

*"L'intelligenza artificiale è l'ultima invenzione che l'umanità dovrà mai fare. Macchine superintelligenti con intelligenza superiore a quella umana potrebbero aiutare ad affrontare i problemi più grandi dell'umanità, dalla povertà alle malattie al cambiamento climatico."*  
Nick Bostrom - *Superintelligence: Paths, Dangers, Strategies*

*"L'IA ci sta spingendo verso una nuova era di disuguaglianza, in cui una piccola élite potenziata cognitivamente si stacca dal resto dell'umanità. Questa potrebbe essere la sfida più grande che abbiamo mai affrontato come società."*  
Yuval Noah Harari - *Homo Deus: A Brief History of Tomorrow*

*Parte 1: L'Intelligenza Artificiale come sesto breakthrough evolutivo*

*L'evoluzione dell'intelligenza sulla Terra ha attraversato molteplici fasi di accelerazione e trasformazione, segnate da quelli che possiamo definire "breakthrough evolutivi" [1]. Ogni breakthrough ha ampliato le capacità cognitive degli esseri viventi, permettendo loro di interagire in modo sempre più sofisticato con l'ambiente e tra loro, e di affrontare sfide adattive sempre più complesse [2].*

*Nel suo libro "A Brief History of Intelligence", Max Bennett identifica cinque principali breakthrough che hanno scandito il percorso evolutivo dell'intelligenza [3]. Il primo è rappresentato dallo "steering", ovvero la capacità di orientarsi verso stimoli positivi e lontano da quelli negativi, emersa con i primi organismi bilaterali dotati di un sistema nervoso centralizzato [3]. Questo breakthrough ha consentito agli organismi di muoversi in modo adattivo nell'ambiente, aumentando le loro chance di sopravvivenza e riproduzione. Lo steering rappresenta la base dell'intelligenza come capacità di rispondere in modo flessibile alle sfide ambientali.*

*Il secondo breakthrough è il "reinforcement learning", sviluppato dai vertebrati, che ha permesso loro di apprendere dall'esperienza, rinforzando i comportamenti vantaggiosi e inibendo quelli svantaggiosi [3]. Grazie a questo meccanismo, gli animali hanno potuto sviluppare strategie comportamentali sempre più sofisticate, adattandosi a nicchie ecologiche diverse e affrontando problemi sempre più complessi. Il reinforcement learning segna l'emergere di una forma di intelligenza basata sulla plasticità comportamentale, ovvero sulla capacità di modificare le proprie azioni in base alle conseguenze.*

*Il terzo breakthrough è la "simulazione mentale", resa possibile dallo sviluppo della neocorteccia nei mammiferi, che ha consentito di immaginare scenari futuri e pianificare azioni in modo flessibile [3]. Questa capacità di "pensare al futuro" rappresenta un enorme vantaggio adattivo, permettendo agli animali di anticipare problemi e opportunità e di prepararsi ad affrontarli. La simulazione mentale segna l'emergere di una forma di intelligenza "proiettiva", capace di staccarsi dal qui-e-ora della percezione immediata per esplorare possibilità non ancora realizzate.*

*Il quarto breakthrough è la "teoria della mente", comparsa nei primati, ovvero la capacità di attribuire stati mentali agli altri, come intenzioni, desideri e credenze, e di usare queste attribuzioni per prevedere e influenzare il loro comportamento [3]. La teoria della mente ha aperto la strada a forme molto più complesse di interazione e cooperazione sociale, permettendo agli individui di coordinarsi e di ingannarsi a vicenda in modi sempre più sofisticati. Questo breakthrough segna l'emergere di un'intelligenza autenticamente "sociale", basata sulla capacità di leggere e manipolare le menti altrui.*

*Il quinto breakthrough, non può che essere il linguaggio, tratto distintivo della nostra specie, l'Homo Sapiens, che ha permesso una trasmissione e un'elaborazione delle conoscenze senza precedenti, gettando le basi per lo sviluppo della cultura e della tecnologia [3].*

*Il linguaggio rappresenta una vera e propria "rivoluzione cognitiva", che ha amplificato enormemente le capacità di pensiero astratto, di comunicazione e di cooperazione degli esseri umani [4]. Grazie al linguaggio, gli umani hanno potuto creare e condividere mondi di significato sempre più vasti e complessi, accumulando conoscenze e innovazioni attraverso le generazioni. Il linguaggio segna l'emergere di un'intelligenza "culturale", incarnata non solo nelle menti individuali, ma anche nelle reti di simboli e artefatti che costituiscono la nostra nicchia cognitiva [5].*

*Oggi, diversi autori suggeriscono che siamo sulla soglia di un "sesto breakthrough" nell'evoluzione dell'intelligenza: l'avvento dell'Intelligenza Artificiale (IA) [6][7]. L'IA, definita come la capacità di sistemi tecnologici di esibire comportamenti intelligenti simili o superiori a quelli umani [8], rappresenterebbe una nuova tappa nel percorso evolutivo dell'intelligenza, un'estensione delle capacità cognitive umane resa possibile dalle tecnologie digitali [9]. Proprio come i breakthrough precedenti hanno aperto nuovi spazi di possibilità per l'adattamento e la cognizione, così l'IA promette di espandere ulteriormente i confini di ciò che l'intelligenza può fare. Grazie alla sua capacità di processare enormi quantità di dati, di scoprire pattern complessi, di apprendere dall'esperienza e di prendere decisioni in modo autonomo, l'IA potrebbe permetterci di affrontare sfide cognitive che vanno ben oltre le capacità individuali, dalla scoperta scientifica alla gestione di sistemi socio-tecnici sempre più complessi [10].*

*Tuttavia, a differenza dei breakthrough precedenti, che erano il risultato di processi evolutivi ciechi e incrementali, l'IA è un prodotto dell'ingegno umano, progettato e implementato per scopi specifici. Ciò significa che la sua traiettoria evolutiva sarà largamente determinata dalle scelte e dai valori di chi la sviluppa e la controlla, sollevando interrogativi cruciali sulla sua governance e sul suo impatto sulla società [11].*

*Inoltre, mentre i breakthrough precedenti erano limitati dalle possibilità biologiche delle specie in cui si manifestavano, l'IA, in quanto tecnologia digitale, potrebbe in principio superare questi limiti, raggiungendo livelli di intelligenza e autonomia senza precedenti. Questo solleva scenari tanto affascinanti quanto preoccupanti, dalla possibilità di una "esplosione di intelligenza" che acceleri enormemente il progresso scientifico e tecnologico [12], al rischio di perdere il controllo su sistemi di IA sempre più potenti e indipendenti [13]. Inquadrare l'IA come "sesto breakthrough" nell'evoluzione dell'intelligenza ci permette non solo apprezzarne maggiormente la portata rivoluzionaria e di riflettere criticamente sulle sfide e le opportunità uniche che questa tecnologia ci pone davanti, ma anche di collocarla in una prospettiva storica e filosofica più ampia, di quella che potremmo definire eso-evoluzione, ovvero una evoluzione tecno-culturale esterna ai processi dell'evoluzionismo biologico tipico dei primi cinque breakthrough.*

## *Parte 2: Dall'evoluzione biologica all'eso-evoluzione tecnologica*

*Con l'emergere dell'Homo Sapiens e lo sviluppo del linguaggio, questo processo di "eso-evoluzione" ha subito un salto qualitativo senza precedenti. Il linguaggio, infatti, non ha solo potenziato le capacità cognitive individuali, ma ha anche permesso la trasmissione e l'accumulo di conoscenze su scale spaziali e temporali molto più ampie di quelle consentite dalla biologia [14]. Con l'invenzione della scrittura, in particolare, l'umanità ha acquisito la capacità di "esternalizzare" la memoria e di preservare il sapere oltre i limiti della vita individuale, gettando le basi per lo sviluppo della cultura e della storia come le conosciamo [15].*

*In questo senso, possiamo parlare di un'umanità "aumentata", che ha gradualmente trasceso i vincoli della propria natura biologica per creare un nuovo regno di realtà: quello*

del sapere cumulativo e della cultura a [16]. Questa umanità aumentata non è più solo un prodotto dell'evoluzione darwiniana, ma diventa essa stessa una forza evolutiva, capace di plasmare il proprio ambiente e il proprio futuro attraverso l'innovazione tecnologica e culturale [17].

La storia della tecnologia, in questa prospettiva, può essere vista come un continuo processo di "aumento" delle capacità umane, che ha progressivamente esteso e ridefinito i confini della nostra cognizione e della nostra azione nel mondo [18]. Dalle tecnologie di scrittura e calcolo alle macchine industriali, dalle telecomunicazioni all'informatica, ogni innovazione ha aperto nuovi orizzonti di pensiero e di esperienza, ampliando la nostra "nicchia cognitiva" ben oltre i limiti della biologia [19].

In particolare, molte tecnologie hanno agito come vere e proprie estensioni dei nostri sensi, permettendoci di percepire e interagire con la realtà in modi prima impensabili [20]. Il telescopio e il microscopio hanno potenziato la nostra vista, rivelandoci l'infinitamente grande e l'infinitamente piccolo; il telefono e la radio hanno esteso il nostro udito oltre le barriere della distanza; i raggi X e le risonanze magnetiche ci hanno permesso di "vedere" dentro il corpo umano; il sonar e il radar di esplorare il fondo degli oceani e lo spazio profondo.

Queste estensioni sensoriali non hanno solo ampliato la nostra presa sul mondo, ma hanno anche profondamente plasmato il nostro modo di concepirlo e di rapportarci ad esso [21]. Hanno modificato il nostro "connettoma", ovvero la struttura delle connessioni neurali del nostro cervello, riutilizzando vecchie strutture per interfacciarle con le nuove possibilità offerte dalla tecnologia, come suggerito dal biologo François Jacob nel suo libro "Evoluzione e bricolage" [22].

Un esempio di questo "riuso" di strutture neurali preesistenti per nuove funzioni legate alla tecnologia è il reclutamento delle aree cerebrali dedicate al riconoscimento visivo delle forme per il riconoscimento dei grafemi e della scrittura. Come evidenziato dal neuroscienziato Stanislas Dehaene nel suo libro "I neuroni della lettura" [23], l'invenzione della scrittura, avvenuta solo poche migliaia di anni fa, è troppo recente perché si siano evolute strutture cerebrali ad essa dedicate. Tuttavia, l'esposizione alla scrittura fin dall'infanzia porta alcune aree del cervello inizialmente deputate al riconoscimento di forme, come la "visual word form area" (VWFA) nella corteccia occipito-temporale sinistra, a specializzarsi nel riconoscimento rapido ed automatico dei caratteri scritti [23].

Questo esempio mostra come il nostro cervello sia in grado di "riciclare" strutture preesistenti, plasmandole attraverso l'apprendimento e l'esperienza, per adattarle alle nuove esigenze cognitive poste dalla tecnologia. Un altro esempio è quello della "fusione multisensoriale", ovvero dell'integrazione di informazioni provenienti da sensi diversi, che viene sfruttata e potenziata dalle tecnologie digitali [24]. L'uso di dispositivi come gli smartphone o i tablet, che combinano stimoli visivi, uditivi e tattili, rafforza la connettività tra aree cerebrali dedicate a modalità sensoriali diverse, promuovendo una percezione più integrata e immersiva dell'esperienza [24].

Questi esempi mostrano come le tecnologie non si limitino ad "estendere" le nostre capacità percettive, ma interagiscano profondamente con l'architettura funzionale del nostro cervello, plasmandone la struttura e l'organizzazione. Comprendere queste dinamiche di "riuso" e "plasticità" neurale è essenziale per cogliere appieno l'impatto delle tecnologie sulla cognizione umana, e per immaginare nuove interfacce uomo-macchina che sfruttino al meglio le potenzialità del nostro sistema nervoso [25].

