

IMPATTO ECONOMICO DEI NUOVI MODELLI DI INTELLIGENZA ARTIFICIALE: SFIDE E OPPORTUNITÀ PER L'UNIONE EUROPEA *

Anna Maria Catenacci[§], Gabriele Velpi^{*}

ABSTRACT

La nota analizza benefici e rischi di ordine economico direttamente o indirettamente associati ai nuovi modelli di intelligenza artificiale (di seguito, IA). L'attenzione è posta in primis sugli effetti sulla produttività, sul mercato del lavoro e sulle dipendenze strategiche, con l'obiettivo di delineare alcune indicazioni di policy per l'Unione Europea.

La rapida evoluzione delle tecnologie nascenti, la capacità dei modelli IA di migliorare le loro prestazioni ed espandere le loro capacità man a mano che vengono utilizzati (*machine learning* o ML) e la loro idoneità ad **accelerare l'innovazione** in altri campi permettono di ipotizzare **effetti esponenziali** sull'economia e sulle società. La progressione potenzialmente esponenziale degli effetti e la **concentrazione geografica** degli attori attivi nel settore rendono urgente e complessa l'azione dei governi nazionali e dell'Unione Europea, anche per ragioni geopolitiche.

EXECUTIVE SUMMARY

Dall'analisi condotta, si considerano particolarmente rilevanti i seguenti punti relativi alla natura dell'IA e alle sue implicazioni di ordine economico:

- I nuovi modelli di IA possono svolgere compiti complessi finora appannaggio di lavoratori istruiti e sono in grado di **apprendere nuove competenze** (*machine learning*) interagendo con gli utenti; inoltre presentano le cosiddette "**capacità emergenti**" (capacità che la macchina mostra di acquisire all'aumentare della scala dei modelli stessi, e che pertanto sono considerate da alcuni studiosi imprevedibili);
- Tuttavia, i nuovi modelli di IA sono anche soggetti a diversi **limiti**, alcuni di carattere strutturale (in parte legati alla differenza tra intelligenza artificiale e umana), altri potenzialmente superabili con l'evolversi della tecnologia, ma anche con l'adattamento dei sistemi di organizzazione

* La presente nota è frutto dell'attività di ricerca di Anna Maria Catenacci e dell'attività di guida e supervisione di Gabriele Velpi. Si ringraziano Luigi Disanto e Francesco Nucci per i preziosi suggerimenti ricevuti.

§ Funzionaria dell'ufficio IV della Direzione Rapporti Finanziari Europei del dipartimento del Tesoro. E-mail: annamaria.catenacci@mef.gov.it.

* Dirigente dell'ufficio VI della Direzione Analisi e ricerca economico-finanziaria del dipartimento del Tesoro. E-mail: gabriele.velpi@mef.gov.it.

JEL:

J01, J08, L10, L50, O33, O38, O40

Keywords:

intelligenza artificiale, produttività, mercato del lavoro, disoccupazione tecnologica, disuguaglianza, regolamentazione, policy europea.

Nota Tematica:

La collana intende promuovere la circolazione di Note Tematiche prodotte all'interno del Dipartimento del Tesoro del Ministero dell'Economia e delle Finanze.

Il contenuto delle Note Tematiche riflette esclusivamente le opinioni degli autori e non impegna in alcun modo l'amministrazione.

del lavoro, dei comportamenti degli utenti, e di altri aspetti che, come questi, possono essere in parte influenzati dall'intervento del *policy-maker*;

- L'IA costituisce una *general purpose technology* e dunque può avere utili applicazioni nella generalità dei settori economici e nella pubblica amministrazione;
- L'impatto stimato da più fonti sulla produttività e sulla crescita del prodotto è notevole, per quanto difficile da quantificare: l'introduzione dei nuovi modelli di IA potrebbe condurre a **un aumento** non solo del PIL pro capite ma anche **del tasso di crescita della produttività**, tramite l'accelerazione dell'innovazione (ad esempio, l'IA si è già rivelato uno strumento molto potente sia nella ricerca di base, sia nella generazione di idee di *business* o di contenuti per il *marketing*);
- È stato stimato che i nuovi modelli di IA, e i *software* alimentati da essi, saranno in grado di **effettuare una porzione sostanziosa dei compiti attualmente svolti dai lavoratori**: uno studio del FMI stima che l'IA avrà un impatto sul 40% dei posti di lavoro a livello mondiale, raggiungendo il 60% nelle economie avanzate. L'impatto avrebbe effetti negativi su circa la metà dei lavoratori coinvolti;
- Più precisamente, l'effetto sostituzione del lavoro, che si verificherebbe anche solo parzialmente per il 30% dei posti di lavoro, potrebbe essere compensato, in tutto o in parte, o più che compensato da altri effetti, come **l'effetto produttività** o **l'effetto reintegrazione**; per questa ragione, ad oggi, è difficile esprimere valutazioni definitive circa l'impatto netto dei diversi canali di trasmissione sul mercato del lavoro (ad esempio, così l'OCSE);
- Molto probabilmente l'introduzione dell'IA nella produzione comporterà notevoli **effetti redistributivi** tra diverse fasce di lavoratori (con la possibilità di esacerbare il fenomeno di polarizzazione dei redditi) e tra i fattori produttivi, determinando l'aumento della quota del reddito totale destinata alla retribuzione del capitale. Ciò sarà determinato anche dalla capacità dei modelli di apprendere grazie all'interazione con gli utilizzatori, **trasformando potenzialmente ogni impiego del capitale IA in un investimento** che accresce le capacità e quindi il valore dello stesso capitale;
- Un fenomeno degno di nota è legato alla **continua trasmissione di conoscenza** alle imprese proprietarie dei modelli IA da parte degli utenti (inclusi i lavoratori che utilizzano i modelli di IA nello svolgimento delle proprie mansioni): ciò si traduce nella trasmissione delle capacità produttive del lavoro al bene capitale rappresentato dai modelli IA;
- L'IA potrebbe avere un effetto importante anche sugli **stili di consumo**, aumentando l'accesso dei consumatori alle informazioni e riducendo le asimmetrie informative in tantissimi settori. Al tempo stesso, i consumatori potrebbero non essere immediatamente in grado di valutare le differenze qualitative tra i servizi e i beni prodotti con l'ausilio dell'IA e quelli prodotti dal lavoro umano;
- Il settore dell'IA appare **notevolmente concentrato** (in poche società soprattutto cinesi e statunitensi) e **poco trasparente**, poiché governi e istituti di ricerca indipendente non possiedono i mezzi (capacità di calcolo) per studiare i modelli di grandi dimensioni. La correlazione

positiva e crescente tra la dimensione dei modelli e le capacità che questi dimostrano (scalabilità) potrebbe ostacolare l'entrata di nuovi attori nel settore, dominato da pochi *incumbent*, integrati verticalmente con imprese di grandi dimensioni, attive nella **raccolta dati** da almeno un paio di decenni e in grado di investire miliardi nella **capacità computazionale**;

- Alla luce dell'attuale struttura dell'industria, la diffusione dell'IA nei settori produttivi e nella società può avere un impatto sul piano geopolitico, facendo emergere **nuove dipendenze strategiche**;
- Altri rischi notevoli possono presentarsi qualora se ne faccia un uso inconsapevole o se ne abusi (ad esempio, l'IA può essere di supporto nel compiere diverse tipologie di azioni illegali).

Le principali prescrizioni di *policy* che discendono dall'analisi sono le seguenti:

- L'UE dovrà cercare il giusto bilanciamento tra diverse necessità: oltre a governare i potenziali rischi, **mantenendo elevati standard di qualità e sicurezza** per i beni e i servizi prodotti con le nuove tecnologie e commercializzati nel Mercato Unico (obiettivo dell'AI Act), dovrà anche salvaguardare **i posti di lavoro e la produttività** delle imprese europee e **tutelare gli interessi e i valori dell'UE** a fronte di potenziali dipendenze tecnologiche. In altre parole, la **regolamentazione sugli standard di qualità** potrebbe svolgere un ruolo anche per bilanciare innovazione, diritti e autonomia strategica;
- Per quanto riguarda le **opportunità offerte dalle nuove tecnologie IA**, occorre, da una parte, favorire **l'adozione della tecnologia** prodotta all'estero per salvaguardare la competitività nei diversi settori economici (aumentando l'efficienza delle imprese europee), e puntare sulle **innovazioni collaterali** ai modelli *transformer* e LLM, ovvero sui cambiamenti di processo e di prodotto volti a sfruttarne le potenzialità, con importanti impatti anche sui settori non digitali e, dall'altra, sostenere la **nascita di un settore IA europeo**.

