti**, come ha fatto l'automobile con la carrozza;

- le **diversificheranno**, come ha fatto la moto con la bicicletta,

- o le **integreranno ed espanderanno**, come ha fatto lo smartwatch digitale con l'orologio analogico.

L'IA è una nuova forma dell'agire, del fare meglio (non consapevole); che può affrontare con successo compiti e problemi, in vista di un obiettivo, senza alcun bisogno di essere intelligente.



È preferibile comprendere l'IA come una riserva di capacità di agire che può essere usata per risolvere problemi ed eseguire compiti con successo. L'IA consegue i propri obiettivi scindendo la capacità di eseguire un compito con successo da qualsiasi esigenza di essere intelligente nel farlo.

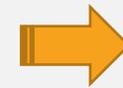
La disgiunzione tra agire, intelligenza e avvolgimento del mondo generano importanti sfide etiche: riguardo ad autonomia, pregiudizi, equità, privacy, responsabilità, trasparenza, fiducia.

Dal 2017, quando sono stati pubblicati i **Principi di Asilomar** (California) per l'IA e la **Dichiarazione di Montréal** per uno sviluppo responsabile dell'IA, molte organizzazioni hanno lanciato un'ampia gamma di iniziative per stabilire principi etici per l'adozione di un'IA socialmente vantaggiosa.

Purtroppo, l'**enorme volume di principi proposti** rischia di risultare soverchiante e fuorviante. O i vari insiemi di principi etici per l'IA sono simili, portando a inutili ripetizioni e ridondanze, oppure, se differiscono in modo significativo, sono suscettibili di ingenerare confusione e ambiguità.



Elon Musk (CEO di Tesla), Stephen Hawking (astrofisico) e altri 2335 ricercatori ed esperti, sotto l'egida del neocostituito Istituto *Future of Life*, hanno approvato un **Cyber-manifesto di 23 "Principi di Asilomar"**.



privacy, democrazia, responsabilità, trasparenza, sicurezza, libertà e condivisione del benessere

L'AI ha già fornito strumenti utili che vengono utilizzati ogni giorno da persone in tutto il mondo. Il suo continuo sviluppo, guidato dai seguenti principi, offrirà incredibili opportunità per l'umanità in futuro.

A) Problemi di ricerca

1) Obiettivo della ricerca sull'Intelligenza Artificiale [AI] non dovrà essere quello di creare un'intelligenza non-orientata [neutrale, indifferente], ma un'intelligenza benefica.

2) Finanziamenti della ricerca dovranno essere mirati a garantirne l'uso benefico, comprese le questioni spinose in informatica, economia, diritto, etica, e studi sociali,

3) Il rapporto tra scienza e politica dovrà essere di sana e costruttiva interazione.

4) Una cultura di cooperazione, fiducia e trasparenza dovrà essere promossa tra i ricercatori e gli sviluppatori.

5) I gruppi di ricerca dovranno cooperare ed evitare scorciatoie sugli standard di sicurezza.

B) Etica e Valori

6) I sistemi di AI devono essere sicuri e protetti per tutta la loro vita operativa e verificabili, se necessario.

7) Se un sistema di AI provoca danni, dovrà essere possibile accertarne con trasparenza il motivo.

8) Qualsiasi coinvolgimento da parte di un sistema di AI giuridico autonomo dovrà essere in grado di fornire una spiegazione soddisfacentemente [per le proprie decisioni] verificabile da un'autorità umana competente (trasparenza giudiziaria).

9) Progettisti e costruttori di sistemi avanzati di AI sono parti interessate nelle implicazioni morali del loro uso, abuso ed azioni, con la responsabilità e l'opportunità di plasmare tali implicazioni.

10) I sistemi di IA autonomi dovranno essere progettati in modo che i loro obiettivi e comportamenti possano essere affidabilmente allineati con i nostri valori in tutto il loro funzionamento.

11) I sistemi di IA dovranno essere compatibili con gli ideali di dignità umana, dei diritti, della libertà e della diversità culturale.

12) Andrà garantito il diritto di accedere, gestire e controllare i dati generati dagli utenti, dato il potere dei sistemi di AI di analizzare e utilizzare tali dati.

13) L'applicazione di IA ai dati personali non deve irragionevolmente limitare la libertà delle persone.

14) Le tecnologie di IA dovranno essere condivise a beneficio del maggior numero di persone possibile.

15) La prosperità creata dall'IA dovrà essere condivisa, in linea di massima, a beneficio di tutta l'umanità.

16) Gli esseri umani dovranno mantenere il controllo e scegliere come e se delegare decisioni ai sistemi di AI.