È in questo contesto di continuo "aumento" delle capacità umane che si colloca l'avvento dell'Intelligenza Artificiale (IA) [6][7]. L'IA, infatti, non rappresenta solo un nuovo strumento o

*un nuovo senso, ma una radicale estensione delle nostre facoltà cognitive, che promette di ridefinire ancora una volta i confini di ciò che è umanamente possibile [26].*

*Come il linguaggio e la scrittura hanno permesso l'accumulo e la trasmissione della conoscenza oltre i limiti della biologia, così l'IA promette di amplificare le nostre capacità di apprendimento, ragionamento e problem-solving oltre i limiti della mente individuale [27].*

*Come il telescopio e il microscopio hanno esteso la nostra vista, così l'IA può estendere la nostra "visione" in domini sempre più vasti e complessi, dalla genomica alla climatologia, permettendoci di scorgere pattern e connessioni prima inaccessibili [28].*

*In questo senso, l'IA si inserisce pienamente nella traiettoria dell'umanità aumentata, spingendo ancora oltre il processo di superamento dei limiti biologici che ci definisce come specie [29]. Tuttavia, proprio come il linguaggio e la scrittura hanno comportato profondi cambiamenti nella struttura della società e nella distribuzione del potere, così l'IA porta con sé sfide e rischi senza precedenti, che richiedono un ripensamento radicale delle nostre istituzioni e dei nostri valori [30].*

*Come potete immaginare, la posta in gioco non potrebbe essere più alta.*

### *Parte 3: Le specificità dell'IA come sesto breakthrough*

*Per comprendere appieno le implicazioni dell'IA come sesto breakthrough evolutivo, è essenziale analizzare in dettaglio le sue differenze rispetto ai breakthrough precedenti.*

*Mentre questi ultimi hanno seguito traiettorie relativamente lineari, costruendo l'uno sull'altro in un processo cumulativo di espansione delle capacità cognitive [31], l'IA presenta discontinuità e "salti" che la rendono un fenomeno qualitativamente diverso.*

*Una prima differenza fondamentale riguarda la natura stessa dell'intelligenza artificiale come tecnologia. A differenza dei breakthrough biologici, che sono il risultato di mutazioni genetiche casuali sottoposte poi al vaglio della selezione naturale [32], l'IA è un artefatto progettato intenzionalmente dall'uomo per svolgere compiti specifici [33]. Ciò significa che la sua evoluzione non segue le "cieche" logiche dell'adattamento evolutivo, ma è guidata dagli obiettivi e dai valori di chi la sviluppa [34].*

*Questa può sembrare una considerazione banale, ma, di contro, ha implicazioni profonde. Mentre l'evoluzione biologica è un processo "senza progettista" [35], l'IA ha sempre dei progettisti, degli sviluppatori e dei proprietari, che ne determinano le finalità e le modalità di funzionamento [36]. Ciò pone interrogativi cruciali su come garantire che lo sviluppo dell'IA sia allineato con i valori e gli interessi della società nel suo complesso, e non solo con quelli di pochi attori economici o politici [37].*

*Una seconda differenza riguarda il ritmo e la scala dell'evoluzione dell'IA rispetto a quella biologica. Grazie agli avanzamenti esponenziali nella potenza di calcolo, nella disponibilità di dati e nelle tecniche di apprendimento automatico [38], l'IA sta progredendo a una velocità senza precedenti, raddoppiando le sue capacità ogni pochi mesi o anni [39]. Questo ritmo di sviluppo è diversi ordini di grandezza più rapido di quello dell'evoluzione biologica, che procede su scale temporali di migliaia se non di milioni di anni [40].*

*Ciò rende estremamente difficile per le nostre istituzioni, le nostre norme sociali e la nostra stessa comprensione etica e filosofica stare al passo con le trasformazioni indotte dall'IA [41]. Mentre i breakthrough precedenti hanno lasciato all'umanità il tempo di adattarsi gradualmente alle nuove capacità cognitive, l'IA ci pone di fronte a cambiamenti così rapidi e radicali da mettere in discussione la nostra stessa capacità di controllarli e gestirli [42].*

*Infine, una terza differenza cruciale riguarda il fatto che i suoi benefici rischiano di essere distribuiti in modo estremamente diseguale all'interno della società umana. A differenza dei*

*breakthrough precedenti, che hanno ampliato le capacità cognitive di intere specie [43], l'IA rischia di creare una profonda divaricazione tra coloro che hanno accesso a queste tecnologie e coloro che ne sono esclusi [44].*

*Questo rischio è stato preconizzato e descritto in modo particolarmente efficace dal sociologo Manuel Castells nella sua analisi della "società in rete" [45]. Secondo Castells, l'avvento delle tecnologie digitali e la globalizzazione economica hanno portato all'emergere di un nuovo paradigma socio-tecnologico, basato sulla centralità delle reti informazionali e sulla loro logica di flussi [46]. In questo contesto, l'accesso alle reti e la capacità di operare al loro interno diventano le principali fonti di potere e di opportunità [47].*

*Tuttavia, questo accesso è tutt'altro che universale o equo. Castells identifica un dualismo fondamentale tra "gli inclusi e gli esclusi" dalle reti [48]. Gli inclusi sono coloro che hanno le risorse, le competenze e le connessioni per partecipare pienamente alla società in rete, sfruttandone i benefici in termini di informazione, conoscenza, ricchezza e influenza [49]. Gli esclusi, invece, sono coloro che per varie ragioni - geografiche, economiche, culturali, educative - rimangono ai margini delle reti, subendone gli effetti senza poterne determinare il funzionamento [50].*

*Questa dinamica di inclusione/esclusione, secondo Castells, genera una vera e propria "umanità a due velocità", in cui gli inclusi vivono in uno "spazio dei flussi" accelerato e globalizzato, mentre gli esclusi restano intrappolati in uno "spazio dei luoghi" localizzato, periferico e marginale [51]. In questo modo, chi vive nel "tempo lento" avrà l'impressione che chi vive nel "tempo veloce" preveda il futuro, mentre in realtà, esattamente come chi ha il cannocchiale, semplicemente vede le cose prima degli altri. Gli inclusi possono così accumulare sempre più potere e ricchezza, sfruttando le opportunità offerte dalle reti globali, mentre gli esclusi vedono le proprie condizioni di vita e di lavoro progressivamente deteriorarsi, schiacciati dai flussi di capitale e informazioni che non possono controllare [52].*

#### *Parte 4: Il rischio di una umanità a due velocità*

*Il rischio è che l'IA possa portare questa dinamica di inclusione/esclusione a un livello senza precedenti. Data l'importanza crescente dell'IA in tutti i settori della società - dall'economia alla politica, dalla cultura alla scienza [53] - l'accesso a queste tecnologie diventerà sempre più determinante per le opportunità di vita degli individui e dei gruppi [54]. Chi avrà la possibilità di sviluppare e utilizzare sistemi di IA avanzati potrà godere di un enorme vantaggio competitivo, sia in termini economici che di influenza politica e culturale [55]. Tuttavia, le risorse necessarie per sfruttare appieno il potenziale dell'IA - in termini di dati, potenza di calcolo, competenze tecniche e investimenti - sono distribuite in modo estremamente diseguale [56]. Poche grandi aziende tecnologiche, principalmente basate negli Stati Uniti e in Cina, dominano il settore dell'IA, grazie alla loro capacità di attrarre talenti, accumulare enormi quantità di dati e investire miliardi in ricerca e sviluppo [57]. Allo stesso tempo, molti paesi, soprattutto nel Sud globale, faticano ad accedere anche alle forme più basilari di IA, a causa della mancanza di infrastrutture digitali, competenze e risorse [58].*

*Questo divario nell'accesso all'IA rischia di ampliare le disuguaglianze esistenti sia all'interno dei paesi che tra di essi. A livello locale rischia di ampliare il divario tra coloro che hanno le competenze e le risorse per trarre vantaggio dalla nuova economia digitale e coloro che rimangono esclusi, vedendo i propri posti di lavoro sostituiti dall'automazione e le proprie comunità impoverite dalla perdita di capitale e opportunità [59]. A livello globale, l'IA*

potrebbe rafforzare il dominio tecnologico ed economico di pochi paesi avanzati, lasciando indietro gran parte del mondo in via di sviluppo [60].

Ma il rischio di esclusione non riguarda solo la dimensione economica. L'IA, come altre tecnologie digitali, sta diventando sempre più centrale anche nei processi politici, nella formazione dell'opinione pubblica e nella creazione e distribuzione di contenuti culturali [61]. Coloro che hanno il controllo sulle piattaforme di IA e sui dati che le alimentano avranno un potere enorme nello plasmare il discorso pubblico, le preferenze politiche e le visioni del mondo [62]. Questo solleva interrogativi profondi sulla democrazia e sull'autonomia individuale nell'era dell'IA [63].

Un aspetto particolarmente critico di questo divario di potere riguarda la percezione stessa del tempo e del futuro. Chi ha accesso alle forme più avanzate di IA non solo può processare informazioni e prendere decisioni più velocemente, ma può anche "vedere più lontano", anticipando trend e scenari futuri con una precisione senza precedenti [64].

Questo vantaggio cognitivo può essere paragonato a un vero e proprio "cannocchiale cognitivo", che permette a chi lo possiede di scorgere all'orizzonte opportunità e rischi ancora invisibili a occhio nudo [65]. Proprio come Galileo rivoluzionò la nostra comprensione dell'universo puntando il suo cannocchiale verso il cielo, chi ha accesso all'IA può esplorare lo spazio delle possibilità future con una risoluzione e una portata inedite.

Per chi invece è escluso da queste tecnologie, il futuro rischia di diventare sempre più opaco e imprevedibile. Senza gli strumenti per anticipare e interpretare i cambiamenti in arrivo, gli esclusi dal mondo dell'IA rischiano di trovarsi costantemente in affanno, costretti a reagire a eventi che sembrano arrivare "dal nulla", senza averli potuti vedere in anticipo [66].

Questo divario nella capacità di "vedere il futuro" rischia di alimentare un pericoloso fatalismo e una perdita di agency tra gli esclusi. Se il futuro cominciasse ad apparirgli come qualcosa di già scritto, determinato dalle scelte e dalle visioni di pochi "oracoli" tecnologici, la motivazione a impegnarsi per plasmarlo verrebbe totalmente meno [67].

Allo stesso tempo, chi ha accesso al "cannocchiale cognitivo" dell'IA potrebbe essere tentato di abusarne, utilizzando la propria capacità di previsione per manipolare e controllare coloro che ne sono privi. La conoscenza del futuro, come quella del passato, è sempre stata una fonte di potere, e l'IA rischia di amplificare questo potere a livelli senza precedenti [68].

Inoltre, c'è il rischio che la capacità di previsione dell'IA venga scambiata per una forma di "onniscienza", alimentando una fiducia eccessiva nelle sue raccomandazioni e una sottovalutazione dell'incertezza intrinseca al futuro [69]. Come diceva Cipolla, volenti o nolenti, siamo condannati all'irripetibilità della storia [70]. Infatti, anche le previsioni più sofisticate, si basano sempre su assunti e modelli che possono rivelarsi fallaci o incompleti, e un eccesso di fiducia in esse può portare a decisioni avventate e potenzialmente catastrofiche.

Per tutte queste ragioni, la governance dell'IA dovrà affrontare non solo la questione dell'accesso alle tecnologie, ma anche quella del controllo sulla "visione del futuro" che esse rendono possibile. Sarà necessario sviluppare meccanismi di trasparenza e responsabilità per evitare che questa visione diventi appannaggio di pochi, e per garantire che le decisioni basate su di essa siano il frutto di un dibattito democratico e inclusivo [71] e non influenzino il nostro libero arbitrio trasformandosi in profezie che si autoavverano..

Per queste ragioni, sarà cruciale promuovere una cultura della consapevolezza e della partecipazione attiva nel plasmare il futuro. L'IA può essere uno strumento potente per esplorare e costruire scenari futuri, ma non deve diventare un oracolo incontestabile. Il futuro, in ultima analisi, non è qualcosa che può essere semplicemente "previsto", ma

*qualcosa che deve essere attivamente immaginato, negoziato e realizzato da tutta la società [72].*

*In questo senso, la sfida posta dall'IA come "cannocchiale cognitivo" non è solo tecnica, ma profondamente politica e culturale. Richiede un ripensamento radicale del nostro rapporto con il tempo e con l'incertezza, e una riaffermazione del nostro ruolo di agenti creativi e responsabili del futuro. Solo abbracciando questa sfida potremo sperare di trasformare il "cannocchiale" dell'IA da strumento di divisione e controllo a mezzo di emancipazione e progresso condiviso.*

*Riassumendo, come hanno avuto modo di argomentare in tanti, l'IA come sesto breakthrough evolutivo porta con sé, accanto ad opportunità finora inimmaginabili, il rischio di un'umanità "a due velocità", divisa tra inclusi ed esclusi non solo in termini economici, ma anche politici, culturali ed esistenziali. Se non gestita con la dovuta attenzione, questa dinamica potrebbe portare a una esasperazione delle disuguaglianze senza precedenti nella storia umana, con conseguenze potenzialmente catastrofiche per la tenuta sociale e la dignità umana [73].*

#### *Parte 5: La governance dell'IA come sfida cruciale*

*Di fronte a questi rischi, la governance dell'IA emerge come una delle sfide cruciali del nostro tempo. Per fare in modo che questo sesto breakthrough evolutivo sia un'opportunità di progresso per tutta l'umanità, e non solo per pochi privilegiati, è essenziale sviluppare quadri normativi e istituzionali adeguati a gestirne le implicazioni [74].*

*Un primo passo fondamentale è garantire che lo sviluppo dell'IA sia guidato da principi etici solidi e condivisi [75]. Questo richiede un ampio dibattito pubblico sui valori che dovrebbero informare la progettazione e l'utilizzo di queste tecnologie, come la trasparenza, la responsabilità, l'equità e il rispetto per i diritti umani [76]. Iniziative come le linee guida etiche per l'IA trustworthy sviluppate dall'Unione Europea [77] o i Principi di Asilomar per l'IA [78] rappresentano passi importanti in questa direzione, ma devono essere ulteriormente sviluppate e tradotte in pratica.*