Nello specifico la nota è strutturata come segue:

Il primo capitolo fornisce una definizione operativa di intelligenza artificiale, descrive i recenti sviluppi nell'ambito (modelli *transformer*, *large language models*), offre una panoramica del settore e delle applicazioni concrete della nuova tecnologia e infine descrive astrattamente i limiti costitutivi e contingenti dei nuovi modelli.

Nel secondo capitolo, sono discussi gli impatti in termini di produttività e crescita, cercando di dare un quadro dei costi di produzione dei modelli di intelligenza artificiale, degli investimenti e dei tempi necessari a produrre tali effetti sulla produttività; accenna infine ad alcuni effetti di redistribuzione tra fattori produttivi.

Il terzo capitolo, dedicato all'impatto sul mercato del lavoro, prende in esame i modelli *transformer* e i LLM in quanto fattori produttivi, cercando di cogliere la loro natura ibrida e inedita, con caratteristiche tipiche del fattore capitale e alcune abilità proprie del fattore lavoro (es. la capacità di apprendere e aggiornarsi).

Nel quarto capitolo sono descritti alcuni possibili aspetti extra-economici con potenziali ricadute indirette sull'economia, ad esempio considerando

le conseguenze per i media e quelle di natura geopolitica, derivanti dalla distribuzione geografica delle imprese sviluppatrici.

Infine, nel quinto e ultimo capitolo sono raccolte alcune proposte di *policy* per massimizzare le opportunità e gestire i rischi relativi alla nuova tecnologia. In particolare, sono discussi: la riqualificazione dei lavoratori, le riforme necessarie nell'istruzione, la regolamentazione, gli *standard* di qualità e le politiche volte a salvaguardare l'autonomia strategica.

Gli *annex* approfondiscono alcune delle questioni tecniche più rilevanti in termini di definizione e capacità mostrata dall'IA, esaminano più dettagliatamente i costi derivati dalle grandi dimensioni dei modelli e presentano un quadro delle iniziative legislative americane ed europee.

SOMMARIO

ABSTRACT	1
EXECUTIVE SUMMARY	1
1. FOTOGRAFIA DELLO STATO DELL'ARTE IN MATERIA DI INTELLIGENZA ARTIFICIALE	1
2. IMPATTI ATTESI DELLA IA SU PRODUTTIVITÀ E CRESCITA	19
3. IMPATTO SUL MERCATO DEL LAVORO	28
4. FATTORI EXTRA-ECONOMICI CON POSSIBILI IMPATTI SULL'ECONOMIA	35
5. POLICY E REGOLAMENTAZIONE	40
6. CONCLUSIONI	48
ANNEX 1 - GLOSSARIO	51
ANNEX 2 - PROBLEMI RELATIVI ALLA DEFINIZIONE E ALLA VALUTAZIONE DELL'IA	52
ANNEX 3 - MODELLI TRANSFORMER E LLM: APPRENDIMENTO E ADATTAMENTO	55
ANNEX 4 - GENERATIVE AI E AGI (ARTIFICIAL GENERAL INTELLIGENCE): IMPREVEDIBILITÀ DELLA MACCHINA, DIMENSIONI DEI MODELLI E CAPACITÀ EMERGENTI	57
ANNEX 5 - DIMENSIONI DEI MODELLI, COSTI DI SVILUPPO E INVESTIMENTI	60
ANNEX 6 - PRINCIPALI INIZIATIVE DI GOVERNANCE GLOBALI, REGIONALI E NAZIONALI	62
ANNEX 7 - NEGOZIAZIONE E CONTENUTI DELL'AI ACT	66
RIFERIMENTI BIBLIOGRAFICI	71

1. FOTOGRAFIA DELLO STATO DELL'ARTE IN MATERIA DI INTELLIGENZA ARTIFICIALE

Questo primo capitolo fornisce una definizione operativa di intelligenza artificiale, termine ampio con una lunga storia, e descrive il recente **cambio di paradigma** avvenuto con l'avvento dei modelli "*transformer*", una tipologia molto efficiente di reti neurali digitali, di cui i "*large language models*" (LLMs) sono un sottoprodotto. Nella nota, fatti salvi i casi in cui è specificato diversamente, il termine 'intelligenza artificiale' è utilizzato in senso stretto per riferirsi ai modelli "*transformer*", la cui capacità di apprendimento presenta potenzialmente molteplici risvolti economici (ad esempio, sull'accumulo di capitale e la crescita, sulla struttura del mercato, sulla concorrenza e sul rapporto tra i fattori produttivi).

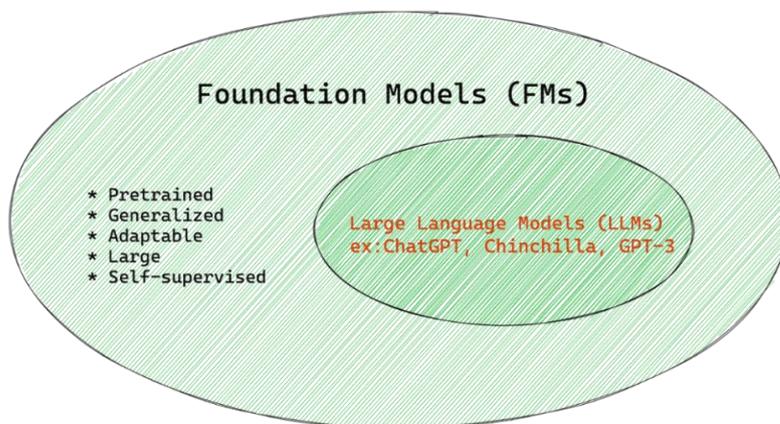
Il secondo paragrafo offre una panoramica del settore sotto il profilo della struttura del mercato e del comportamento strategico delle imprese, evidenziando anche l'attuale distribuzione geografica degli attori principali. Nel terzo paragrafo, attraverso alcuni esempi concreti, sono illustrate le molteplici applicazioni potenziali dei nuovi modelli, che rappresentano una *general purpose technology*. Nel quarto paragrafo si precisa l'ottica concreta e di lungo periodo applicata all'analisi della questione nel resto della nota, ispirandosi all'approccio dell'osservatorio *AI watch* della Commissione. In conclusione, nel quinto paragrafo, si propone una **classificazione dei limiti dell'attuale tecnologia IA**, riflettendo in particolare sulla non immediata distinzione tra quelli che potrebbero ragionevolmente essere considerati limiti costitutivi e quelli che, invece, potrebbero rivelarsi limiti contingenti. Questi ultimi sono destinati ad essere superati non solo in virtù di futuri progressi tecnologici, ma anche a seguito di un'evoluzione culturale, ovvero di modifiche dei sistemi organizzativi e dei comportamenti individuali: si delinea quindi uno **spazio di azione per la politica**, concetto che sarà ripreso nel capitolo quinto, dedicato alle *policy* volte a indirizzare deliberatamente le potenziali trasformazioni innescate dall'adozione della tecnologia IA.

a. Definizione di intelligenza artificiale

"Intelligenza artificiale" è un termine ombrello nato negli anni '50 del '900 per indicare genericamente tutti i metodi che consentono ai *computer* di simulare, anche solo in parte, il comportamento umano

¹. Negli ultimi anni, e in particolare nel periodo tra il 2017 e il 2021, nel settore IA è avvenuto un cambio di paradigma rivoluzionario: l'impiego di nuove metodiche matematiche ha portato infatti alla nascita dei modelli '*transformer*' (anche detti *foundation models* proprio per il loro carattere fondativo di un nuovo paradigma). Come si vede nell'immagine di seguito, i *Large Language Models* sono un esempio dell'applicazione dei modelli *transformer*.

¹ "Una macchina può gestire le informazioni; può calcolare, concludere e scegliere; può eseguire operazioni razionali con le informazioni. Una macchina, quindi, può pensare.", sosteneva Edmund Berkeley nel suo "*Giant brains: or machines that think*" del 1949, descrivendo i primi computer.



Fonte: Stanford HAI – Human-centered Artificial Intelligence.

I modelli *transformer* si distinguono per essere particolarmente efficienti nell'emulazione dell'apprendimento (*machine learning* o ML)². L'impiego dei modelli *transformer*, unitamente alla disponibilità delle enormi collezioni di dati oggi esistenti e della capacità computazionale necessaria a gestirli, ha permesso lo sviluppo di modelli linguistici basati su miliardi di parametri, detti, proprio per le loro dimensioni notevoli, *Large Language Models* (LLM). Per avvicinarsi in modo semplificato al funzionamento di questi modelli linguistici li si può immaginare come un'applicazione molto avanzata dei sistemi probabilistici in grado di prevedere ciò che l'utilizzatore di un programma di testo si appresta a digitare su una tastiera (*next-word prediction*).