17) Il potere conferito dal controllo dei sistemi di IA altamente avanzati dovrà rispettare e migliorare, e non sovvertire, i processi sociali e civili da cui dipende la salute della società.

18) Una corsa agli armamenti in armi letali autonome dovrà essere evitata.

Problemi a lungo termine

19) Non essendoci un consenso, dovremo evitare di supporre quali possano essere i limiti superiori delle future capacità dell'IA.

20) L'arrivo delle AI avanzate potrebbero rappresentare un cambiamento profondo nella storia della vita sulla Terra, e dovrà essere pianificata e gestita con l'attenzione e le risorse necessarie.

21) I rischi associati ai sistemi di IA, in particolare i rischi catastrofici o esistenziali, devono essere oggetto di pianificazione e mitigazione commisurati con il loro potenziale impatto.

22) I sistemi di IA progettati per auto-migliorarsi o auto-replicarsi in maniera da portare ad un rapido aumento della loro qualità o quantità, devono essere soggetti a severe misure di sicurezza e di controllo.

23) La Super intelligenza dovrà essere sviluppata solo al servizio di ideali etici ampiamente condivisi e per il beneficio di tutta l'umanità, piuttosto che di un singolo Stato o organizzazione.

L'affermazione dell'IA come campo di ricerca accademica risale agli anni '50. Il dibattito etico è quasi altrettanto vecchio. Tuttavia, è solo negli ultimi anni che i notevoli progressi nelle capacità e nelle applicazioni dei sistemi di IA hanno portato maggiormente all'attenzione le **opportunità e i rischi** che comporta per la società.

L'esigenza crescente di riflessione e politiche chiare sull'impatto dell'IA sulla società ha prodotto un eccesso di iniziative.

Ogni nuova iniziativa produce un'ulteriore dichiarazione di principi, valori o ideali per guidare lo sviluppo e l'adozione dell'IA, anche **aggiornando le regole del gioco**, sebbene l'etica – in particolare l'etica digitale – non può essere una mera suppellettile, un ripensamento, un ultimo arrivato, una **civetta di Minerva** (metafora usata da Hegel) sulla filosofia, **che si alza in volo solo quando le ombre della notte si stanno addensando.**

Le posizioni degli scienziati su questo punto sono diverse, ma sostanzialmente riconducibili a due fronti:

1) **Da un lato** un numero crescente di scienziati ritiene che ci dovrebbe essere una supervisione normativa, magari a livello nazionale e internazionale, che si assicuri che non facciamo qualcosa di sbagliato e dannoso.

2) **Dall'altro lato** c'è chi pensa che ci troviamo, di fatto, di fronte a un falso problema.

L'IA è solo computer, i computer sono solo macchine di Turing, le macchine di Turing sono solo motori sintattici e i motori sintattici non possono pensare, non possono sapere, non possono essere coscienti.

Non esiste una vera IA, e per questo non ci sono a fortiori problemi causati da essa. Possiamo rilassarci e goderci tutti questi meravigliosi dispositivi elettrici.

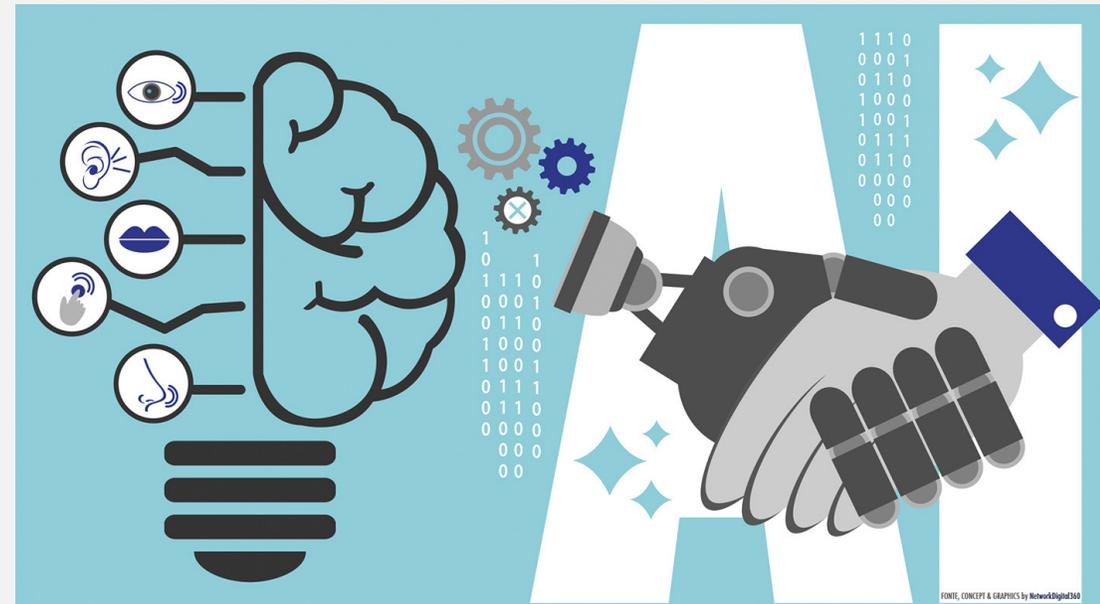
CONCLUSIONI: MACHINA SAPIENS?