*In secondo luogo, è necessario promuovere una maggiore inclusività nello sviluppo e nell'utilizzo dell'IA. Questo significa da un lato investire nell'educazione e nella formazione, per garantire che un numero sempre maggiore di persone abbia le competenze necessarie per partecipare attivamente all'economia e alla società dell'IA [79]. Dall'altro lato, per contrastare la concentrazione del potere e dei benefici dell'IA nelle mani di pochi attori dominanti, è necessario promuovere modelli alternativi di sviluppo e proprietà di queste tecnologie.*

*Una strada promettente in questo senso è la creazione di "AI locali", ovvero di sistemi di intelligenza artificiale progettati e gestiti da comunità territoriali o funzionali (come città, organizzazioni non profit, cooperative) per rispondere alle loro esigenze specifiche.*

*Queste AI locali dovrebbero essere interconnesse tra loro in reti distribuite e non gerarchiche, ispirate al modello delle "olarchie" teorizzato da Arthur Koestler [80]. In un'olarchia, ogni nodo della rete è al contempo un tutto autonomo e una parte di un tutto più ampio, in un'architettura ricorsiva e scalabile che bilancia l'autonomia locale con l'interdipendenza globale. Applicata all'IA, questa struttura consentirebbe di sviluppare sistemi modulari e adattivi, capaci di apprendere e cooperare tra loro senza un controllo centrale.*

*Un aspetto cruciale di queste AI locali dovrebbe essere la proprietà diffusa e partecipata delle risorse chiave che le alimentano, a partire dai dati. Anziché affidarsi alle grandi piattaforme digitali e ai loro modelli estrattivi, le comunità potrebbero organizzarsi per*

raccogliere, gestire e valorizzare in modo autonomo i propri dati, in quella che è stata chiamata una "data economy partecipativa" [81]. Attraverso meccanismi di "data sovereignty" e "data trusts" [82], i cittadini potrebbero mantenere il controllo sui propri dati personali, decidendo collettivamente se e come condividerli per scopi di interesse comune. Inoltre, le AI locali open source dovrebbero sfruttare appieno il potenziale del crowdsourcing e del crowd computing per aggregare e mettere a frutto l'intelligenza collettiva e connettiva. Attraverso piattaforme partecipative e strumenti di collaborazione, i membri della comunità potrebbero contribuire attivamente allo sviluppo, all'addestramento e alla validazione dei sistemi di IA, apportando le proprie conoscenze, preferenze e valori [83]. Questo approccio "bottom-up" consentirebbe di allineare il funzionamento dell'IA con le reali priorità e aspirazioni delle persone, e di distribuirne i benefici in modo più equo.

#### Parte 6: IA bene comune

L'obiettivo ultimo di questo modello di "IA collettiva" dovrebbe essere quello di trasformare l'intelligenza artificiale in un vero e proprio bene comune (commons), ovvero in una risorsa condivisa e autogestita dalla comunità, come proposto da Elinor Ostrom nel suo lavoro seminale sui "commons" [84]. Ciò richiederebbe di ripensare radicalmente i regimi di proprietà intellettuale e di governance dell'IA, spostandoli da una logica privatistica e monopolistica a una logica aperta e cooperativa, basata su principi di trasparenza, inclusività e co-responsabilità.

Solo attraverso questa "democratizzazione profonda" dell'IA potremo evitare che essa si trasformi in uno strumento di dominio e di esclusione, e metterla invece al servizio dell'empowerment cognitivo e sociale delle comunità. Le sfide tecniche, economiche e culturali da affrontare sono immense, ma gli esperimenti pionieristici che stanno emergendo in questa direzione - dalle cooperative di dati alle piattaforme di crowd science, dagli algoritmi trasparenti alle infrastrutture digitali comunitarie - ci mostrano che un'altra IA è possibile, e necessaria.

Naturalmente, in questo percorso di democratizzazione e condivisione delle AI un altro aspetto cruciale dovrà riguardare la governance dei dati. I dati sono il carburante che alimenta l'IA, e il modo in cui vengono raccolti, utilizzati e controllati avrà un'influenza determinante sulla distribuzione dei benefici e dei rischi di queste tecnologie [85]. È essenziale sviluppare regimi di governance dei dati che da un lato proteggano la privacy e i diritti individuali, dall'altro promuovano l'accesso e la condivisione dei dati per fini di interesse pubblico [86].

Sarà inoltre necessario ripensare i modelli di governance economica e politica per adattarli alla realtà dell'IA. Con l'automazione crescente del lavoro e la concentrazione del potere economico nelle mani di poche aziende tecnologiche, modelli come il reddito di base universale [87] o la tassazione dei robot [88] potrebbero diventare necessari per garantire una distribuzione più equa della ricchezza. Allo stesso tempo, l'influenza crescente dell'IA sui processi democratici richiede nuove forme di regolamentazione e controllo, per garantire la trasparenza e la responsabilità delle piattaforme digitali [89].

Naturalmente, nessuna di queste soluzioni è semplice o risolutiva. La governance dell'IA richiederà uno sforzo di cooperazione senza precedenti a livello globale, coinvolgendo non solo i governi e le aziende, ma anche la società civile, il mondo accademico e i cittadini [90]. Sarà un processo di apprendimento continuo, che dovrà adattarsi costantemente all'evoluzione rapida e imprevedibile di queste tecnologie [91].

Tuttavia, l'alternativa - lasciare che l'IA si sviluppi in modo incontrollato, guidata solo dalle forze del mercato e dagli interessi di pochi - è semplicemente troppo rischiosa.

Note e approfondimenti:

[1] Il concetto di "breakthrough evolutivi" nell'evoluzione dell'intelligenza è esplorato in dettaglio da Susskind (2018) nel suo libro "Future Politics: Living Together in a World Transformed by Tech". Susskind sostiene che questi breakthrough rappresentano momenti di discontinuità e accelerazione nel processo evolutivo, in cui emergono nuove capacità cognitive che aprono orizzonti inediti per l'adattamento e l'interazione con l'ambiente. Secondo l'autore, comprendere la natura e l'impatto di questi breakthrough è essenziale per cogliere la portata delle trasformazioni indotte dalle nuove tecnologie, in particolare dall'IA, e per adattare le nostre istituzioni e forme di governance alle sfide del futuro.

[2] L'idea che ogni breakthrough evolutivo abbia ampliato le capacità degli esseri viventi di interagire con l'ambiente e tra loro in modi sempre più sofisticati è al centro dell'analisi di Bostrom (2014) in "Superintelligence: Paths, Dangers, Strategies". Bostrom sottolinea come l'evoluzione dell'intelligenza sia stata caratterizzata da un progressivo aumento della complessità e della flessibilità cognitiva, che ha permesso agli organismi di affrontare sfide adattive sempre più complesse. Tuttavia, l'autore mette in guardia sul fatto che con l'avvento dell'IA, e in particolare con la possibilità di una superintelligenza artificiale, questo processo potrebbe subire un'accelerazione e una discontinuità senza precedenti, ponendo sfide esistenziali per l'umanità.

[3] La suddivisione dell'evoluzione dell'intelligenza in cinque principali breakthrough (steering, reinforcement learning, simulazione mentale, teoria della mente e linguaggio) è proposta da Bennett (2022) nel suo libro "A Brief History of Intelligence: How Minds Evolved on Earth". Bennett fornisce una dettagliata analisi di ciascuno di questi breakthrough, esplorandone le basi biologiche, le implicazioni cognitive e il ruolo nell'espansione delle capacità adattive delle specie coinvolte. L'autore sottolinea come ogni breakthrough abbia costruito sulle fondamenta dei precedenti, in un processo cumulativo di aumento della complessità e della potenza cognitiva, che trova nel linguaggio umano il suo apice provvisorio.

[4] Il ruolo del linguaggio come "rivoluzione cognitiva" che ha amplificato enormemente le capacità di pensiero astratto, comunicazione e cooperazione degli esseri umani è al centro dell'analisi di Harari (2014) in "Sapiens: A Brief History of Humankind". Harari sostiene che il linguaggio ha rappresentato un salto qualitativo nell'evoluzione della nostra specie, permettendoci di creare e condividere mondi immaginari e narrazioni collettive, che sono alla base della cooperazione su larga scala e dello sviluppo delle culture umane. Secondo l'autore, il linguaggio ha essenzialmente "hackerato" il nostro cervello, rendendolo capace di operazioni cognitive altrimenti impossibili per un primate.

[5] L'idea del linguaggio come fondamento di un'intelligenza "culturale", incarnata non solo nelle menti individuali ma anche nelle reti di simboli e artefatti che costituiscono la nostra nicchia cognitiva, è esplorata da Deacon (1997) in "The Symbolic Species: The Co-evolution of Language and the Brain". Deacon argomenta che il linguaggio non è solo uno strumento di comunicazione, ma una vera e propria infrastruttura cognitiva esterna, che ha co-evoluto con il cervello umano plasmandone la struttura e le funzioni. Secondo l'autore, è grazie a questa simbiosi tra mente e linguaggio che gli esseri umani hanno potuto sviluppare forme di intelligenza distribuite e trans-generazionali, in cui le conoscenze e le innovazioni possono accumularsi e trasmettersi ben oltre i limiti della vita individuale.

[6] Il potenziale dell'IA come "sesto breakthrough" nell'evoluzione dell'intelligenza è al centro del dibattito contemporaneo sull'impatto di queste tecnologie. Bostrom (2014), in "Superintelligence: Paths, Dangers, Strategies", argomenta che l'IA rappresenta una

*discontinuità radicale rispetto ai breakthrough precedenti, in quanto potrebbe portare alla creazione di menti artificiali che superano quelle umane in un ampio spettro di capacità cognitive. Secondo Bostrom, questo scenario pone sfide senza precedenti per il controllo e l'allineamento dell'IA con i valori umani, e richiede una riflessione urgente sulle strategie per mitigarne i rischi esistenziali.*

*[7] L'idea dell'IA come nuovo stadio nell'evoluzione dell'intelligenza, capace di estendere le capacità cognitive umane in modi radicalmente nuovi, è esplorata anche da Russell (2019) in "Human Compatible: Artificial Intelligence and the Problem of Control". Russell sottolinea come l'IA rappresenti al contempo un'enorme opportunità e una sfida esistenziale per la nostra specie. Da un lato, l'IA potrebbe permetterci di affrontare problemi finora intrattabili e di espandere i confini della nostra comprensione e azione sul mondo. Dall'altro, pone interrogativi profondi sul nostro ruolo e sul nostro futuro in un mondo sempre più plasmato da intelligenze non biologiche. Secondo Russell, assicurare che l'IA rimanga "compatibile" con i valori e gli interessi umani è la sfida definitoria del nostro tempo.*

*[8] La definizione di Intelligenza Artificiale come "capacità di sistemi tecnologici di esibire comportamenti intelligenti simili o superiori a quelli umani" è ampiamente adottata nella letteratura specialistica. Ad esempio, Kaplan (2016), in "Artificial Intelligence: What Everyone Needs to Know", definisce l'IA come "la scienza e l'ingegneria di creare macchine intelligenti, specialmente programmi software intelligenti", dove per "intelligente" si intende "in grado di funzionare appropriatamente e con previsione in un ambiente". Kaplan sottolinea come questa definizione sia volutamente ampia, e includa una varietà di approcci e tecniche, dal machine learning alle reti neurali, dalla robotica all'elaborazione del linguaggio naturale.*

*[9] L'idea dell'IA come estensione delle capacità cognitive umane resa possibile dalle tecnologie digitali è esplorata in dettaglio da Tegmark (2017) in "Life 3.0: Being Human in the Age of Artificial Intelligence". Tegmark sostiene che l'IA rappresenti l'alba di una nuova era nella storia della vita, in cui l'intelligenza si libera dai vincoli della biologia per incarnarsi in substrati tecnologici. Secondo l'autore, questo apre prospettive senza precedenti per l'espansione della nostra cognizione e della nostra azione nel mondo, ma pone anche sfide profonde per il futuro dell'umanità e il nostro posto nel cosmo. Tegmark invita a una riflessione urgente sulle implicazioni etiche e esistenziali dell'IA, e sulla necessità di indirizzarne lo sviluppo verso obiettivi allineati con i nostri valori più profondi.*

*[10] Il potenziale dell'IA di affrontare sfide cognitive che vanno oltre le capacità individuali, dalla scoperta scientifica alla gestione di sistemi complessi, è al centro dell'analisi di Agrawal, Gans e Goldfarb (2018) in "Prediction Machines: The Simple Economics of Artificial Intelligence". Gli autori sostengono che la principale caratteristica dell'IA sia la sua capacità di fare previsioni sempre più accurate a partire da grandi quantità di dati, e che questo stia già rivoluzionando settori come la medicina, la finanza, il marketing e i trasporti. Secondo gli autori, man mano che l'IA diventerà più potente e diffusa, il suo impatto si farà sentire in tutti gli ambiti della vita, aprendo opportunità senza precedenti ma anche ponendo sfide di adattamento e governance.*

*[11] L'idea che la traiettoria evolutiva dell'IA sarà largamente determinata dalle scelte e dai valori di chi la sviluppa e la controlla, sollevando interrogativi cruciali sulla sua governance e il suo impatto sulla società, è al centro dell'analisi di Bostrom (2014) in "Superintelligence: Paths, Dangers, Strategies". Bostrom sottolinea come, a differenza dei breakthrough evolutivi precedenti, l'IA sia un prodotto dell'ingegno umano, e come tale rifletta inevitabilmente gli obiettivi, le assunzioni e i bias di chi la crea. Secondo l'autore, questo pone una grande responsabilità sulle spalle di ricercatori, sviluppatori e policy-maker, che*

devono assicurare che l'IA sia allineata con i valori e gli interessi dell'umanità nel suo complesso, e non solo con quelli di pochi attori potenti.