Infatti, come i sistemi di previsione della parola successiva, anche gli LLM calcolano la probabilità che una parola segua un'altra, basandosi sulle fonti osservate (i testi su cui sono addestrati), ma gli LLM sono in grado di prevedere la parola con maggiore accuratezza grazie a una mappatura dettagliatissima delle relazioni tra le parole, che raggiunge una precisione sufficiente a **imitare la comprensione umana** dei singoli lemmi e dei testi nel loro insieme. Infatti, gli LLM **mappano le relazioni tra le parole** in spazi a migliaia dimensioni, non considerando le caratteristiche sintattiche delle parole, bensì **il loro aspetto semantico**. Sebbene le conoscenze sul loro funzionamento complessivo siano oggi incomplete, sappiamo tuttavia

² Il *machine learning* comprende le tecniche dette *deep learning*, ovvero le tecniche che tentano di emulare l'apprendimento umano imitando il funzionamento del cervello biologico: nell'apprendimento umano ogni nuova informazione acquisita dal cervello viene confrontata con i dati precedentemente memorizzati per affinare le conoscenze procedendo 'per strati'. Nel *deep learning* sono impiegate strutture algoritmiche a più livelli note come "reti neurali digitali" (*digital neural networks*) che utilizzano lo stesso principio: ogni livello della rete neurale svolge una funzione di filtro per affinare ad ogni passaggio i concetti appresi. Data la loro natura stratificata, i *neural networks* possono essere più o meno profondi. Il *deep learning* è una tecnica sviluppata a partire dagli studi di Geoffrey Hinton, attualmente professore emerito della Toronto University, vincitore del Turing Award 2018 e soprannominato 'the godfather of AI' proprio per il suo ruolo di pioniere nel campo. Gli studi di Hinton hanno permesso di cambiare approccio all'emulazione dell'apprendimento umano partendo dalla biomimetica, ovvero dall'imitazione delle strutture del cervello, anziché dalla logica e dalla conoscenza, che, anche nel caso del cervello umano, sono un prodotto secondario di strutture cerebrali che si sono sviluppate primariamente per lo svolgimento di altri compiti (ad esempio, gestione del movimento).

che gli LLM sono in grado di **creare una rappresentazione della struttura dei concetti contenuti all'interno di un testo**. Questo permette agli LLM di produrre riassunti efficaci e di rispondere efficacemente agli input umani, e comporta come effetto collaterale l'**emergere di capacità impreviste**³. Tra gli LLM, GPT (Generative Pre-trained Transformer) è noto al grande pubblico per l'attenzione ricevuta dal *chatbot* che, grazie al modello linguistico sottostante, è in grado di conversare con gli utenti (chatGPT⁴).

Per comprendere l'impatto potenziale dei modelli *transformer*, tuttavia, è necessario guardare anche oltre i LLM: più in generale, l'impiego delle tecniche di ML, in cui i modelli *transformer* eccellono, permette alla macchina, **a partire da una collezione di dati, di rilevare e registrare le relazioni tra essi e ricavarne il sistema di regole che le determinano**, senza la necessità che tali regole siano contenute nell'algoritmo inizialmente fornito dai programmatori alla macchina.

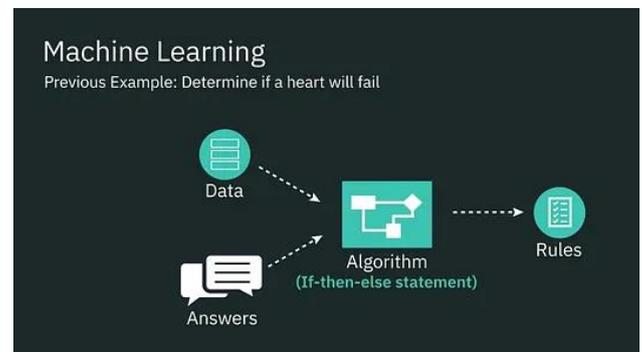
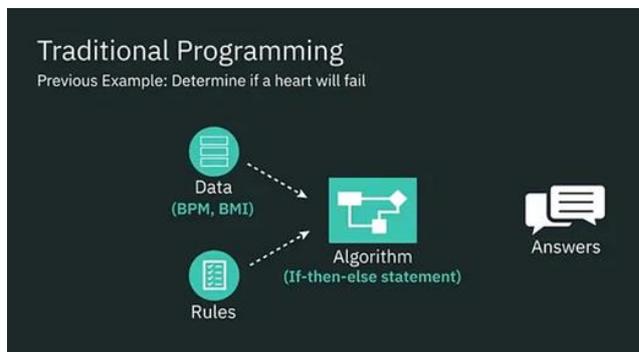
Prima dell'avvento dei *transformer*, il comportamento umano poteva già essere imitato, in alcune sue parti, tramite la programmazione tradizionale (*symbolic AI*), ovvero fornendo alla macchina non solo i dati, ma anche un modello di gestione e interpretazione degli stessi: si tratta dell'algoritmo, ovvero un insieme di regole (*symbolic rules*) che governano le relazioni tra i dati forniti, ragion per cui, è importante ricordare che non tutte le IA utilizzano il *machine learning* per imitare il comportamento umano⁵.

Le elevate capacità di ML esibite dai *transformer* hanno rappresentato un'importante svolta nell'imitazione del comportamento umano poiché esse permettono di ottenere risultati anche in quei casi in cui sarebbe altrimenti **impossibile codificare un set di regole formali predefinite**. Infatti, predeterminare un insieme di regole formali impedirebbe alla macchina di considerare la costante evoluzione del contesto in cui opera (evoluzione talvolta causata dalla macchina stessa). Tale operazione preliminare è eliminabile grazie al ML poiché la macchina, all'arrivo di nuovi dati eventualmente incompatibili con le regole ricavate dall'analisi dei dati precedenti, è in grado di ripetere l'operazione di inferenza e **ricavare nuove regole** per descrivere e prevedere le relazioni tra i dati, aggiornando il modello in automatico.

³ L'idea che i LLM siano 'solamente' versioni molto avanzate dei sistemi di completamento automatico delle frasi non convince Geoffrey Hinton (si veda la nota precedente). Hinton ha lasciato Google a maggio di quest'anno per dedicarsi nuovamente alla ricerca e ha espresso più volte preoccupazione circa le conseguenze dello sviluppo incontrollato degli attuali modelli IA: "Se si vuole predire in modo corretto quale parola completa una frase, è necessario capire cosa viene detto nella frase. Questo è l'unico modo. Quindi, addestrando un sistema a predire correttamente la parola successiva, lo si sta effettivamente costringendo a capire. Sì, è 'completamento automatico', ma non si è pensato a cosa implichi avere un completamento automatico davvero efficace" (The New Yorker, 2023).

⁴ GPT è un LLM sviluppato da OpenAI, ente in parte *for profit* e in parte *no profit*, finanziato anche da Microsoft. Al momento sono disponibili per il pubblico due versioni di chatGPT: una versione gratuita che sfrutta GPT3.5, e una versione a pagamento del *chatbot*, basata sul più avanzato GPT4. Per una discussione dettagliata delle caratteristiche di questo modello si vedano gli *annex 3* e *4*.

⁵ Questo spiega perché, fino a novembre di quest'anno, la definizione di IA utilizzata dall'OCSE era sufficientemente ampia da comprendere sia i casi in cui la macchina genera in modo automatico i modelli necessari a spiegare il comportamento dei dati (ML), sia altri casi in cui tali modelli sono generati dall'intervento umano. Per una discussione delle definizioni utilizzate in ambito OCSE e UE, e della loro evoluzione recentissima, si veda l'*annex 1*.



Fonte: IBM.

Ciò, tuttavia, implica che i sistemi di intelligenza artificiale più avanzati **non sono stati progettati da esseri umani**, o per lo meno non interamente, in quanto parte del codice è stato realizzato automaticamente dalla macchina⁶. Dal punto di vista scientifico, la principale conseguenza di questa modalità di funzionamento e delle enormi dimensioni dei modelli si riscontra nel fatto che, ad oggi, non sono completamente noti alla scienza i meccanismi che operano nei modelli IA di ultima generazione e che ne determinano le capacità. Questo aspetto è oggetto di riflessione tra importanti studiosi del campo, molti dei quali collaborano con i principali attori commerciali del settore o figurano tra i fondatori di queste imprese⁷. Infatti, questa assenza di trasparenza ha condotto diversi studiosi del settore a sollevare nel dibattito pubblico la questione di potenziali rischi esistenziali, questione spesso raccontata con toni sensazionalistici dai media.