L'**IA e i robot** sono la concretizzazione dell'evoluzione tecnologica dell'informazione e del mondo compreso come **una serie di dati**: siamo in grado di costruire macchine che possono prendere decisioni autonome e coesistere con l'uomo (macchine a guida autonoma; sistemi di radio chirurgia; robot destinati al lavoro affianco all'uomo nei processi produttivi in fabbrica, ecc.).

Si tratta, dunque, di nuove tecnologie pervasive perché stanno insinuandosi in ogni ambito della nostra esistenza: tanto nei sistemi di produzione, incarnandosi in robot, quanto nei sistemi di gestione, sostituendo i server e gli analisti.

Anche nella vita quotidiana i sistemi di IA sono sempre più pervasivi: gli smartphone di ultima generazione sono tutti venduti con un assistente dotato di intelligenza artificiale, **Cortana, Siri o Google Hello**, che trasforma il telefono da un *hub* di servizi e applicazioni a un vero e proprio partner che interagisce in maniera cognitiva con l'utente.

Sono in fase di sviluppo crescente sistemi di intelligenza artificiale, i bot, che saranno disponibili come partner virtuali da interrogare via voce o in chat che sono in grado di fornire servizi e prestazioni che primano erano esclusiva di particolari professioni: avvocati, medici e psicologi, ecc.



1) Se da un lato le preoccupazioni per la crescita esponenziale e potenzialmente incontrollata dell'IA (del suo utilizzo e delle sue applicazioni) sono motivate, è anche vero che 2) il mondo del lavoro conosce oggi una nova frontiera: le interazioni e la coesistenza tra uomini e intelligenze artificiali.

Questo dato ci deve far considerare un implicito culturale che rischia di sviare la nostra comprensione del tema.

Nello sviluppo delle IA la divulgazione dei successi ottenuti da queste macchine è sempre stata presentata secondo un **modello competitivo rispetto all'uomo**:

- IBM ha presentato *Deep Blue* come l'intelligenza artificiale che nel 1996 riuscì a sconfiggere a scacchi il campione del mondo in carica, Garry Kasparov;

- IBM nel 2011 ha realizzato *Watson* che ha sconfitto i campioni di un noto gioco televisivo sulla cultura generale *Jeopardy!*.

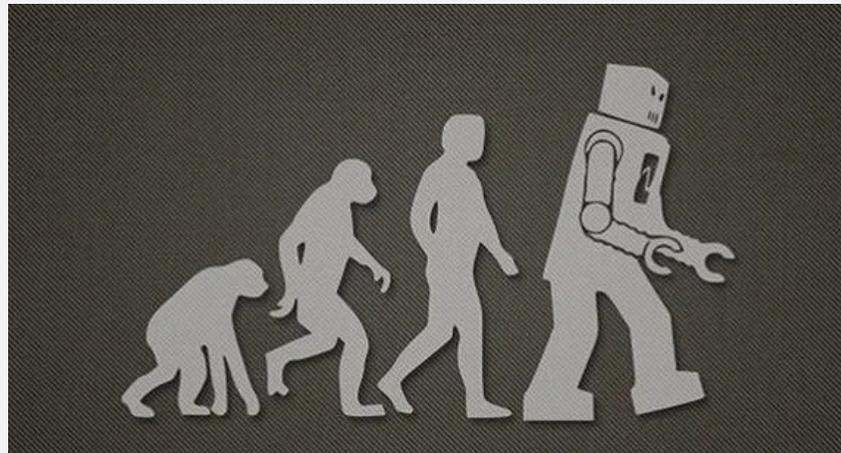
Si tratta di esempi che potrebbero farci pensare che questi sono sistemi che competono con l'uomo e che tra *Homo sapiens* e questa nuova *Macchina sapiens* si sia instaurata una rivalità di natura evolutiva che vedrà un solo vincitore e condannerà lo sconfitto a una inesorabile estinzione.