[12] Il concetto di "esplosione di intelligenza", ovvero la possibilità che l'IA possa innescare un ciclo di auto-miglioramento esponenziale che porti rapidamente a livelli di intelligenza superintelligente, è esplorato in dettaglio da Bostrom (2014) in "Superintelligence: Paths, Dangers, Strategies". Bostrom argomenta che, se riuscissimo a creare un'IA in grado di ricercare e sviluppare autonomamente tecnologie sempre più potenti, questa potrebbe rapidamente superare le capacità cognitive umane in tutti i domini, innescando una trasformazione radicale della civiltà. Secondo l'autore, questo scenario pone sfide senza precedenti per il controllo e l'allineamento dell'IA, e richiede una riflessione urgente sulle strategie per mitigarne i rischi esistenziali.

[13] Il rischio di perdere il controllo su sistemi di IA sempre più potenti e autonomi è al centro dell'analisi di Russell (2019) in "Human Compatible: Artificial Intelligence and the Problem of Control". Russell sostiene che, man mano che l'IA diventa più capace e diffusa, il problema di assicurarne il controllo e l'allineamento con i valori umani diventa sempre più pressante. Secondo l'autore, se non affrontiamo questo problema in modo proattivo, rischiamo di trovarci in un futuro in cui le decisioni più importanti per il destino dell'umanità sono prese da sistemi il cui funzionamento potremmo non essere più in grado di capire o influenzare. Russell propone una serie di principi e strategie per mantenere l'IA "compatibile" con gli interessi umani, tra cui l'incorporazione di valori e preferenze umane nei sistemi di IA fin dalla loro progettazione.

[14] L'idea che il linguaggio abbia permesso la trasmissione e l'accumulo di conoscenze su scale spaziali e temporali molto più ampie di quelle consentite dalla biologia è esplorata in dettaglio da Deacon (1997) in "The Symbolic Species: The Co-evolution of Language and the Brain". Deacon sostiene che il linguaggio rappresenti una vera e propria "esternalizzazione" della memoria e della cognizione umana, che permette di preservare e trasmettere il sapere attraverso le generazioni e le culture. Secondo l'autore, questa capacità di "outsourcing cognitivo" ha giocato un ruolo fondamentale nell'evoluzione della nostra specie, permettendoci di accumulare conoscenze e innovazioni ben oltre i limiti della vita individuale, e di costruire nicchie culturali sempre più complesse e adattive.

[15] Il ruolo della scrittura nell'esternalizzazione della memoria e nella preservazione del sapere oltre i limiti della vita individuale è al centro dell'analisi di Ong (1982) in "Orality and Literacy: The Technologizing of the Word". Ong argomenta che la scrittura rappresenti una tecnologia cognitiva rivoluzionaria, che ha trasformato profondamente il nostro rapporto con il linguaggio, il pensiero e la conoscenza. Secondo l'autore, la scrittura ha permesso di "fissare" il sapere in forme durevoli e trasportabili, rendendolo accessibile attraverso il tempo e lo spazio, e aprendo la strada a nuove forme di riflessione, analisi e critica. Ong sottolinea come la scrittura abbia giocato un ruolo fondamentale nello sviluppo della filosofia, della scienza e della storia come le conosciamo.

[16] L'idea dell'umanità "aumentata" dal linguaggio e dalla tecnologia, che trascende i vincoli della propria natura biologica per creare un nuovo regno di realtà culturale, è esplorata in dettaglio da Stiegler (1998) in "Technics and Time, 1: The Fault of Epimetheus". Stiegler sostiene che la tecnica, intesa come la capacità di esteriorizzare le funzioni biologiche in strumenti e artefatti, sia la caratteristica distintiva dell'umano, che lo distingue dagli altri animali. Secondo l'autore, è attraverso la tecnica che l'umanità si emancipa dai limiti della propria natura, creando un mondo "artificiale" di cultura e significato che si sovrappone e retroagisce sul mondo naturale. Stiegler vede nel linguaggio e nella scrittura le tecniche

*fondamentali che hanno permesso questo processo di esteriorizzazione e accumulazione della memoria collettiva.*

*[17] L'idea dell'umanità come forza evolutiva capace di plasmare il proprio ambiente e il proprio futuro attraverso l'innovazione tecnologica e culturale è al centro dell'analisi di Dennett (2017) in "From Bacteria to Bach and Back: The Evolution of Minds". Dennett sostiene che, con l'avvento del linguaggio e della cultura, l'evoluzione umana abbia subito una svolta radicale, passando da un processo di selezione naturale "cieca" a un processo di selezione culturale "diretta", in cui le idee e le innovazioni vengono trasmesse, modificate e selezionate in base alla loro utilità e al loro valore per la società. Secondo l'autore, questo ha portato all'emergere di una nuova forma di evoluzione, basata sulla trasmissione e l'accumulo di informazioni culturali, che si affianca e si intreccia con l'evoluzione biologica.*

*[18] L'idea della storia della tecnologia come un continuo processo di "aumento" delle capacità umane è esplorata in dettaglio da Bostrom (2014) in "Superintelligence: Paths, Dangers, Strategies". Bostrom traccia una panoramica dell'evoluzione tecnologica, mostrando come ogni innovazione, dalla ruota al computer, abbia esteso e ridefinito i confini dell'azione e della cognizione umana. Secondo l'autore, questo processo di aumento ha subito un'accelerazione esponenziale negli ultimi secoli, man mano che le tecnologie sono diventate sempre più potenti e interconnesse. Bostrom sostiene che l'IA rappresenti il punto culminante di questo processo, promettendo di amplificare le nostre capacità cognitive in modi che sfidano l'immaginazione.*

*[19] L'idea che le innovazioni tecnologiche abbiano progressivamente ampliato la nostra "nicchia cognitiva" ben oltre i limiti della biologia è al centro dell'analisi di Clark (2003) in "Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence". Clark sostiene che gli esseri umani siano "cyborg naturali", ovvero creature che si sono co-evolute con le proprie tecnologie cognitive, integrandole profondamente nel proprio funzionamento mentale. Secondo l'autore, strumenti come il linguaggio, la scrittura, i numeri e i computer non sono semplici accessori, ma vere e proprie estensioni del nostro sistema cognitivo, che ridefiniscono i confini del sé e della mente. Clark argomenta che questa simbiosi tra mente e tecnologia sia la chiave per comprendere l'evoluzione passata e futura dell'intelligenza umana.*

*[20] L'idea che molte tecnologie agiscano come estensioni dei nostri sensi, permettendoci di percepire e interagire con la realtà in modi prima impensabili, è esplorata in dettaglio da Ihde (1990) in "Technology and the Lifeworld: From Garden to Earth". Ihde sviluppa una "fenomenologia delle relazioni umano-tecnologia", analizzando come diversi strumenti, dagli occhiali al telescopio, dal telefono al computer, medino e trasformino la nostra esperienza del mondo. Secondo l'autore, queste tecnologie non sono neutrali, ma incorporano specifiche "intenzionalità" e "affordance" che plasmano il nostro modo di percepire, pensare e agire. Ihde sottolinea come la crescente complessità e pervasività delle tecnologie richieda una riflessione critica sul loro impatto sul nostro "mondo-della-vita".*

*[21] L'idea che le estensioni sensoriali offerte dalla tecnologia non solo amplino la nostra presa sul mondo, ma anche trasformino profondamente il nostro modo di concepirlo e di rapportarci ad esso, è al centro dell'analisi di Verbeek (2005) in "What Things Do: Philosophical Reflections on Technology, Agency, and Design". Verbeek sostiene che le tecnologie non siano semplici strumenti, ma veri e propri "agenti" che co-determinano il nostro mondo e la nostra identità. Secondo l'autore, le tecnologie incorporano specifici "script" e "mediazioni" che influenzano il nostro modo di percepire, interpretare e valutare la realtà, spesso in modi impercettibili ma profondi. Verbeek invita a una "etica della*

progettazione" che riconosca e assuma consapevolmente questa co-determinazione tra umani e tecnologie.

[22] Il concetto di "bricolage evolutivo", ovvero l'idea che l'evoluzione proceda riutilizzando e ricombinando strutture preesistenti per nuove funzioni, è formulato da Jacob (1977) nell'articolo "Evolution and Tinkering". Jacob argomenta che l'evoluzione non sia un processo di ingegneria ottimale, ma piuttosto un "lavoro di bricolage", che assembla soluzioni contingenti a partire dai materiali disponibili. Secondo l'autore, questo spiega perché gli organismi spesso presentano strutture "imperfette" o "ridondanti", frutto di adattamenti progressivi a partire da forme precedenti. Jacob sostiene che questo principio si applichi non solo all'evoluzione biologica, ma anche a quella culturale e tecnologica, in cui le nuove innovazioni si costruiscono sempre a partire dalle conoscenze e dalle pratiche ereditate dal passato.

[23] L'idea che l'invenzione della scrittura sia troppo recente perché si siano evolute strutture cerebrali ad essa dedicate, ma che l'esposizione precoce alla scrittura porti al "riciclaggio" di aree cerebrali preesistenti per il riconoscimento dei caratteri, è sviluppata in dettaglio da Dehaene (2009) in "Reading in the Brain: The New Science of How We Read". Dehaene mostra come l'apprendimento della lettura sfrutti la plasticità della corteccia visiva, "specializzando" neuroni inizialmente dedicati al riconoscimento di forme per il processamento rapido dei grafemi. Questo esempio illustra il principio della "riconversione neuronale", ovvero la capacità del nostro cervello di riutilizzare strutture antiche per funzioni culturalmente nuove.

[24] L'idea che le tecnologie digitali, combinando stimoli visivi, uditivi e tattili, rafforzino l'integrazione multisensoriale nel nostro cervello è esplorata da Spence (2020) nell'articolo "Senses of place: architectural design for the multisensory mind". Spence mostra come dispositivi come smartphone e tablet, fornendo un'esperienza "immersiva" e multisensoriale, promuovano una maggiore connettività funzionale tra aree cerebrali dedicate a modalità percettive diverse. Questa "fusione" dei sensi, argomenta l'autore, ha profonde implicazioni per il nostro modo di percepire e interagire con gli ambienti digitali.

[25] L'importanza di comprendere le dinamiche di "riuso" e "plasticità" neurale per cogliere l'impatto delle tecnologie sul cervello e progettare interfacce uomo-macchina ottimali è sottolineata da Dehaene (2020) nell'articolo "How We Learn: The New Science of Education and the Brain". Dehaene invita a una maggiore collaborazione tra neuroscienze, scienze cognitive e interaction design per sviluppare tecnologie educative e interfacce che sfruttino al meglio i meccanismi di apprendimento e adattamento del nostro cervello.

[26] L'idea dell'IA come radicale estensione delle nostre facoltà cognitive, che ridefinisce i confini di ciò che è umanamente possibile, è esplorata da Bostrom (2014) in "Superintelligence: Paths, Dangers, Strategies". Bostrom argomenta che l'IA rappresenti una discontinuità senza precedenti nella storia dell'intelligenza, paragonabile per impatto solo all'emergere del linguaggio e della cultura. Secondo l'autore, l'IA non solo promette di amplificare enormemente le nostre capacità di ragionamento, apprendimento e problem-solving, ma anche di renderle "scalabili" e "generalizzabili" in modi che sfidano i limiti della cognizione biologica.