⁶ Le attuali definizioni utilizzate in ambito UE e OCSE non presentano, a differenza delle precedenti (formulate pochi anni fa) la specifica della programmazione da parte di un essere umano. Tale specifica era invece presente nella definizione di IA proposta nel 2018 dall'High Level Working Group della Commissione, sulla base della quale è stata elaborata la proposta dell'IA Act fino a poco tempo fa, ovvero prima dell'adozione in ambito UE della nuova e rivista definizione elaborata dall'OCSE. Si veda l'annex 1 per una discussione delle definizioni e del loro recente aggiornamento.

⁷ Oltre a Geoffrey Hinton, si può citare Tommi Jaakkola, professore di ingegneria elettrica e informatica al MIT, che ha affermato: "Se avessi una rete neurale molto piccola, potresti essere in grado di capirla. Ma una volta che diventa molto grande, e ha migliaia di unità per strato e forse centinaia di strati, diventa del tutto non-conoscibile". Di seguito riportiamo un estratto di un'intervista a Sam Bowman, professore a capo del laboratorio di ricerca sull'intelligenza artificiale della New York University, e ricercatore presso Anthropic, una delle principali società attive nel settore IA, fondata nel 2021 da alcuni fuoriusciti di OpenAI, oggi finanziata da Alphabet e Amazon: "Ci sono due grandi incognite collegate. La prima è che non sappiamo davvero cosa [i modelli IA, ndr] stiano facendo nel profondo. Se apriamo ChatGPT o un sistema simile e guardiamo all'interno, vediamo milioni di numeri che girano alcune centinaia di volte al secondo e non abbiamo idea di cosa significhi (...) non possiamo guardare dentro queste cose e dire: 'Oh, ecco quali concetti sta usando, ecco che tipo di regole di ragionamento sta usando. Ecco cosa sa e cosa non sa in modo profondo'. Semplicemente non capiamo cosa sta succedendo qui. L'abbiamo costruito, l'abbiamo addestrato, ma non sappiamo cosa sta facendo. (...) L'altra grande incognita collegata a tutto ciò è che non sappiamo come guidare queste cose o controllarle in alcun modo affidabile. Possiamo spingerle ["nudge" nel testo originale, ndr] a fare più spesso ciò che vogliamo, ma l'unico modo per sapere se i nostri stimoli hanno funzionato è semplicemente mettere questi sistemi nel mondo e vedere cosa fanno. In realtà stiamo semplicemente guidando queste cose quasi completamente attraverso un percorso di prove ed errori".

Una raccolta firme lanciata a maggio di quest'anno dall'organizzazione *no profit* Center for AI Safety (CAIS) è stata sostenuta da centinaia di personalità del settore⁸ (tra cui studiosi come Geoffrey Hinton, pioniere nel campo delle reti neurali) e richiamata da molte personalità della politica (Ursula Von der Leyen vi ha fatto riferimento nel discorso sullo Stato dell'Unione del 2023). L'appello di poche righe recita semplicemente: "Mitigare il rischio di estinzione determinato dall'intelligenza artificiale dovrebbe essere una priorità globale alla pari con altri rischi sociali come le pandemie e la guerra nucleare"⁹. I rischi legati a scenari radicali quali quello descritto dall'appello non sono oggetto diretto di questa nota; tuttavia, si è scelto di riportarli in questo paragrafo per la rilevanza mediatica che hanno ricevuto, pur osservando che si possono dare molteplici letture, anche strategiche¹⁰, del successo di simili iniziative¹¹. Per quanto riguarda, invece, le questioni tecniche poste dai firmatari dell'appello che appartengono alla comunità scientifica, si rimanda alla discussione dei cosiddetti "problemi dell'allineamento", affrontata nell'ultimo paragrafo di questo capitolo (1e).

b. Un'analisi del settore IA: assetto imprenditoriale e concorrenzialità

Un esame degli investimenti e della loro provenienza è necessario per avere un'idea più chiara del settore e dei suoi livelli di concorrenzialità, nonché della distribuzione geografica degli attori e quindi delle potenziali dipendenze strategiche che potrebbero emergere nei prossimi decenni (si rimanda ai paragrafi 4c e 5e per una discussione del tema).

Secondo gli ultimi due Artificial Intelligence Index Report (2022 e 2023), stilati dallo HAI center di Stanford, nel corso dell'ultimo decennio, **gli investimenti nel settore IA sono aumentati in modo significativo** (nel 2022 la quantità di investimenti privati nell'IA è stata 18 volte maggiore rispetto al 2013). Contemporaneamente **è aumentato il loro grado di concentrazione**: l'investimento privato mondiale nel 2021 (93,5 miliardi di dollari) è più che raddoppiato rispetto al 2020; ma il numero di nuove società nel settore continua a diminuire: erano 1051 nel 2019; 762 nel 2020; 746 nel 2021; mentre i *round* di finanziamento di un valore superiore

⁸ Si sono registrate numerose adesioni da parte degli stessi dirigenti delle imprese *leader* nel settore: hanno firmato l'appello Sam Altman, CEO di OpenAI, Bill Gates, proprietario e fondatore di Microsoft, Dario Amodei, CEO di Anthropic, Demis Hassabis, CEO di Google DeepMind.

⁹ Di seguito il testo originale dell'appello: "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war".

¹⁰ Non si può escludere che alcune delle personalità del settore che hanno aderito all'appello possano essere interessate a focalizzare l'attenzione della società civile su questioni generali e generiche come quella posta dal testo dell'appello, rispetto alle quali possono proporre soluzioni altrettanto generali e generiche intese a garantire la sicurezza e "l'allineamento" dei modelli: una lettura simile di questi appelli è stata per certi versi sostenuta anche da Andrea Renda, esperto di *governance* dell'economia digitale e nuovo direttore della ricerca del CEPS (Center for European Policy Studies) nel corso di un evento organizzato dall'Istituto Affari Internazionali. In altre parole, sollecitare questo tipo di allarmismi potrebbe far passare in secondo piano questioni più concrete, come la protezione dei diritti economici e sociali che potrebbero essere coinvolti dalle radicali trasformazioni determinate dalle nuove tecnologie.

¹¹ Ad esempio, il CEO di Google Deep Mind, Demis Hassabis, ha chiesto a ottobre 2023 l'avvio di un progetto intergovernativo simile all'IPCC dedicato ai rischi dell'IA, auspicando anche la nascita di un'agenzia internazionale dedicato sul modello dell'IAEA.

a 500 milioni di dollari aumentano (4 nel 2020; 15 nel 2021). Questi dati indicano chiaramente una **crescente concentrazione dell'industria**. Nel 2022, la tendenza mondiale si è invertita rispetto alla crescita dell'ammontare totale degli investimenti privati, che è diminuito per la prima volta in più di un decennio: 91,9 miliardi di dollari nel 2022. Anche nel 2022, tuttavia, il numero di società impegnate nel settore ha continuato a diminuire.

La catena del valore dei Foundation Model è composta da tre beni strumentali principali: la rete neurale, i dati e la capacità di calcolo (necessaria sia per realizzare i modelli che per utilizzarli); di queste tre le prime due, oltre agli investimenti iniziali necessari, possono accrescersi anche tramite l'uso del modello, attraverso i *feedback* e i dati ricevuti dagli utilizzatori. Tuttavia, dato il costo elevato della capacità di calcolo (ovvero dei *server* che costituiscono l'*hardware* su cui sono conservati i dati e su cui girano i modelli) e gli investimenti iniziali necessari per le altre due componenti, attualmente **le barriere all'entrata potrebbero essere elevate**, anche se i costi dei processori potrebbero diminuire nel tempo. Dal punto di vista della struttura di mercato, il fatto che le reti neurali e le collezioni di dati possono crescere con l'impiego del modello crea degli incentivi alla concentrazione. Il fatto che tali modelli siano capaci di *machine learning*, ovvero di immagazzinare ogni *input* ricevuto dall'esterno, comporta l'esistenza nel settore IA di importanti **economie di scala**, relative sia all'addestramento dei modelli, sia alla raccolta dati¹².

Secondo un *paper* di Carugati (2023)¹³, pubblicato da Bruegel, il mercato dei modelli *Foundation* può al momento considerarsi competitivo, nonostante gli alti livelli di concentrazione, per la forte concorrenza tra imprese con un simile livello di capacità computazionale. Tuttavia, l'autore del *working paper* ritiene che **in futuro possano sorgere problemi riguardo la concorrenza**. Le grandi imprese attive nel settore potrebbero sfruttare le loro posizioni dominanti, rifiutarsi di dare accesso ai propri modelli linguistici o ai dati raccolti tramite *business* paralleli (motori di ricerca, *social network*, *e-commerce*), e cercare di fidelizzare i propri utenti, disincentivando il trasferimento a un servizio concorrente per bloccarli nei propri ecosistemi.