In realtà queste macchine non sono mai state costruite per competere con l'uomo ma per realizzare una nuova simbiosi tra l'uomo e i suoi artefatti: (*Homo+Machina*) *sapiens*.

Non sono le AI la minaccia di estinzione dell'uomo, anche se la tecnologia può essere pericolosa per la nostra sopravvivenza come specie. L'uomo ha già rischiato di estinguersi perché battuto da una macchina molto stupida come la bomba atomica.

Esistono tuttavia delle sfide estremamente delicate nella società contemporanea in cui **la variabile più importante non è l'intelligenza ma il poco tempo a disposizione per decidere e le macchine cognitive trovano qui grande interesse applicativo.**

Il pericolo maggiore non viene dalle IA in se stesse ma dal non conoscere queste tecnologie e dal lasciare decidere sul loro impiego a una classe dirigente assolutamente non preparata a gestire il tema.



UNA NUOVA REALTÀ?

Una serie di questioni che sorgono da questi nuovi artefatti e che interrogano anche la filosofia riguardano la **differenza tra naturale e artificiale**.

Ci siamo dotati di un nuovo strumento di indagine. Tre secoli fa con le lenti concave abbiamo realizzato il **telescopio** e il **microscopio**, imparando a vedere il mondo in modo diverso. Microscopio e telescopio costituirono gli strumenti tecnologici con cui la rivoluzione scientifica del '600 e del '700 ha ottenuto le sue scoperte: abbiamo reso visibile l'estremamente lontano e l'estremamente piccolo.

Oggi con i dati abbiamo realizzato un nuovo "strumento": il macroscopio. Con i **big data** riusciamo a vedere in maniera nuova e sorprendente l'**estremamente complesso** delle relazioni sociali, individuando relazioni e connessioni dove prima non vedevamo nulla.

L'IA applicata a questi enormi set di dati è il macroscopio con cui studiare meccanicisticamente l'estremamente complesso.

Lo sviluppo delle **biotecnologie**, che iniziano a capire la vita stessa come un processo informazionale espresso dal DNA, mettono in crisi il significato stesso di una distinzione tradizionale con cui comprendevamo la realtà: quella tra naturale ed artificiale.

I prodigi della **biologia sintetica** sembrano far emergere l'incertezza, se non addirittura l'impossibilità, di definire un qualsivoglia limite. **I confini tra naturale e artificiale stanno sfumando, rendendosi indistinguibili**.

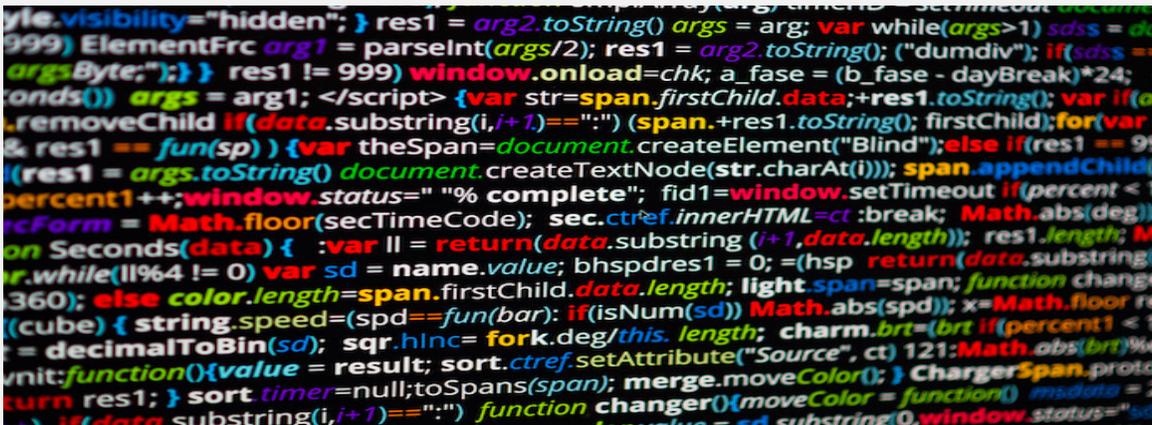
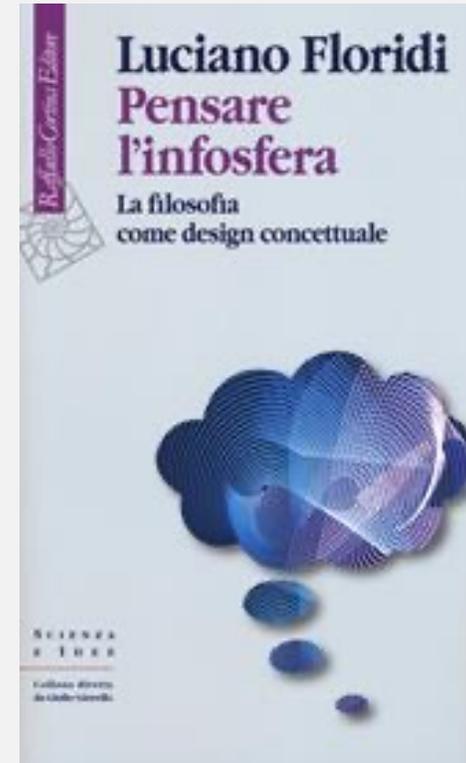
L'avvento del **cyborg**, cioè la **fusione tra uomo e macchina**, la nascita della biologia sintetica e lo sviluppo della carne in provetta mostrano come tutto sia divenuto manipolabile e dominabile: **gli artefatti che produciamo non sono meramente artificiali, né sono naturali, ma hanno una natura spesso ibrida**.