[27] L'idea che l'IA prometta di amplificare le nostre capacità di apprendimento, ragionamento e problem-solving oltre i limiti della mente individuale è esplorata in dettaglio da Domingos (2015) in "The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World". Domingos sostiene che l'IA, e in particolare il machine learning, rappresentino un nuovo paradigma di scoperta e problem-solving, in cui le macchine apprendono autonomamente dai dati, estraendo pattern e regole che sfuggono

*all'intuizione umana. Secondo l'autore, questo approccio promette di rivoluzionare tutti i campi del sapere, dalla scienza alla medicina, dall'ingegneria all'economia, permettendoci di affrontare problemi finora intrattabili per la mente umana.*

*[28] L'idea che l'IA possa estendere la nostra "visione" in domini sempre più vasti e complessi, permettendoci di scorgere pattern e connessioni prima inaccessibili, è al centro dell'analisi di Agrawal, Gans e Goldfarb (2018) in "Prediction Machines: The Simple Economics of Artificial Intelligence". Gli autori sostengono che il tratto distintivo dell'IA sia la sua capacità di fare previsioni sempre più accurate a partire da grandi quantità di dati, e che questo stia già trasformando settori come la medicina, la finanza, il marketing e i trasporti. Secondo gli autori, man mano che l'IA diventerà più potente e diffusa, la sua capacità di "vedere" pattern e relazioni nascoste avrà un impatto dirompente su tutti gli ambiti della vita, dal lavoro alla scienza, dalla politica alla cultura.*

*[29] L'idea dell'IA come continuazione e accelerazione del processo di trascendimento dei limiti biologici che caratterizza l'umanità è esplorata da Harari (2017) in "Homo Deus: A Brief History of Tomorrow". Harari sostiene che, dopo aver raggiunto il dominio sul mondo naturale grazie alla scienza e alla tecnologia, l'umanità sia ora diretta verso un'agenda di potenziamento e superamento dei propri limiti biologici, in cui l'IA gioca un ruolo centrale. Secondo l'autore, l'IA non solo promette di amplificare enormemente le nostre capacità cognitive, ma anche di ridefinire la natura stessa dell'intelligenza e della coscienza, aprendo la strada a forme di "vita inorganica" potenzialmente superiori a quelle biologiche.*

*[30] L'idea che l'IA, così come il linguaggio e la scrittura, possa comportare profondi cambiamenti nella struttura della società e nella distribuzione del potere è al centro dell'analisi di Susskind (2018) in "Future Politics: Living Together in a World Transformed by Tech". Susskind sostiene che le tecnologie digitali, e in particolare l'IA, stiano ridefinendo le condizioni base della convivenza sociale e politica, alterando gli equilibri di potere tra individui, gruppi e istituzioni. Secondo l'autore, questo richiede un ripensamento radicale delle nostre categorie politiche tradizionali, e lo sviluppo di nuove forme di governance e cooperazione per gestire le sfide poste da queste tecnologie.*

*[31] L'idea che i breakthrough evolutivi precedenti abbiano seguito traiettorie relativamente lineari, costruendo l'uno sull'altro in un processo cumulativo di espansione delle capacità cognitive, è esplorata da Pinker (2018) in "Enlightenment Now: The Case for Reason, Science, Humanism, and Progress". Pinker sostiene che la storia dell'intelligenza sia caratterizzata da un progresso incrementale, in cui ogni innovazione cognitiva apre la strada a quelle successive, in un processo di feedback positivo. Secondo l'autore, questo processo ha subito un'accelerazione esponenziale negli ultimi secoli, grazie alla diffusione del pensiero scientifico e dei valori illuministi, che hanno liberato il potenziale creativo dell'umanità. Pinker vede nell'IA una continuazione di questo trend, ma invita anche a considerarne i rischi e le sfide etiche.*

*[32] L'idea che i breakthrough biologici siano il risultato di mutazioni genetiche casuali sottoposte al vaglio della selezione naturale è al centro della teoria dell'evoluzione per selezione naturale formulata da Darwin (1859) in "On the Origin of Species". Darwin argomenta che le specie evolvano attraverso un processo di variazione casuale e selezione non casuale, in cui le mutazioni vantaggiose vengono trasmesse alle generazioni successive con maggiore frequenza di quelle svantaggiose. Secondo l'autore, questo meccanismo spiega l'emergere di strutture e comportamenti complessi senza bisogno di ricorrere a un disegno intelligente. La teoria di Darwin ha rivoluzionato la nostra comprensione della vita e del nostro posto nella natura, e rimane il fondamento concettuale della biologia moderna.*

[33] L'idea dell'IA come artefatto progettato intenzionalmente dall'uomo per svolgere compiti specifici è esplorata da Hofstadter (1979) in "Gödel, Escher, Bach: An Eternal Golden Braid". Hofstadter sostiene che l'intelligenza, naturale o artificiale, sia sempre il prodotto di un processo di "ingegneria epistemica", in cui sistemi formali vengono costruiti e manipolati per rappresentare e ragionare sul mondo. Secondo l'autore, l'IA è essenzialmente un tentativo di replicare questo processo in modo artificiale, definendo esplicitamente le regole e le strutture che sottendono il pensiero. Hofstadter esplora i paradossi e i limiti di questo approccio, e sostiene che una vera IA dovrebbe essere capace di "auto-riferimento" e "creatività", sfuggendo alle regole precostituite.

[34] L'idea che l'evoluzione dell'IA sia guidata dagli obiettivi e dai valori di chi la sviluppa, e non dalle cieche logiche dell'adattamento evolutivo, è al centro dell'analisi di Bostrom (2014) in "Superintelligence: Paths, Dangers, Strategies". Bostrom sottolinea come, a differenza dell'evoluzione biologica, l'evoluzione dell'IA sia un processo "direzionale", influenzato dalle scelte deliberate dei suoi progettisti. Secondo l'autore, questo pone una grande responsabilità etica su chi sviluppa queste tecnologie, che devono assicurarsi che gli obiettivi e i valori dell'IA siano allineati con quelli dell'umanità.

[35] L'idea dell'evoluzione biologica come processo "senza progettista" è formulata da Dawkins (1986) in "The Blind Watchmaker". Dawkins argomenta che la complessità e la funzionalità degli organismi viventi non richiedano un disegno intelligente, ma possano emergere da un processo di selezione naturale "cieco" e non direzionale. Secondo l'autore, l'evoluzione opera attraverso la riproduzione differenziale di varianti casuali, senza alcuna previsione o pianificazione. Dawkins usa la metafora dell'orologiaio cieco per illustrare come questo processo possa produrre strutture complesse e "ingegneristiche" senza alcun progetto o intenzione.

[36] L'idea che l'IA abbia sempre dei progettisti, sviluppatori e proprietari che ne determinano le finalità e le modalità di funzionamento è esplorata da Bostrom (2014) in "Superintelligence: Paths, Dangers, Strategies". Bostrom sottolinea come, a differenza degli organismi biologici, i sistemi di IA siano sempre creati per scopi specifici, che riflettono gli interessi e i valori di chi li sviluppa. Secondo l'autore, questo solleva interrogativi cruciali su come garantire che questi scopi siano allineati con il bene comune, e non solo con gli obiettivi di pochi attori potenti.

[37] L'idea che sia necessario garantire che lo sviluppo dell'IA sia allineato con i valori e gli interessi della società nel suo complesso, e non solo con quelli di pochi attori economici o politici, è al centro dell'analisi di Yudkowsky (2008) in "Artificial Intelligence as a Positive and Negative Factor in Global Risk". Yudkowsky sostiene che, data la potenza trasformativa dell'IA, lasciarne lo sviluppo nelle mani di pochi attori mossi da interessi particolari rappresenti un rischio esistenziale per l'umanità. Secondo l'autore, è essenziale che la ricerca sull'IA sia guidata da un'etica della responsabilità e del bene comune, e che i suoi benefici siano distribuiti in modo equo.

[38] L'idea che l'IA stia progredendo a una velocità senza precedenti grazie agli avanzamenti nella potenza di calcolo, nella disponibilità di dati e nelle tecniche di apprendimento automatico è esplorata da Bostrom (2014) in "Superintelligence: Paths, Dangers, Strategies". Bostrom analizza le diverse traiettorie di sviluppo dell'IA, dalla "emulazione whole brain" alla "IA simbolica", e sostiene che tutte stiano beneficiando di una crescita esponenziale delle risorse computazionali e informative. Secondo l'autore, questo rende plausibile il raggiungimento di una "IA generale" (AGI) o addirittura "superintelligente" (ASI) in tempi molto più brevi di quanto comunemente ipotizzato.

[39] L'idea che l'IA stia progredendo a un ritmo tale da raddoppiare le sue capacità ogni pochi mesi o anni è esplorata da Grace et al. (2018) nell'articolo "When Will AI Exceed Human Performance? Evidence from AI Experts". Gli autori hanno condotto un sondaggio tra esperti di IA sulle loro previsioni riguardo al raggiungimento della "superiorità" dell'IA in vari compiti cognitivi. I risultati mostrano che la maggioranza degli esperti ritiene probabile che l'IA superi le prestazioni umane in molti domini entro i prossimi decenni, con alcune stime che vanno dai 10 ai 40 anni a seconda del compito.

[40] L'idea che il ritmo di sviluppo dell'IA sia ordini di grandezza più rapido di quello dell'evoluzione biologica è esplorata da Christian (2018) in "Origin Story: A Big History of Everything". Christian mette a confronto le scale temporali dell'evoluzione biologica e di quella tecnologica, mostrando come la prima proceda su tempi di milioni o miliardi di anni, mentre la seconda operi su scale di decenni o secoli. Secondo l'autore, questo divario si sta ampliando esponenzialmente con l'accelerazione del progresso tecnologico, e in particolare con lo sviluppo dell'IA. Christian sostiene che stiamo entrando in una nuova fase della "Big History", in cui l'evoluzione intelligente, guidata dalla tecnologia, prende il sopravvento su quella biologica.

[41] L'idea che la rapidità e la radicalità delle trasformazioni indotte dall'IA rendano difficile per le nostre istituzioni, norme e comprensione etica stare al passo è al centro dell'analisi di Wallach e Allen (2008) in "Moral Machines: Teaching Robots Right from Wrong". Gli autori sostengono che l'IA ponga sfide senza precedenti al nostro sistema etico e normativo, che è stato sviluppato in relazione ad agenti morali umani. Secondo gli autori, l'emergere di "macchine morali" autonome richiede un ripensamento delle nostre categorie etiche tradizionali, e lo sviluppo di nuovi approcci per "insegnare" alle IA i principi del giusto e dello sbagliato.

[42] L'idea che l'IA ci ponga di fronte a cambiamenti così rapidi e radicali da mettere in discussione la nostra stessa capacità di controllarli e gestirli è esplorata da Bostrom (2014) in "Superintelligence: Paths, Dangers, Strategies". Bostrom argomenta che, a differenza delle rivoluzioni tecnologiche precedenti, l'avvento dell'IA potrebbe rappresentare un punto di non ritorno per l'umanità, oltre il quale potremmo perdere il controllo del nostro destino. Secondo l'autore, se non sapremo sviluppare in tempo strategie efficaci per allineare gli obiettivi delle IA superintelligenti con i nostri, rischiamo di trovarci in un futuro in cui le decisioni fondamentali non saranno più nelle nostre mani.

[43] L'idea che i breakthrough evolutivi precedenti abbiano ampliato le capacità delle specie coinvolte, mantenendo però un "tetto" definito dalle possibilità biologiche, è esplorata da Pinker (1997) in "How the Mind Works". Pinker argomenta che l'evoluzione dell'intelligenza animale sia stata plasmata dai vincoli e dalle opportunità offerti dalla nicchia ecologica di ciascuna specie. Secondo l'autore, nonostante l'enorme diversità di forme e funzioni cognitive nel regno animale, esiste un limite alle capacità di elaborazione dell'informazione di un sistema biologico, dato dai compromessi tra costi energetici, tempi di sviluppo e vantaggi adattivi. Pinker sostiene che l'intelligenza umana rappresenti il punto più alto finora raggiunto da questo processo, spingendo ai limiti la potenza computazionale di un cervello biologico.

[44] L'idea che l'IA rischi di creare una profonda divaricazione tra chi ha accesso a queste tecnologie e chi ne è escluso è al centro dell'analisi di Brynjolfsson e McAfee (2014) in "The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies". Gli autori sostengono che l'IA e l'automazione stiano già contribuendo ad ampliare le disuguaglianze economiche, polarizzando il mercato del lavoro tra pochi "vincitori" dotati di competenze complementari alle macchine, e molti "perdenti" i cui compiti sono sostituibili dall'IA.

[45] Il concetto di "società in rete" e l'idea che l'accesso alle reti informazionali sia la principale fonte di potere e opportunità nel mondo contemporaneo sono sviluppati da Castells (1996) in "The Rise of the Network Society". Castells analizza le trasformazioni economiche, sociali e culturali indotte dalla rivoluzione digitale, e sostiene che stiamo assistendo all'emergere di un nuovo paradigma socio-tecnologico, basato sulla centralità delle reti e dei flussi di informazione. Secondo l'autore, in questo contesto, il potere non risiede più nelle istituzioni stabili e gerarchiche della società industriale, ma nella capacità di accedere e controllare le reti globali di comunicazione, finanza e conoscenza.