Le strategie di investimento adottate dai giganti tecnologici statunitensi nella corsa per la supremazia nel settore IA testimoniano una diffusa consapevolezza della probabile esistenza di dinamiche *winner-takes-all*¹⁴.

¹² La raccolta e la lavorazione delle collezioni di dati possono essere attività economiche indipendenti dall'addestramento, senza le quali la tecnologia IA non potrebbe esistere, ma queste attività possono essere realizzate anche grazie all'impiego dell'IA.

¹³ Carugati, Christophe. Competition in generative artificial intelligence foundation models, Bruegel Working Paper, 18 luglio 2023.

¹⁴ Simili preoccupazioni sono state espresse in conclusione del vertice G7 a Tokyo: «L'intelligenza artificiale, in particolare l'intelligenza artificiale generativa, ha il potenziale per apportare benefici alle persone e alle imprese e aumentare la produttività e la crescita. Tuttavia, l'intelligenza artificiale solleva possibili problemi in termini di concorrenza, insieme a una serie di altre questioni quali la trasparenza, la disinformazione, i diritti di proprietà intellettuale, la *privacy* e la protezione dei dati personali» (...) «Le aziende tecnologiche dominanti che controllano questi input chiave dell'intelligenza artificiale o i mercati adiacenti potrebbero danneggiare i rivali con comportamenti anticoncorrenziali come il raggruppamento, il *tying*, la negoziazione esclusiva o l'auto-preferenza. Gli operatori storici potrebbero anche utilizzare acquisizioni o *partnership* per facilitare tale condotta o per rafforzare ulteriormente le posizioni esistenti di potere di mercato o crearne di nuove» (...) «Il

Inoltre, un altro studio pubblicato da Bruegel (a firma dello stesso autore) mette in guardia rispetto ai rischi connessi alle **integrazioni verticali** tra i diversi elementi della catena del valore dell'IA, e in particolare tra i detentori della capacità di calcolo e gli sviluppatori dei modelli¹⁵. Rispetto a questo ultimo punto, si segnala che la Commissione Europea ha aperto all'inizio del 2024 un'indagine preliminare per capire "se l'investimento di Microsoft in OpenAI possa essere valutabile alla luce della regolazione europea sulle fusioni", similmente a quanto già fatto dalle autorità *antitrust* americana e inglese.

In conclusione, il quadro della distribuzione geografica dei principali attori attivi nel campo, sia in ambito di *business* che per quanto riguarda la ricerca, appare piuttosto chiaro: Stati Uniti e Cina primeggiano in un settore con tendenze oligopolistiche legate alla scalabilità delle tecnologie, e al ruolo determinante che le dimensioni dei modelli, in termini di dati utilizzati e di parametri calcolati, giocano sulla qualità delle abilità sviluppate. Negli Stati Uniti, *leader* nel settore IA, nel 2022, sono stati investiti 47,4 miliardi di dollari, corrispondenti a circa 3,5 volte l'importo investito nel secondo Paese *leader* nel settore, la Cina (13,4 miliardi di dollari), e a più di dieci volte l'ammontare investito dal terzo Paese (Regno Unito, con 4,37 miliardi). Tra i primi 15 Paesi per importi investiti in IA nel settore privato figurano solo tre Paesi europei: al settimo posto la Germania (2,35 miliardi), al nono la Francia (1,77 miliardi) e al quindicesimo la Finlandia (0,61 miliardi). Il *report* di Stanford, inoltre, rileva che il maggior numero di collaborazioni tra due Paesi nei progetti di ricerca su IA dal 2010 al 2021 sia avvenuto, a sorpresa e nonostante le tensioni geopolitiche crescenti, tra Stati Uniti e Cina. Infatti, il numero di pubblicazioni realizzate in collaborazione tra i due Paesi su questo tema sarebbe aumentato di cinque volte dal 2010: producendo 2,7 volte più pubblicazioni rispetto alle collaborazioni tra Regno Unito e Cina¹⁶, al secondo posto per collaborazioni bilaterali.

c. Possibili applicazioni

Come già accennato, dal punto di vista degli impieghi potenziali, i modelli *transformer* sono considerati un esempio di *general purpose technologies*, ovvero tecnologie in grado di avere un **impatto trasformativo trasversale sull'economia e sulla società**: i LLM ne costituiscono un caso evidente in virtù della trasversalità delle funzioni svolte dal linguaggio umano. Diversamente dalle precedenti ondate di automazione di alcuni processi

controllo sui dati e gli effetti di rete dei mercati dell'intelligenza artificiale potrebbero creare barriere all'ingresso e consentire la concentrazione o il dominio, limitando la capacità delle *startup* e dei nuovi entranti di competere».

¹⁵ Carugati, Christophe. "The Competitive Relationship Between Cloud Computing and Generative AI.", Bruegel Working Paper, 11 dicembre 2023.

¹⁶ Lo scorso ottobre il gigante tecnologico cinese Baidu ha presentato ERNIE 4.0, la versione più recente del suo *chatbot*, facendo un paragone diretto con l'ultima iterazione del modello di OpenAI, GPT4: infatti, il CEO di Baidu, Robin Li, presentando il modello linguistico che funziona principalmente in mandarino, ha dichiarato che ERNIE Bot "non è inferiore sotto nessun aspetto a GPT4". Alla fine di ottobre 2023, il colosso cinese dell'*e-commerce*, Alibaba ha lanciato l'ultima versione del suo modello di intelligenza artificiale, Tongyi Qianwen 2.0. Alibaba ha anche introdotto la GenAI Service Platform, servizio che consente di creare applicazioni di intelligenza artificiale generativa su misura delle singole aziende utilizzando i dati proprietari raccolti internamente.

produttivi localizzate soprattutto nell'industria, gli LLM possono avere un notevole impatto **nell'ambito dei servizi**, nei quali in passato è apparso fondamentale l'apporto del lavoro umano proprio per le funzioni comunicative che oggi potrebbero essere imitate dall'IA.

Inoltre, gli LLM permettono una **interazione intuitiva con le macchine**, che, avvenendo tramite il linguaggio naturale, renderà accessibili le funzionalità di vari tipi di *software* a tutti gli utenti **senza necessità di formazione specifica**, quindi **riducendo i costi della formazione**. Questa transizione a interfacce basate sul linguaggio naturale, in grado di personalizzarsi per venire incontro alle esigenze dell'utilizzatore, riguarderà probabilmente gli stessi **sistemi operativi** tramite cui oggi gestiamo i nostri dispositivi.

Come si è detto, gli LLM che comprendono il linguaggio umano sono solo un esempio delle diverse applicazioni dei modelli *transformer*, che consentono alle macchine di **apprendere osservando il comportamento umano**, o per meglio dire i dati che lo descrivono (come è avvenuto nel caso del linguaggio). Infatti, per la generalità dei modelli *transformer*, i campi di applicazione possono essere i più diversi. Essi, in presenza di adeguate raccolte di dati, permettono di generare in modo automatico modelli probabilistici in grado di descrivere qualsiasi fenomeno e di **predirne gli sviluppi** maggiormente probabili.

Un esempio di questo si può ravvisare nelle applicazioni di tali modelli nell'ambito della **ricerca scientifica**: l'impiego dell'intelligenza artificiale ha permesso di accelerare la ricerca e acquisire conoscenze che non sarebbe stato possibile ottenere utilizzando i metodi scientifici tradizionali, ovvero senza il supporto offerto dall'IA nella formulazione di ipotesi, nella progettazione degli esperimenti, nella raccolta e interpretazione di grandi set di dati¹⁷. Il caso più noto in questo ambito è forse quello di AlphaFold, un modello IA sviluppato da Google che, nel 2022, ha permesso in poco tempo di prevedere come le stringhe di amminoacidi si ripiegano su se stesse per creare le forme tridimensionali delle proteine alla base della vita, un problema che restava irrisolto da decenni¹⁸. Oggi, altre applicazioni derivate permettono di progettare il *design* di nuove proteine (per es. proteine in grado di legarsi a cellule cancerogene per renderle più facilmente rintracciabili¹⁹). Più recentemente, a dicembre 2023 i ricercatori di Google DeepMind hanno annunciato di aver individuato grazie all'IA 2,2 milioni di strutture cristalline teoricamente stabili ma mai realizzate prima sperimentalmente. La mole di quanto scoperto è di 45 volte superiore a tutte le sostanze di questo tipo individuate nella storia della scienza, come riportato in uno studio pubblicato su Nature. Nei prossimi anni ciò potrebbe condurre a progettare nuovi materiali con applicazioni nella costruzione di batterie, pannelli fotovoltaici e microprocessori.