Spetta dunque a noi capire che tipo di conoscenza stiamo generando: se questa forma di conoscenza sia scientifica e in che senso sia deterministica o predittiva. Tuttavia la rivoluzione conoscitiva è già in atto.

Assistiamo al sorgere di **nuove sfide**:

- Per alcuni si tratta della sfida di cominciare a concepire l'uomo come animale informazionale al fianco di altri, inserito all'interno dell'infosfera;

- Per altri si tratta della sfida di farsi carico di una nuova società, la *società dell'informazione*, cresciuta molto più rapidamente della capacità dell'uomo di sviluppare solide radici concettuali, etiche e culturali in grado di **comprenderla, gestirla e orientarla** verso il bene comune e lo sviluppo.



UNA NUOVA UMANITÀ?

I profondi cambiamenti indotti dall'irruzione dell'informazione e dagli artefatti biotecnologici suscita **nuove domande sull'uomo e sulla sua identità**: la questione antropologica diventa un luogo chiave dove la filosofia e la scienza si devono confrontare, soprattutto per affrontare le sfide del *post-umanesimo* e del *trans-umanesimo*.



La visione dell'**uomo come di un essere malleabile, dall'identità aperta, modulare, porosa**, è uno dei principali cardini del pensiero *post-umano* e *trans-umano*, suscettibile di innumerevoli trasformazioni, una *vita liquida*.

* Il valore dell'umano non è più la persona ma le informazioni che abitano il suo corpo biologico: **ogni uomo è visto come un insieme di informazioni contenute in un *medium* che è il suo corpo.**

Il valore dell'uomo è espresso in termini di informazioni e **la sua essenza diviene un qualcosa di computabile e gestibile come un flusso di informazioni.**

La vita stessa diviene la capacità di conservare ed elaborare informazioni.

La *malleabilità* dell'uomo si trasforma in una sostanziale svalutazione del corpo e della corporeità visti come *accidens* dell'esistenza.

Alla luce di quanto emerso appare evidente come **il fenomeno tecnologico sia anzitutto un “luogo” filosofico che interroga la nostra comprensione del mondo e dell’umano.**

In quanto appartenenti alla specie *Homo sapiens* siamo esseri simbolici, dotati di linguaggio duttile e flessibile.

Questa caratteristica ci ha permesso di avere la capacità senza precedenti di cooperare tra grandi numeri di individui.

Siamo in grado di evidenziare una caratteristica peculiare e irriducibile che *distacca* l’uomo da ogni altra specie esistente: **grazie a questa *condizione cognitiva* l’uomo può rivedere e modificare il proprio comportamento con rapidità, conformandosi al mutare delle necessità.**

L’evoluzione culturale soppianta la lenta e imprevedibile evoluzione genetica: grazie a questa caratteristica la nostra specie ha iniziato un cammino che mostra nei segni archeologici della storia una distanza sempre più marcata da tutte le altre specie.





Per concludere

«Siamo un *hapax legomenon* nel Libro della Natura di Galileo, un po' come l'espressione *gopher* ("legno di cipresso"), che si riferisce al materiale originario con cui è stata costruita l'arca di Noè e che ricorre una sola volta in tutta la Bibbia. Con una metafora più digitale e contemporanea, siamo un bellissimo errore nel grande software dell'universo, non l'app di maggior successo. Resteremo un errore unico e riuscito, mentre l'IA sarà ancora di più un elemento peculiare nel libro matematico della natura di Galileo».

(Luciano Floridi)