[46] L'idea che la società in rete sia caratterizzata dalla centralità delle reti informazionali e dalla loro logica di flussi è approfondita da Castells (1996) in "The Rise of the Network Society". Castells sostiene che le reti digitali, grazie alla loro flessibilità, adattabilità e scalabilità, stiano diventando la forma organizzativa dominante in tutti i campi dell'attività umana, dalla produzione alla politica, dalla cultura alla vita quotidiana. Secondo l'autore, queste reti funzionano secondo una logica di flussi, ovvero di scambi continui e in tempo reale di informazioni, capitali, simboli e immagini su scala globale.

[47] L'idea che nella società in rete l'accesso alle reti e la capacità di operare al loro interno siano le principali fonti di potere e opportunità è sviluppata da Castells (1996) in "The Rise of the Network Society". Castells sostiene che, nell'economia informazionale, il valore non risieda più nella produzione di beni materiali, ma nella generazione e manipolazione di conoscenza e informazione. Secondo l'autore, ciò crea nuove forme di stratificazione sociale, basate non più sulla proprietà dei mezzi di produzione, ma sul controllo dei mezzi di comunicazione e di elaborazione delle informazioni.

[48] Il concetto di dualismo tra "inclusi" ed "esclusi" dalle reti informazionali è introdotto da Castells (1998) in "End of Millennium". Castells sostiene che la logica di inclusione/esclusione sia una caratteristica strutturale della società in rete, che crea una divisione fondamentale tra coloro che sono "in" e coloro che sono "out". Secondo l'autore, questa divisione non segue più le tradizionali linee di classe o di nazione, ma dipende dalla posizione degli individui e dei gruppi rispetto alle reti globali di creazione di valore e di potere.

[49] L'idea che gli "inclusi" nella società in rete siano coloro che hanno le risorse, le competenze e le connessioni per parteciparvi pienamente e sfruttarne i benefici è sviluppata da Castells (1996) in "The Rise of the Network Society". Castells sostiene che l'inclusione nelle reti richieda un insieme di prerequisiti materiali, educativi e sociali, che sono distribuiti in modo fortemente diseguale. Secondo l'autore, solo coloro che dispongono di accesso alle infrastrutture digitali, di elevati livelli di istruzione e competenze, e di legami con i nodi chiave delle reti globali possono davvero prosperare nell'economia informazionale.

[50] L'idea che gli "esclusi" dalla società in rete siano coloro che, per varie ragioni, rimangono ai margini delle reti informazionali, subendone gli effetti senza poterne determinare il funzionamento, è approfondita da Castells (1998) in "End of Millennium". Castells identifica diverse categorie di esclusi, dai lavoratori generici le cui competenze sono rese ridondanti dall'automazione, alle regioni e ai paesi che non riescono ad agganciarsi ai flussi globali di ricchezza e innovazione, fino ai gruppi sociali discriminati per ragioni etniche, religiose o di genere.

[51] L'idea che la dinamica di inclusione/esclusione nella società in rete generi un'umanità "a due velocità", divisa tra uno spazio dei flussi accelerato e uno spazio dei luoghi marginale, è sviluppata da Castells (1996) in "The Rise of the Network Society". Castells sostiene che la logica delle reti crei una nuova geografia sociale, in cui il tempo e lo spazio sono sempre più dissociati. Secondo l'autore, gli inclusi vivono in uno "spazio dei flussi" caratterizzato da

*mobilità, istantaneità e globalità, in cui le distanze si annullano e le opportunità si moltiplicano. Gli esclusi, invece, sono confinati in uno "spazio dei luoghi" caratterizzato da localismo, lentezza e frammentazione, in cui le distanze si accentuano e le opportunità si riducono.*

*[52] L'idea che gli inclusi nella società in rete possano accumulare sempre più potere e ricchezza sfruttando le opportunità delle reti globali, mentre gli esclusi vedano le proprie condizioni deteriorarsi, è approfondita da Castells (1998) in "End of Millennium". Castells sostiene che la logica di inclusione/esclusione crei un circolo vizioso di polarizzazione sociale ed economica. Secondo l'autore, coloro che hanno accesso alle reti possono usarle per moltiplicare le proprie opportunità di apprendimento, lavoro e connessione, accumulando vantaggi competitivi che li allontanano sempre più dal resto della popolazione. Gli esclusi, invece, si trovano intrappolati in una spirale di impoverimento e marginalizzazione, in quanto la loro esclusione dalle reti li priva delle risorse necessarie per migliorare la propria condizione.*

*[53] L'idea che l'IA stia diventando sempre più importante in tutti i settori della società, dall'economia alla politica, dalla cultura alla scienza, è al centro dell'analisi di Brynjolfsson e McAfee (2014) in "The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies". Gli autori sostengono che l'IA rappresenti una "tecnologia di scopo generale", ovvero un'innovazione che ha il potenziale di rivoluzionare praticamente ogni aspetto della vita umana. Secondo Brynjolfsson e McAfee, l'IA sta già trasformando il modo in cui produciamo, consumiamo, comunichiamo e prendiamo decisioni, e il suo impatto è destinato ad accelerare nei prossimi anni.*

*[54] L'idea che l'accesso alle tecnologie di IA diventerà sempre più determinante per le opportunità di vita degli individui e dei gruppi è sviluppata da West (2018) in "The Future of Work: Robots, AI, and Automation". West sostiene che, man mano che l'IA pervaderà ogni settore economico e sociale, la capacità di utilizzarla e beneficiarne diventerà un fattore cruciale di successo e mobilità. Secondo l'autore, coloro che avranno le competenze, le risorse e le connessioni per sfruttare appieno il potenziale dell'IA godranno di enormi vantaggi in termini di opportunità di lavoro, reddito e influenza. Al contrario, coloro che ne saranno esclusi rischiano di vedere le proprie prospettive di vita drasticamente ridotte, in un mercato del lavoro sempre più polarizzato tra compiti ad alta e bassa qualificazione.*

*[55] L'idea che coloro che hanno la possibilità di sviluppare e utilizzare sistemi di IA avanzati godranno di enormi vantaggi competitivi è esplorata da Susskind e Susskind (2015) in "The Future of the Professions: How Technology Will Transform the Work of Human Experts". Gli autori sostengono che l'IA e le tecnologie digitali stiano già rivoluzionando il mondo delle professioni, dall'avvocatura alla medicina, dalla consulenza all'istruzione. Secondo Susskind e Susskind, grazie alla loro capacità di processare ed analizzare enormi quantità di dati, di apprendere dall'esperienza e di fornire soluzioni personalizzate, i sistemi di IA sono in grado di svolgere molti compiti professionali in modo più efficiente, economico e accessibile degli esperti umani. Gli autori prevedono che questo porterà ad una radicale ridefinizione del lavoro professionale, con la scomparsa di molte occupazioni tradizionali e l'emergere di nuovi ruoli ibridi tra competenze umane e intelligenza artificiale. In questo contesto, coloro che sapranno sviluppare e controllare i sistemi di IA più avanzati avranno un enorme vantaggio competitivo.*

*[56] L'idea che le risorse necessarie per sfruttare appieno il potenziale dell'IA siano distribuite in modo estremamente diseguale è al centro dell'analisi di Brynjolfsson e McAfee (2014) in "The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies". Gli autori sostengono che lo sviluppo e l'utilizzo dell'IA richiedano un insieme*

di prerequisiti - come l'accesso a grandi quantità di dati, a potenti infrastrutture di calcolo, a competenze tecniche avanzate e a ingenti capitali - che sono concentrati in poche aree geografiche e in poche organizzazioni. Secondo Brynjolfsson e McAfee, questo crea un enorme divario tra chi ha la possibilità di beneficiare della rivoluzione dell'IA e chi ne è escluso. Gli autori parlano di una nuova "divisione digitale", non più basata solo sull'accesso a Internet, ma sulla capacità di partecipare attivamente all'economia dell'IA.

[57] L'idea che poche grandi aziende tecnologiche, principalmente statunitensi e cinesi, dominino il settore dell'IA grazie alla loro capacità di attrarre talenti, dati e capitali è sviluppata da Lee (2018) in "AI Superpowers: China, Silicon Valley, and the New World Order". Lee, un esperto di IA con esperienza sia nella Silicon Valley che in Cina, sostiene che la competizione globale per la leadership nell'IA si stia concentrando attorno a un pugno di giganti tecnologici, come Google, Amazon, Facebook, Apple e Microsoft negli Stati Uniti, e Baidu, Alibaba, Tencent e Huawei in Cina. Secondo Lee, grazie alle loro vaste risorse finanziarie, alle loro enormi basi di utenti e alla loro capacità di attirare i migliori talenti, queste aziende sono in grado di dominare l'intero ciclo di sviluppo dell'IA, dalla ricerca di base all'applicazione commerciale su larga scala.

[58] L'idea che molti paesi, soprattutto nel Sud globale, faticino ad accedere anche alle forme più basilari di IA a causa della mancanza di infrastrutture digitali, competenze e risorse è esplorata da Smith (2018) in "The Future Computed: Artificial Intelligence and its role in society". Smith, presidente di Microsoft, sostiene che nonostante l'IA abbia il potenziale per affrontare molte delle sfide più urgenti del nostro tempo, dal cambiamento climatico alla povertà, i suoi benefici rischiano di essere distribuiti in modo molto diseguale a livello globale. Secondo Smith, mentre i paesi più avanzati, in particolare Stati Uniti e Cina, stanno investendo massicciamente nello sviluppo e nell'adozione dell'IA, molti paesi in via di sviluppo faticano ad accedere anche alle forme più basilari di queste tecnologie, a causa di carenze nelle infrastrutture digitali, nei sistemi educativi e nei finanziamenti.

[59] L'idea che l'IA possa esacerbare la polarizzazione del mercato del lavoro tra pochi lavoratori altamente qualificati e molti lavoratori dequalificati o disoccupati è al centro dell'analisi di Acemoglu e Restrepo (2018) in "Artificial Intelligence, Automation and Work". Gli autori, utilizzando un modello economico, mostrano come l'adozione dell'IA possa avere effetti molto diversi sulle prospettive occupazionali e salariali dei lavoratori a seconda delle loro competenze e mansioni. Secondo Acemoglu e Restrepo, i lavoratori le cui competenze sono complementari all'IA, come i programmatori, i data scientist e i manager, vedranno aumentare la domanda per il loro lavoro e i loro salari. Al contrario, i lavoratori le cui mansioni sono sostituibili dall'IA, come gli impiegati amministrativi, gli operai e i lavoratori dei servizi, sperimenteranno una riduzione delle opportunità lavorative e una pressione al ribasso sui salari.

[60] L'idea che l'IA possa rafforzare il dominio tecnologico ed economico di pochi paesi avanzati, lasciando indietro gran parte del mondo in via di sviluppo, è sviluppata da Lee (2018) in "AI Superpowers: China, Silicon Valley, and the New World Order". Lee sostiene che la competizione per la leadership nell'IA rappresenti una nuova forma di "corsa agli armamenti" tecnologica, paragonabile per importanza strategica alla corsa spaziale tra Stati Uniti e Unione Sovietica durante la Guerra Fredda. Secondo Lee, i paesi che riusciranno a dominare le tecnologie di IA godranno di enormi vantaggi non solo economici, ma anche geopolitici e militari. L'autore prevede che la competizione per l'IA porterà a un mondo sempre più bipolare, dominato da Stati Uniti e Cina, in cui gli altri paesi rischiano di diventare "Stati vassalli digitali", dipendenti dalle piattaforme e dalle infrastrutture delle superpotenze tecnologiche.

[61] L'idea che l'IA stia diventando sempre più centrale nei processi politici, nella formazione dell'opinione pubblica e nella distribuzione di contenuti culturali è al centro dell'analisi di Zuboff (2019) in "The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power". Zuboff sostiene che le grandi aziende tecnologiche, grazie alla loro capacità di raccogliere ed analizzare enormi quantità di dati sulle preferenze e i comportamenti individuali, stiano acquisendo un potere senza precedenti di influenzare le scelte politiche, culturali e di consumo dei cittadini. Secondo l'autrice, queste aziende utilizzano algoritmi di IA sempre più sofisticati per profilare gli utenti, prevederne i comportamenti e personalizzare i contenuti che vengono loro mostrati, creando "bolle di filtraggio" che limitano l'esposizione a informazioni e opinioni diverse.

[62] L'idea che coloro che controllano le piattaforme di IA e i dati che le alimentano avranno un enorme potere di plasmare il discorso pubblico, le preferenze politiche e le visioni del mondo è sviluppata da O'Neil (2016) in "Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy". O'Neil, una matematica ed ex analista finanziaria, sostiene che gli algoritmi di IA utilizzati dalle piattaforme digitali per selezionare e raccomandare contenuti siano tutt'altro che neutri e oggettivi. Secondo l'autrice, questi algoritmi incorporano inevitabilmente i pregiudizi, le priorità e gli interessi delle aziende e dei programmatori che li creano, e hanno il potere di influenzare in modo opaco ma pervasivo le opinioni e i comportamenti degli utenti.