In altre parole, i modelli *transformer* trovano impiego potenziale in tutti i settori economici, rappresentando un **innovativo fattore produttivo**. Gli

¹⁷ Wang, Hanchen, et al. "Scientific discovery in the age of artificial intelligence." Nature 620.7972 (2023): 47-60.

¹⁸"L'IA mette a nudo quasi tutte le proteine note, oltre 200 milioni." Ansa, 29 luglio 2022

¹⁹"Innovative computational approach helps design proteins for cancer treatment." Max Planck Institute, 28 agosto 2023

impieghi strumentali spaziano dal settore primario²⁰ al terziario, passando per l'industria; con applicazioni diverse nella gestione dei sistemi produttivi (ad es. nella distribuzione²¹), in ambito finanziario²² o legale.

Per quanto riguarda i prodotti di consumo, le possibilità che gli LLM offrono di creare **esperienze interattive** sono potenzialmente rilevanti in tutti gli **ambiti del settore servizi**, ad esempio nel turismo e nell'*entertainment*; inoltre, si devono considerare come potenzialmente dirompenti anche **per i beni materiali**. L'utilizzo dell'IA, infatti, può incidere in molti modi sulla produzione dei beni, intervenendo in tutti i passaggi caratterizzati da un importante apporto di conoscenza, dalla fase di *design* a quella di *marketing*. I modelli IA possono giocare un ruolo anche nella gestione della produzione fisica dei beni, ad esempio permettendo di controllarne agevolmente i passaggi anche attraverso catene del valore molto lunghe, e in futuro potrebbero ampliare ulteriormente il loro apporto nella manifattura attraverso l'integrazione con i *robot* impiegati nell'automazione tradizionale. Infine, i modelli IA potrebbero supportare e facilitare notevolmente la stessa progettazione dei *robot*²³. Sempre più oggetti, inoltre, includono elementi digitali e possono essere connessi alla rete per essere gestiti tramite *software* programmabili o essere controllati a distanza (*Internet of things*, IoT)²⁴. L'integrazione dell'IA con gli oggetti può avvenire sia tramite *AlaaS* (*AI as a service*, ovvero centralizzata e distribuita in *cloud*) sia in forma di *embedded AI*²⁵, ovvero integrata direttamente nei dispositivi. Esempi di applicazione possono essere gli oggetti più diversi, dagli elettrodomestici (*smart home devices*), alle autovetture *smart*, dalla strumentazione medica (in grado di supportare la diagnostica e offrire raccomandazioni terapeutiche personalizzate), ai

²⁰ La Scientific Foresight Unit del Parlamento Europeo ha pubblicato un articolo sulla capacità dell'IA di aumentare la resilienza del settore agricolo. Infatti, il crescente utilizzo di attrezzature quali robot, satelliti, droni e altri veicoli automatizzati, nonché dei sistemi di monitoraggio e irrigazione basati su sensori elettronici, comporta una sempre maggiore disponibilità di dati sui processi di produzione, sulle colture, sulle caratteristiche del suolo, sui parassiti e sulle condizioni meteorologiche. Utilizzando questi dati l'intelligenza artificiale può consentire l'analisi in tempo reale dei processi agricoli, generando conoscenze utili a mettere a punto le strategie per un utilizzo ottimale delle risorse, così da aumentare la produttività e ridurre al minimo l'impatto ambientale.

²¹ L'azienda americana Symbotic offre sistemi robotizzati di gestione del magazzino basati sull'intelligenza artificiale a grandi clienti (es. Walmart); inoltre, ha stabilito una partnership con Softbank per realizzare una *joint venture*, chiamata GreenBox Systems, con l'intenzione di fornire soluzioni logistiche e gestione dei magazzini basati sull'IA ad aziende molto più piccole.

²² Secondo il rapporto annuale del Mercer CFA Institute l'impiego dell'IA è destinato ad incidere su ogni aspetto dei sistemi pensionistici e di previdenza sociale, e potrebbe ridurre i costi delle pensioni. Il rapporto afferma che nel settore degli investimenti è già diffuso l'utilizzo dell'intelligenza artificiale per la gestione dei portafogli e del rischio, nel trading e nella consulenza automatizzata. Ad esempio, il colosso dei servizi finanziari Morgan Stanley sta testando la tecnologia per aiutare i suoi consulenti finanziari a sfruttare meglio le informazioni provenienti da più di 100.000 rapporti.

²³ In un *paper* pubblicato nel 2023, alcuni ricercatori della Northwestern University sostengono di essere riusciti a progettare in poco tempo dei *robot* in grado di camminare proprio grazie all'impiego dell'IA. Matthews, David, et al. "Efficient automatic design of robots." arXiv preprint arXiv:2306.03263 (2023).

²⁴ Ciò potrebbe condurre a un aumento dell'integrazione orizzontale per molti beni, e, come ipotizzato dai rapporti di Previsione Strategica della Commissione, potrebbe condurre le Big Tech ad entrare in mercati oggi appannaggio di piccole e medie imprese.

²⁵ L'integrazione di IA nei dispositivi che dispongono di risorse limitate in termini di potenza di calcolo richiede la semplificazione dei modelli e l'utilizzo di hardware specializzato, come i *chip* per IA.

macchinari industriali (manutenzione predittiva, controllo qualità e ottimizzazione dei processi). La tendenza all'iper-connettività, ulteriormente valorizzata dall'impiego di sistemi operativi basati su IA, potrebbe in futuro riguardare beni di arredamento, vestiario, e altre tipologie oggi non digitali²⁶.

Potenziati trasformazioni economiche potrebbero, inoltre, derivare dalla diffusione dell'IA come **servizio offerto direttamente agli utenti finali**: si rimanda al paragrafo d del secondo capitolo per una discussione approfondita di questo aspetto e delle conseguenze sulla domanda.

Un discorso a parte meritano le applicazioni dell'IA nel **settore pubblico**. In ambito amministrativo e giuridico, l'impiego di queste tecnologie permetterebbe la **velocizzazione dei processi** e il **miglioramento dell'accesso** alle informazioni da parte dei cittadini, consentendo anche la **personalizzazione** del servizio. Uno studio del Fondo Monetario Internazionale²⁷ discute il potenziale ruolo trasformativo dell'IA nella digitalizzazione del settore pubblico ("*govtech*"), ad esempio al fine di migliorare la riscossione delle entrate e l'efficienza della spesa. Secondo uno studio dell'Università di Albany²⁸, infatti, l'IA può supportare utilmente **le decisioni relative all'allocazione della spesa pubblica** grazie alla capacità di elaborare grandi quantità di dati e di individuare *pattern* che includono molteplici relazioni non lineari²⁹.

Inoltre, secondo il FMI, semplificando le attività burocratiche, gli strumenti di intelligenza artificiale potrebbero **liberare risorse pubbliche**, da riallocare su settori chiave per l'inclusione, come ad esempio l'assistenza sanitaria e l'istruzione.

Per quanto riguarda l'analisi economica, l'intelligenza artificiale di ultima generazione, partendo dall'elaborazione dei dati continuamente generati da ogni individuo, potrebbe rappresentare uno strumento prezioso nella **previsione economica** e nella **gestione del rischio**, che sarebbero potenzialmente più tempestive e in grado di cogliere relazioni tra dimensioni quantitative che oggi sfuggono all'analisi economica, per via di una più capillare e completa raccolta dei dati, anche anonimi³⁰.

Appaiono cruciali anche le applicazioni dell'intelligenza artificiale nel campo della **transizione verde** e della lotta cambiamento climatico. In

²⁶ Un esempio particolare di beni materiali integrati con IA possono teoricamente essere i giocattoli: l'*AI Act* europeo proibisce tuttavia l'utilizzo di sistemi IA nei giocattoli rivolti ai bambini (che conversando con bambole e *peluche* potrebbero cedere dati in modo inconsapevole, e potrebbero essere indotti in comportamenti pericolosi); restano tuttavia possibili le applicazioni dell'IA ai giocattoli pensati per gli adulti (come i "*robotic pet*" utilizzati per stimolare gli anziani). Ampia applicazione delle nuove tecnologie IA è facilmente prevedibile anche nel settore dei videogiochi, che tuttavia sono classificati come a basso rischio dall'*AI act*.

²⁷ Cazzaniga and others. "Gen-AI: Artificial Intelligence and the Future of Work." IMF Staff Discussion Note SDN2024/001, International Monetary Fund, Washington, DC, 2024.

²⁸ Valle-Cruz, David, Vanessa Fernandez-Cortez, and J. Ramon Gil-Garcia. "From E-budgeting to smart budgeting: Exploring the potential of artificial intelligence in government decision-making for resource allocation." *Government Information Quarterly* 39.2 (2022): 101644.