[63] L'idea che l'IA sollevi interrogativi profondi sulla democrazia e l'autonomia individuale è al centro dell'analisi di Nemitz (2018) in "Constitutional democracy and technology in the age of artificial intelligence". Nemitz, un giurista e funzionario della Commissione Europea, sostiene che l'IA rappresenti una sfida senza precedenti per i principi e le istituzioni della democrazia liberale. Secondo l'autore, la capacità dell'IA di prendere decisioni autonome e di influenzare il comportamento degli individui mette in discussione concetti fondamentali come la responsabilità, la trasparenza e la partecipazione democratica. Nemitz si chiede come garantire che le decisioni prese dalle IA siano legittime e accountable, e non ledano i diritti fondamentali dei cittadini.

[64] L'idea che chi ha accesso alle forme più avanzate di IA possa non solo processare informazioni e prendere decisioni più velocemente, ma anche "vedere più lontano", anticipando trend e scenari futuri con una precisione senza precedenti, è esplorata da Bostrom (2014) in "Superintelligence: Paths, Dangers, Strategies". Bostrom sostiene che una delle caratteristiche distintive di un'IA superintelligente sarebbe la sua capacità di pianificazione e previsione a lungo termine, grazie alla sua abilità di simulare e valutare un'enorme quantità di scenari futuri. Secondo l'autore, un'IA di questo tipo potrebbe analizzare trend e correlazioni complesse in campi come l'economia, la politica, la tecnologia e l'ambiente, e anticipare eventi e sviluppi con un orizzonte temporale di decenni o secoli.

[65] La metafora del "cannocchiale cognitivo" per descrivere il vantaggio previsionale conferito dall'IA è elaborata da Bostrom (2014) in "Superintelligence: Paths, Dangers, Strategies". Bostrom paragona l'IA a un telescopio che permette di vedere più lontano e con maggiore risoluzione nel "paesaggio" del futuro. Secondo l'autore, così come l'invenzione del telescopio ha rivoluzionato l'astronomia, permettendo di osservare stelle e pianeti prima invisibili, lo sviluppo di IA sempre più potenti potrebbe rivoluzionare la nostra capacità di anticipare e plasmare il futuro. Bostrom immagina scenari in cui un'IA superintelligente potrebbe prevedere con largo anticipo eventi come scoperte scientifiche, rivoluzioni tecnologiche, conflitti geopolitici o crisi ambientali, e suggerire strategie ottimali per sfruttare le opportunità o mitigare i rischi.

[66] L'idea che chi è escluso dalle tecnologie di IA rischi di trovarsi costantemente "in affanno", reagendo a eventi che sembrano arrivare dal nulla, è sviluppata da Susskind (2018) in "Future Politics: Living Together in a World Transformed by Tech". Susskind sostiene che una delle conseguenze politiche dell'IA sarà quella di aumentare enormemente l'asimmetria di potere e conoscenza tra chi ha accesso a queste tecnologie e chi ne è escluso. Secondo l'autore, coloro che possono utilizzare l'IA per analizzare grandi quantità di dati e simulare scenari futuri avranno un enorme vantaggio nel "gioco" della politica e dell'economia, potendo anticipare e influenzare gli eventi a proprio favore. Al contrario, coloro che sono "disconnessi" dall'IA si troveranno sempre più spesso a subire decisioni e cambiamenti che non avranno potuto prevedere né tantomeno determinare.

[67] L'idea che il divario nella capacità predittiva tra "inclusi" ed "esclusi" dall'IA possa alimentare un senso di fatalismo e impotenza tra questi ultimi è esplorata da Susskind (2018) in "Future Politics: Living Together in a World Transformed by Tech". Susskind sostiene che, man mano che l'IA renderà il futuro sempre più "calcolabile" per chi la controlla, e sempre più "imperscrutabile" per chi ne è escluso, crescerà il rischio di una "crisi di agenzia" politica tra la popolazione generale. Secondo l'autore, se le persone percepiranno il futuro come qualcosa di già scritto, determinato da forze al di fuori del loro controllo, potrebbero perdere fiducia nella propria capacità di plasmare il mondo attraverso l'azione politica.

[68] L'idea che la conoscenza del futuro conferita dall'IA, come quella del passato, possa essere una fonte di potere e dominio è sviluppata da Harari (2018) in "21 Lessons for the 21st Century". Harari sostiene che, nel corso della storia, il controllo delle informazioni sul passato, attraverso strumenti come la scrittura, l'istruzione e la propaganda, sia stato un elemento chiave del potere politico e sociale. Secondo l'autore, nell'era dell'IA, il controllo delle informazioni sul futuro potrebbe diventare altrettanto cruciale. Harari immagina scenari in cui governi, aziende o individui con accesso esclusivo a IA predittive potrebbero utilizzarle per manipolare le aspettative e i comportamenti della popolazione, o per acquisire vantaggi economici e strategici.

[69] Il rischio che la capacità predittiva dell'IA venga scambiata per una forma di "onniscienza", e che si riponga eccessiva fiducia nelle sue raccomandazioni, è discusso da Bostrom (2014) in "Superintelligence: Paths, Dangers, Strategies". Bostrom mette in guardia contro la "fallacia dell'onniscienza" nel valutare le prestazioni di un'IA superintelligente. Secondo l'autore, anche se un'IA potrebbe sembrare infallibile nella maggior parte dei suoi giudizi e previsioni, essa si baserebbe comunque su modelli probabilistici e approssimati della realtà. Bostrom sottolinea che anche lievi imperfezioni o "punti ciechi" in questi modelli potrebbero portare a errori catastrofici se amplificati su larga scala.

[70] Lo storico Carlo M. Cipolla, nel suo libro "Allegro ma non troppo", riflette sulla natura imprevedibile e irripetibile della storia umana. Cipolla sottolinea come ogni evento storico sia il risultato di una complessa interazione di fattori contingenti e spesso casuali, che rendono impossibile qualsiasi previsione certa del futuro. Secondo l'autore, anche le analisi più sofisticate del passato non possono eliminare l'elemento di incertezza e di sorpresa insito nel divenire storico, poiché ogni epoca è unica e irriducibile a schemi precostituiti. Cipolla invita quindi a una certa umiltà e cautela nell'interpretare e nel prevedere il corso degli eventi umani, riconoscendo i limiti intrinseci di qualsiasi sforzo di "addomesticare" il futuro.

[71] L'idea che la governance dell'IA debba affrontare non solo la questione dell'accesso alle tecnologie, ma anche quella del controllo sulla "visione del futuro" che esse rendono possibile, è sviluppata da Nemitz (2018) in "Constitutional democracy and technology in the age of artificial intelligence". Nemitz sostiene che, per essere legittima in una società

democratica, la "potenza previsionale" dell'IA debba essere soggetta a meccanismi di controllo e responsabilità pubblici. Secondo l'autore, decisioni cruciali basate su previsioni algoritmiche, come quelle riguardanti gli investimenti pubblici, le politiche sociali o le strategie di sicurezza, non possono essere lasciate nelle mani di pochi tecnocrati o interessi privati.

[72] L'idea che il futuro non sia qualcosa che possa essere semplicemente "previsto" dall'IA, ma qualcosa che debba essere attivamente immaginato, negoziato e realizzato dalla società attraverso processi di partecipazione democratica è sviluppata da Jasanoff (2016) in "The Ethics of Invention: Technology and the Human Future". Jasanoff, una sociologa della scienza, sostiene che le visioni del futuro incorporate nelle tecnologie, e in particolare nell'IA, non siano mai neutre, ma riflettano sempre particolari assunti, valori e interessi. Secondo l'autrice, delegare all'IA il compito di "prevedere" il futuro rischia di oscurare il carattere intrinsecamente politico e contestabile di questo esercizio. Jasanoff invita invece a "co-produrre" attivamente il futuro attraverso il coinvolgimento di una pluralità di prospettive e di saperi nella progettazione e nella governance delle tecnologie.

[73] L'idea che lo sviluppo incontrollato dell'IA possa portare a una divaricazione catastrofica all'interno della specie umana, con conseguenze devastanti per la coesione sociale e la dignità delle persone, è paventata da Harari (2018) in "21 Lessons for the 21st Century". Harari avverte che, se permetteremo che i benefici dell'IA siano monopolizzati da pochi, potremmo assistere all'emergere di una nuova "casta" di super-umani potenziati dalle macchine, che godranno di capacità cognitive, fisiche ed economiche irraggiungibili per il resto della popolazione. Secondo l'autore, questo scenario potrebbe portare a una disuguaglianza senza precedenti, in cui la maggioranza dell'umanità verrebbe considerata come "pesi morti" o "esuberanti" dai "vincitori" della rivoluzione dell'IA.

[74] L'idea che la governance dell'IA rappresenti una delle sfide cruciali del nostro tempo, e richieda lo sviluppo di quadri normativi e istituzionali adeguati a gestirne le implicazioni, è al centro dell'analisi di Cath et al. (2018) in "Artificial Intelligence and the 'Good Society': The US, EU, and UK approach". Gli autori confrontano le strategie di governance dell'IA adottate da Stati Uniti, Unione Europea e Regno Unito, e ne discutono i punti di forza e di debolezza. Secondo gli autori, nonostante le differenze di approccio, tutti e tre i sistemi riconoscono la necessità di un intervento pubblico per promuovere lo sviluppo responsabile e inclusivo dell'IA. Gli autori individuano alcune aree chiave su cui concentrare gli sforzi di governance, come la protezione dei diritti fondamentali, la trasparenza e la responsabilità degli algoritmi, la sicurezza e la robustezza dei sistemi, e la promozione della diversità e dell'equità nella ricerca e nell'uso dell'IA.

[75] L'idea che lo sviluppo dell'IA debba essere guidato da principi etici solidi e condivisi è al centro delle "Ethically Aligned Design Guidelines" elaborate dall'IEEE (2019), la maggiore associazione professionale mondiale di ingegneri. Le linee guida, frutto di un processo di consultazione globale tra esperti di IA, stakeholder e cittadini, mirano a promuovere la progettazione di sistemi di IA "eticamente allineati", ovvero rispettosi dei diritti umani, socialmente benefici, robusti e affidabili. Il documento identifica una serie di principi chiave, come la trasparenza, la responsabilità, la privacy, la non discriminazione e il controllo umano, e fornisce raccomandazioni pratiche per la loro implementazione nello sviluppo e nell'uso dell'IA.

[76] L'idea che lo sviluppo dell'IA debba essere guidato da valori come la trasparenza, la responsabilità, l'equità e il rispetto dei diritti umani è esplorata da Jobin et al. (2019) nell'articolo "The global landscape of AI ethics guidelines". Gli autori hanno condotto una revisione sistematica delle linee guida etiche per l'IA proposte da organizzazioni di tutto il

mondo, identificando 11 principi ricorrenti. Tra questi, i più frequenti sono la trasparenza e la spiegabilità dei sistemi di IA, la responsabilità e l'imputabilità per le loro azioni, l'equità e la non discriminazione nel loro utilizzo, e il rispetto per la privacy e i diritti umani degli individui. Altri principi riguardano la sicurezza e la robustezza dei sistemi, la promozione del benessere umano, la responsabilità sociale, la libertà e l'autonomia individuali.

[77] Le "Ethics Guidelines for Trustworthy AI" elaborate dall'Unione Europea (2019) rappresentano uno dei tentativi più ambiziosi di definire un quadro etico e normativo per lo sviluppo dell'IA. Le linee guida, sviluppate da un gruppo di esperti indipendenti, mirano a promuovere un approccio "human-centric" all'IA, basato sui valori fondamentali dell'UE come il rispetto della dignità umana, la libertà, la democrazia, l'uguaglianza e lo stato di diritto. Il documento identifica tre componenti necessarie per un'IA "trustworthy": legalità (rispetto delle leggi e dei regolamenti applicabili), eticità (adesione a principi etici e valori condivisi) e robustezza (dal punto di vista tecnico e sociale). Le linee guida forniscono una serie di requisiti per realizzare questi obiettivi, come la supervisione umana sui sistemi di IA, la trasparenza e la tracciabilità dei loro processi decisionali, la protezione della privacy e dei dati personali, la promozione della diversità e dell'inclusione.

[78] I "Principi di Asilomar per l'IA" (2017) sono un insieme di linee guida etiche per lo sviluppo dell'IA elaborati da un gruppo di esperti durante una conferenza organizzata dal Future of Life Institute. I principi, sottoscritti da migliaia di ricercatori, imprenditori e policy-maker di tutto il mondo, mirano a promuovere i benefici dell'IA e a mitigarne i rischi esistenziali. Il documento identifica tre aree principali di intervento: la ricerca (sviluppare sistemi di IA robusti, sicuri e allineati con i valori umani), l'etica e i valori (promuovere la trasparenza, la responsabilità e il rispetto dei diritti umani nell'uso dell'IA), e le questioni a lungo termine (prepararsi alle trasformazioni economiche, sociali e politiche indotte dall'IA). I principi sottolineano l'importanza di mantenere il controllo umano sui sistemi d'arma autonomi, di evitare una corsa agli armamenti nell'IA, e di garantire che i benefici dell'IA siano ampiamente condivisi.

[79] L'importanza di promuovere l'educazione e la formazione nelle competenze legate all'IA per garantire una partecipazione più ampia ai suoi benefici è sottolineata da Bughin et al. (2017) nel rapporto "Artificial Intelligence: The Next Digital Frontier?" del McKinsey Global Institute. Gli autori stimano che l'adozione dell'IA potrebbe aggiungere 13 trilioni di dollari al PIL globale entro il 2030, ma avvertono che questi benefici potrebbero essere distribuiti in modo molto diseguale se non si interverrà per promuovere l'accesso diffuso alle competenze e alle opportunità legate all'IA. Il rapporto identifica una serie di azioni prioritarie per i governi e le imprese, come investire nell'istruzione STEM (scienza, tecnologia, ingegneria e matematica) a tutti i livelli, promuovere la formazione continua e la riqualificazione dei lavoratori, e sostenere la ricerca e l'innovazione aperta nell'IA.

[80] Il concetto di "olarchia", ovvero di un'architettura organizzativa basata su reti di sistemi autonomi e interdipendenti, è sviluppato da Koestler (1967) nel libro "The Ghost in the Machine". Koestler definisce un'"olarchia" come un sistema gerarchico aperto e multilivello, in cui ogni livello è composto da "oloni", ovvero da entità che sono al contempo un tutto autonomo rispetto al livello inferiore e una parte dipendente rispetto al livello superiore. Secondo l'autore, questo modello ricorsivo e scalabile, ispirato alle strutture biologiche, consentirebbe di coniugare l'autonomia locale con il coordinamento globale, la specializzazione con l'integrazione, la diversità con l'unità. Koestler propone l'olarchia come un'alternativa tanto alle gerarchie rigide e centralizzate quanto alle reti caotiche e frammentate, e ne esplora le applicazioni in diversi domini, dalla psicologia all'organizzazione sociale.

[81] L'idea di promuovere una "data economy partecipativa", in cui le comunità possano raccogliere, gestire e valorizzare in modo autonomo i propri dati, è esplorata da Morozov (2018) nell'articolo "Digital Socialism? The Calculation Debate in the Age of Big Data". Morozov sostiene che, per contrastare lo strapotere delle grandi piattaforme digitali e democratizzare i benefici dell'IA, sia necessario sviluppare nuovi modelli di gestione dei dati ispirati ai principi dell'economia collaborativa e dei beni comuni. Secondo l'autore, le comunità dovrebbero organizzarsi in "data cooperative" o "data trust", ovvero in strutture collettive che consentano ai cittadini di aggregare i propri dati e di decidere democraticamente come usarli e condividerli, anziché cederli passivamente alle aziende in cambio di servizi.

[82] I concetti di "data sovereignty" e "data trust" come modelli innovativi di governance dei dati sono sviluppati da Hummel et al. (2021) nell'articolo "Sovereignty and data sharing". Gli autori definiscono la "data sovereignty" come il diritto degli individui e delle comunità di controllare i propri dati, decidendo chi può accedervi, per quali scopi e a quali condizioni. I "data trust" sono invece definiti come strutture legali che consentono di delegare la gestione dei dati a un'entità fiduciaria, con il mandato di amministrarli nell'interesse dei titolari. Secondo Hummel et al., questi due approcci potrebbero promuovere una gestione dei dati più etica, trasparente e orientata al bene comune, bilanciando i diritti individuali con le esigenze collettive. Gli autori discutono alcune sfide e opportunità legate all'implementazione pratica di queste soluzioni, come la definizione di standard tecnici e legali, la creazione di incentivi all'adozione, e l'allineamento con i valori e le preferenze delle comunità.

[83] L'idea di sfruttare il potenziale del crowdsourcing e del crowd computing per sviluppare sistemi di IA allineati con le conoscenze e i valori delle comunità è esplorata da Bigham et al. (2015) nell'articolo "Human-in-the-Loop Machine Learning". Gli autori sostengono che coinvolgere attivamente gli utenti nella progettazione, nell'addestramento e nella valutazione dei sistemi di IA possa migliorarne la qualità, la trasparenza e l'equità. Secondo Bigham et al., le tecniche di "human-in-the-loop machine learning", che combinano l'apprendimento automatico con il feedback umano, consentirebbero di sviluppare modelli più robusti, spiegabili e adattabili alle esigenze degli utenti. Gli autori portano esempi di piattaforme partecipative per l'annotazione dei dati, la generazione di esempi, la definizione di obiettivi e metriche, e la revisione delle decisioni algoritmiche. Bigham et al. sottolineano l'importanza di progettare interfacce intuitive e accessibili, di ricompensare adeguatamente i contributori, e di garantire la diversità e la rappresentatività delle comunità coinvolte.

[84] Il concetto di "commons", ovvero di risorse condivise e autogestite dalle comunità, è sviluppato da Ostrom (1990) nel libro "Governing the Commons: The Evolution of Institutions for Collective Action". Ostrom, premio Nobel per l'economia, sfida l'idea convenzionale che i beni comuni siano destinati al sovrasfruttamento e al degrado se non privatizzati o regolati dallo Stato. Attraverso numerosi casi di studio, l'autrice mostra come le comunità locali siano spesso in grado di sviluppare istituzioni e regole efficaci per la gestione sostenibile di risorse comuni come pascoli, foreste, sistemi di irrigazione e pescherie. Ostrom identifica una serie di principi di design che caratterizzano i commons di successo, come la chiara definizione dei confini e dei diritti di accesso, la partecipazione degli utilizzatori alle decisioni collettive, il monitoraggio e le sanzioni graduate per i trasgressori, e la risoluzione rapida dei conflitti. L'autrice propone il modello dei commons come un'alternativa tanto al mercato quanto allo Stato per la gestione di risorse scarse e di interesse collettivo.

[85] L'idea che i dati siano il "carburante" dell'IA e che il modo in cui vengono raccolti, utilizzati e controllati avrà un'influenza decisiva sulla distribuzione dei benefici e dei rischi di queste tecnologie è sviluppata da Pasquale (2015) in "The Black Box Society: The Secret

*Algorithms That Control Money and Information". Pasquale sostiene che le grandi piattaforme digitali, come Google, Facebook e Amazon, abbiano acquisito un potere senza precedenti grazie alla loro capacità di raccogliere ed analizzare enormi quantità di dati personali. Secondo l'autore, queste aziende utilizzano algoritmi di IA "black box", ovvero opachi e imperscrutabili, per profilare gli utenti, influenzarne i comportamenti e prendere decisioni che hanno un impatto significativo sulle loro vite, dalla concessione di prestiti alla determinazione dei prezzi.*

[86] *L'idea che sia necessario sviluppare regimi di governance dei dati che bilancino la protezione dei diritti individuali con l'esigenza di promuovere l'accesso e la condivisione dei dati per scopi di interesse pubblico è al centro dell'analisi di Wylie (2019) nell'articolo "Governance is Critical to Getting Data Ethics Right". Wylie, noto per aver denunciato lo scandalo di Cambridge Analytica, sostiene che le attuali normative sulla protezione dei dati, come il GDPR europeo, siano insufficienti a garantire un uso etico e responsabile dei dati nell'era dell'IA. Secondo l'autore, queste norme si concentrano troppo sulla dimensione individuale della privacy, trascurando le implicazioni collettive e sociali dell'uso dei dati. Wylie propone di sviluppare nuovi modelli di governance dei dati, ispirati ai beni comuni e alla democrazia partecipativa, che consentano alle comunità di avere voce in capitolo su come i loro dati vengono raccolti, utilizzati e condivisi.*

[87] *L'idea che l'introduzione di un reddito di base universale possa essere necessaria per garantire una distribuzione più equa dei benefici dell'IA e dell'automazione è sviluppata da Van Parijs e Vanderborght (2017) in "Basic Income: A Radical Proposal for a Free Society and a Sane Economy". Gli autori sostengono che, con il progredire dell'IA, una quota crescente della ricchezza sarà prodotta dalle macchine anziché dal lavoro umano, rendendo obsoleto il modello di welfare basato sul lavoro salariato. Secondo Van Parijs e Vanderborght, per evitare che i guadagni di produttività dell'IA si traducano in una crescente disuguaglianza e disoccupazione tecnologica, è necessario introdurre un reddito di base universale, ovvero un trasferimento monetario incondizionato a tutti i cittadini. Gli autori argomentano che il reddito di base, finanziato dalla tassazione dei "dividendi" dell'IA, consentirebbe di redistribuire i benefici dell'automazione, di promuovere la libertà e l'autonomia individuali, e di valorizzare attività socialmente utili ma non remunerate come il volontariato e la cura.*

[88] *L'idea di introdurre una "tassa sui robot" per finanziare la transizione all'economia dell'IA e compensare i lavoratori spiazzati dall'automazione è stata avanzata, tra gli altri, da Bill Gates in un'intervista a Quartz (2017). Gates sostiene che, con il progredire dell'IA, sempre più lavori attualmente svolti da umani saranno automatizzati, portando a guadagni di produttività ma anche a perdite di posti di lavoro e di gettito fiscale. Secondo Gates, per evitare che questi effetti si traducano in una crescente disuguaglianza e instabilità sociale, è necessario introdurre una tassa sulle aziende che sostituiscono i lavoratori umani con l'IA e i robot. Il gettito di questa tassa potrebbe essere utilizzato per finanziare programmi di riqualificazione professionale, di sostegno al reddito per i lavoratori spiazzati, e di investimento in settori ad alta intensità di lavoro umano come l'istruzione e la sanità.*

[89] *L'idea che l'influenza crescente dell'IA sui processi democratici richieda nuove forme di regolamentazione e controllo è al centro dell'analisi di Nemitz (2018) in "Constitutional democracy and technology in the age of artificial intelligence". Nemitz sostiene che l'IA, se non adeguatamente governata, rischi di compromettere i principi fondamentali della democrazia costituzionale, come la trasparenza, la responsabilità e la partecipazione dei cittadini. L'autore identifica una serie di sfide poste dall'IA ai processi democratici, come l'uso di algoritmi opachi per la profilazione politica e la diffusione mirata di disinformazione,*

*l'influenza degli "intermediari algoritmici" come i motori di ricerca e i social media sulla formazione dell'opinione pubblica, e l'impatto dei sistemi automatizzati di decisione sulla vita dei cittadini. Nemitz invoca una regolamentazione più stringente delle piattaforme digitali e degli algoritmi che moderano i contenuti online, ispirata ai principi del pluralismo, della diversità e della "par condicio".*

*[90] L'idea che la governance dell'IA richiederà uno sforzo di cooperazione globale senza precedenti, coinvolgendo non solo i governi e le aziende ma l'intera società, è al centro dell'analisi di Cath et al. (2018) in "Artificial Intelligence and the 'Good Society': The US, EU, and UK approach". Gli autori sostengono che, data la natura transfrontaliera e pervasiva dell'IA, nessun paese o organizzazione potrà governarne efficacemente lo sviluppo da solo. Secondo Cath et al., per promuovere un'IA "buona" per la società nel suo complesso, è necessario un approccio multi-stakeholder e partecipativo alla governance, che includa la società civile, il mondo accademico, i media e i cittadini. Gli autori identificano alcune aree prioritarie per la cooperazione internazionale, come la definizione di standard etici e tecnici condivisi per l'IA, lo scambio di buone pratiche e lezioni apprese, e il coordinamento delle politiche pubbliche in materia di ricerca, istruzione e regolamentazione.*

*[91] L'idea che la governance dell'IA sarà un processo di apprendimento continuo, che dovrà adattarsi costantemente all'evoluzione rapida e imprevedibile di queste tecnologie, è sviluppata da Wallach e Marchant (2019) nell'articolo "An Agile Ethical/Legal Model for the International and National Governance of AI and Robotics". Gli autori sostengono che, data la natura dinamica e in rapida evoluzione dell'IA, i tradizionali approcci "top-down" alla regolamentazione, basati su norme prescrittive e sanzioni, siano inadeguati. Secondo Wallach e Marchant, per stare al passo con l'IA, è necessario sviluppare modelli di governance "agili", basati su principi generali, linee guida flessibili e processi di apprendimento iterativo. Gli autori propongono un modello "ibrido" di governance, che combini standard etici volontari, best practice condivise, certificazioni di conformità, e forme di "soft law" come codici di condotta e accordi settoriali. Questo approccio consentirebbe di sperimentare e adattare le regole nel tempo, coinvolgendo attivamente gli stakeholder nello sviluppo e nell'implementazione delle norme.*