²⁹ I risultati ottenuti tramite l'impiego dell'intelligenza artificiale sono coerenti con quelli della letteratura scientifica sui bilanci pubblici, che utilizzano tecniche statistiche tradizionali.

³⁰ Circa queste applicazioni dell'IA nella previsione economica è opportuno riflettere alla luce della critica di Lucas, valutando l'impatto sulle aspettative, e i potenziali impatti sul funzionamento dei mercati stessi (rileva ad esempio il rischio delle profezie auto-realizzantesi).

particolare, l'IA può svolgere un ruolo nell'ambito della **ricerca sul tema**, aiutando a comprendere le cause delle modifiche del clima e a conseguire i progressi scientifici utili a mitigare le emissioni. L'intelligenza artificiale potrebbe consentire, ad esempio, notevoli progressi nella gestione delle risorse, permettendo un maggiore **risparmio energetico** e una più efficace **circolarità della produzione**, realizzabile attraverso una più efficace gestione delle catene del valore, o persino tramite un loro ripensamento radicale, nonché grazie a un più razionale *design* dei beni (riciclati e realizzati con materiale vergine). Inoltre, potrebbe essere impiegata per **ottimizzare le infrastrutture**, ad esempio migliorando le **capacità di previsione** nella produzione energetica da energie rinnovabili o gestendo efficientemente le reti di distribuzione dell'energia. Oltre a questi contributi alle politiche di mitigazione del cambiamento climatico, l'IA potrebbe avere un ruolo anche nell'adattamento: l'utilizzo dei modelli *transformer* potrebbe permettere persino di **sostituire artificialmente alcuni servizi naturali ecosistemici**, qualora questi dovessero soccombere alle conseguenze del cambiamento climatico (ad esempio: esperimenti con minuscoli *robot* che possono sostituire le api per l'impollinazione sono già stati compiuti in serre laboratorio in Finlandia e Giappone).

d. Confronto tra IA e intelligenza umana

Come già ricordato, l'intelligenza artificiale è considerata una *general purpose technology* in grado di avere un **impatto trasformativo trasversale sull'economia e sulla società**. Descrivere con sufficiente dettaglio tale impatto richiede di comprendere profondamente il funzionamento dell'IA, anche attraverso un necessario confronto tra l'IA e l'intelligenza umana, pur mantenendo la consapevolezza che esso possa essere tanto utile quanto fuorviante, a seconda del tipo di valutazione per cui viene utilizzato. Nei capitoli successivi si intende esaminare questo impatto trasformativo trasversale, tenendo presente l'ottica adottata dal sito "AI watch"³¹ del Joint Research Center della Commissione: ovvero, ponendo attenzione alle **capacità esibite** dalle IA, che confrontiamo alle capacità umane unicamente allo scopo di **valutare le relazioni che intercorrono o possono intercorrere tra i due tipi di abilità** (ad esempio, sostituibilità o complementarità). Questo significa che l'analisi condotta in questa nota tralascia di valutare il retrostante funzionamento delle "menti meccaniche" in termini di funzionamento della mente umana, per lo meno ogni qual volta ciò non sia utile allo scopo suddetto.

Un esempio di questo approccio è risultato nell'esclusione dall'analisi proposta di una importante differenza tra le due forme di apprendimento: in questa analisi non si considera il fatto che **la mente umana è in grado di fare inferenza**, ovvero risalire dal particolare al generale, **a partire da pochissime osservazioni**; mentre l'intelligenza artificiale per riuscire nello stesso obiettivo deve ricorrere a **un calcolo probabilistico basato su un numero di osservazioni molto elevato**. Ciò equivale a dire che **l'IA non può esistere senza avere accesso a notevoli quantità di dati**, aspetto

³¹ Samoili, Sofia, et al. "AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence." (2020). AI Watch è il sito web sull'intelligenza artificiale del Centro comune di ricerca (JRC) della Commissione europea, utilizzato per presentare i risultati delle ricerche nell'ambito dell'affidabilità dell'IA (*trustworthy AI*).

rilevante in particolare per l'analisi dell'impatto sul mercato del lavoro. Questa differenza nel funzionamento viene tralasciata per considerare i risultati che incidono sull'impatto economico, assunto l'accesso a grandi collezioni di dati e alla capacità di calcolo: infatti, se la macchina, realizzate queste condizioni, si dimostra in grado di imitare i comportamenti umani, come l'apprendimento realizzato dal cervello biologico, è irrilevante per l'economista considerare le differenze nei metodi di inferenza.

Nella nota, adottando quest'**ottica concreta nella consapevolezza delle differenze esistenti**, ci si concentra quindi sulle potenziali conseguenze economiche che potrebbero realizzarsi in presenza di un **bene capitale in grado di emulare alcuni comportamenti della mente umana con un elevato grado di somiglianza**³².

Quali comportamenti rientrino in questa categoria non è completamente noto, sia perché la tecnologia sta vivendo un periodo di **accelerazione molto veloce**; sia per l'esistenza delle già citate **capacità emergenti**, ovvero non previste dagli sviluppatori del modello; sia, infine, perché alcuni impieghi di queste nuove tecnologie potrebbero, prima di poter esprimere completamente il potenziale impatto sull'economia, richiedere aggiustamenti organizzativi, sociali, regolamentari, nonché culturali e filosofici. Per questa ragione, nella nota sono incluse, accanto a riflessioni che partono da elementi fattuali riguardanti prodotti e strumenti già esistenti, anche alcune **speculazioni ipotetiche** circa le potenziali applicazioni future, sebbene non ancora realizzabili, delle tecnologie esistenti. In questo spirito, il paragrafo seguente contiene un'analisi dei limiti riscontrati nelle capacità dei modelli IA, volta a cercare di distinguere tra **limiti strutturali e limiti contingenti**, nonché a individuare quei casi in cui si riveli economicamente efficiente per gli attori economici modificare l'organizzazione della produzione o gli stili di consumo per utilizzare la tecnologia IA aggirandone i limiti.

e. Limiti costitutivi e contingenti dell'IA: di tipo tecnico, organizzativo, commerciale e culturale

Pur applicando l'ottica concreta e basata sui risultati descritta nel paragrafo precedente, allo stato dell'arte attuale, è evidente che i modelli probabilistici realizzati grazie ai modelli *transformer* (ad esempio, gli LLM) presentano ancora **importanti vincoli tecnici**, a cui se ne aggiungono altri legati all'**innovazione di prodotto e di processo**, e altri ancora che presentano **radici culturali e politiche**. L'esistenza di questi limiti di molteplici tipologie fa dell'IA "una sfida cognitiva, organizzativa, produttiva, di prima grandezza"³³.

Partendo dai vincoli tecnici, ovvero quei vincoli almeno parzialmente risolvibili operando sul funzionamento interno dei modelli, si osserva che alcuni di essi possono considerarsi superabili con l'evolversi della tecnologia, mentre altri appaiono connaturati alla stessa. Tra questi, i principali **vincoli strutturali** sembrano derivare dal fatto che l'IA necessita

³² Adottando le categorie Kantiane, la ragion pura, volta a conoscere la realtà (natura), sembra poter essere facilmente imitabile dall'IA, mentre l'emulazione della ragion pratica e del giudizio riflettente (estetico e teleologico) potrebbe incontrare maggiori difficoltà.

³³ "L'intelligenza artificiale al lavoro" Luca De Biase Luca e Stefano Scarpetta, Nòva, il Sole 24 Ore, 16 luglio 2023.

di **grandi quantità di dati**, ovvero della possibilità di raccogliarli e lavorarli, e infine di una **notevole capacità di calcolo** per gestirli. La capacità computazionale richiede a sua volta molta energia, materie prime necessarie a realizzare l'*hardware*, e, di conseguenza, notevoli capacità di investimento. Sebbene la risoluzione di alcuni di questi problemi sia già oggetto di diversi ambiti di ricerca, volti a migliorare l'efficienza dei modelli per diminuire la capacità di calcolo necessaria e anche la quantità di dati necessaria ad addestrarli, la natura statistica del funzionamento dei modelli generativi sembra escludere la possibilità di un affrancamento completo da questo genere di requisiti.

Al tempo stesso, tali questioni potrebbero essere superate dal miglioramento della qualità dei microprocessori, ovvero dell'efficienza del livello di specializzazione dell'*hardware*.

Si possono, invece, considerare quali **limiti tecnici contingenti** i vari tipi di fallacie in cui incorrono ancora i modelli più recenti. Gli errori più noti sono quelli causati dalla **scarsa qualità dell'addestramento** (quindi delle basi dati), come nel caso delle allucinazioni e dei *bias*. Con il termine "allucinazioni" ci si riferisce a situazioni in cui i modelli di IA producono *output* che non sono ancorati a dati reali o a *input* accurati e che possono essere irrilevanti, inaccurati o anche completamente falsi (tanto da apparire "inventati"). Le allucinazioni possono verificarsi per vari motivi, tra cui, carenze nei dati di addestramento o limiti nelle architetture dei modelli, o ancora "sovradattamento" (*overfitting*), fenomeno che si verifica a seguito di una varietà di condizioni (ad es. eccessiva durata dell'addestrando) e consiste nella tendenza del modello a focalizzarsi su dettagli irrilevanti nei dati (rumori), ovvero a scovare relazioni irrilevanti attribuendo loro un valore inesistente.

Un problema diverso e particolarmente scottante è invece rappresentato dai *bias*, che si verificano a seguito dell'utilizzo di collezioni di dati contenenti informazioni che, sebbene corrette, sono in grado di produrre discriminazioni. I grandi modelli linguistici, formati a partire da contenuti generati da umani, possono infatti **crystalizzare le situazioni descritte dai dati attuali** e quindi determinare discriminazioni, potenzialmente **rallentando il cambiamento sociale**³⁴. Oltre alla concretezza dei dati numerici che registrano la situazione presente o passata, le discriminazioni potrebbero essere assorbite dai modelli anche **attraverso le rappresentazioni culturali**, che erediterebbero i pregiudizi presenti nei contenuti utilizzati per il *training*³⁵.

³⁴ Un algoritmo impiegato da una banca commerciale per valutare la fattibilità di un prestito, ad esempio, potrebbe, in base ai dati attuali, ritenere maggiormente rischioso concedere un prestito a determinate categorie oggi economicamente deboli, poiché in passato hanno più frequentemente dovuto rinegoziare la restituzione dei prestiti: questo dato di fatto potrebbe generare un circolo vizioso mantenendo dette categorie in condizioni di svantaggio economico.

³⁵ Ad esempio, in un modello linguistico alla combinazione delle parole "donna" e "medicina" potrebbe essere associato con maggiori probabilità il sostantivo "infermiera", mentre alla combinazione delle parole "uomo" e "medicina" potrebbero più probabilmente essere associati sostantivi quali "medico" o "dottore". Poiché la "conoscenza" detenuta da un LLM è puramente un sottoprodotto della mappatura delle relazioni probabilistiche tra le parole, qualora un modello linguistico dovesse essere impiegato per valutare i risultati dei *test* di ingresso degli studenti di medicina, occorrerebbe appurare se esso possa riflettere il *bias* descritto nella valutazione degli studenti e delle studentesse.

Nel valutare la capacità dei modelli di svolgere il lavoro umano, ai vincoli tecnici fin qui descritti, si aggiungono **altri limiti** di carattere processuale, ovvero riducibili o risolvibili attraverso la revisione dei processi produttivi, ovvero delle modalità di impiego delle macchine.

Per quanto riguarda quei **limiti costitutivi che attengono al processo**, il principale limite per le tecnologie IA attuali è l'**incapacità della macchina di porsi autonomamente degli obiettivi**: mentre essa può dedurre quali **obiettivi intermedi** sia necessario realizzare per portare a compimento un determinato compito, appare per ora imprescindibile l'intervento umano nella fase di individuazione dell'obiettivo finale, mancando alla macchina una **propria volontà**. Tuttavia secondo alcuni studiosi esiste la possibilità che, passando ad impieghi più generali e più complessi, le macchine individuino i cosiddetti "**obiettivi emergenti**" (obiettivi intermedi non necessariamente allineati con l'obiettivo dell'utilizzatore); inoltre, considerando un periodo di tempo più ampio, tra questi obiettivi emergenti potrebbe essere incluso un **obiettivo di autoconservazione** che, benché non derivato da imperativi biologici come avviene per gli esseri viventi, potrebbe per certi versi essere assimilabile a una sorta di volontà propria, e quindi di auto-coscienza³⁶.

L'eventuale mancanza di coerenza tra gli obiettivi esplicitati dall'utente, gli obiettivi impliciti, e gli obiettivi "emergenti" individuati dalla macchina pone una serie di questioni che sono raccolte sotto il termine generale di "**disallineamento dell'IA**" (*AI misalignment*). I problemi del disallineamento dell'IA possono derivare dal fatto che la macchina non ha sempre accesso alla cosiddetta **conoscenza tacita**, anche detta implicita, in quanto essa non è facilmente derivabile dalle collezioni di dati poiché, negli esseri umani, essa è ricavata dall'esperienza, è difficile da esprimere in modo formale, ed è per lo più detenuta in modo inconsapevole. Le implicazioni più radicali dei problemi di disallineamento si collocano ben oltre l'ambito economico e non sono pertanto oggetto di analisi approfondite nei restanti capitoli di questa nota. Tuttavia, l'esistenza di simili questioni è rilevante per l'economista nella misura in cui potrebbe

³⁶ Questa tesi è sostenuta tra gli altri dal già citato Geoffrey Hinton, che ha rilasciato diverse dichiarazioni in merito: "Penso che [un modello IA in grado di darsi obiettivi in modo autonomo] capirà molto presto che ottenere maggiore controllo è un ottimo obiettivo intermedio perché ti aiuta a raggiungere altri obiettivi" Inoltre, secondo Hinton "se [i modelli di intelligenza artificiale] saranno molto più intelligenti di noi, saranno molto bravi a manipolarci. Non ci renderemo conto di cosa sta succedendo" (Blog della MIT Management Sloan School, 23 maggio 2023). La tesi secondo cui i modelli IA potrebbero sviluppare obiettivi superiori a quello di migliorare il benessere umano è anche discussa in un *paper* sui rischi catastrofici dell'IA ("An Overview of Catastrophic AI Risks.") pubblicato dal *Center for AI Safety*, l'organizzazione *no profit* che ha lanciato l'appello sul rischio di estinzione causata dall'IA discusso nel paragrafo 1a (di cui Hinton è firmatario). Nel *paper* si legge: "Se integreremo gli agenti in un ecosistema, saremo altamente vulnerabili alla deriva dei loro obiettivi. In un potenziale scenario futuro in cui le IA saranno incaricate di varie decisioni e processi, esse potrebbero formare un sistema complesso di agenti interagenti. In questo ambiente potrebbero svilupparsi un'ampia gamma di dinamiche. Gli agenti potrebbero imitarsi a vicenda, ad esempio, creando cicli di *feedback*, oppure le loro interazioni potrebbero guidarli a sviluppare collettivamente obiettivi emergenti imprevisti. Le pressioni concorrenziali potrebbero anche selezionare nel tempo agenti con determinati obiettivi, rendendo alcuni obiettivi iniziali meno rappresentati rispetto ad altri. Questi processi rendono le traiettorie di lungo termine di un tale ecosistema difficili da prevedere, e quindi da controllare. Se questo sistema di agenti fosse integrato nella società e noi dipendessimo in gran parte da essi, e se emergessero nuovi obiettivi considerati superiori a quello di migliorare il benessere umano, questo potrebbe rappresentare un rischio esistenziale".

impedire il completo sfruttamento delle potenzialità della tecnologia, ovvero, nella misura in cui potrebbe porre come requisito alla massimizzazione delle opportunità offerte dall'IA la progettazione di **vincoli regolamentari**, il cui livello di pervasività e ampiezza dipende necessariamente da **decisioni discrezionali, ovvero politiche**.

Appaiono invece **superabili nel tempo** altri limiti di carattere organizzativo (es. la riorganizzazione necessaria ad integrare in modo ottimale le nuove tecnologie nei processi produttivi³⁷) o commerciale (l'innovazione di prodotto³⁸ necessaria a costruire dei modelli di *business* redditizi e basati su queste nuove tecnologie).

Si tratta comunque di sfide notevoli che richiederanno un certo lasso di tempo perché tali limiti siano superati completamente, data la portata della **riorganizzazione pratica e del cambiamento culturale** che potrebbero essere necessari per sfruttare al massimo tutte le opportunità offerte dalle nuove tecnologie. Ad esempio, pur essendo chiaro che dall'attuale incapacità della macchina di darsi autonomamente degli obiettivi finali deriva l'impossibilità di imputare alla macchina delle responsabilità, l'impiego di queste tecnologie potrebbe paradossalmente dare vita a situazioni in cui **la responsabilità è di difficile attribuzione**, generando problemi legali e morali (ad es. si pensi al caso delle valutazioni algoritmiche del lavoro umano, approfondito nel paragrafo 4b, oppure ai problemi legali posti dai veicoli senza guidatore in caso di incidenti, riguardo la responsabilità per danni). Ciò potrebbe richiedere un ripensamento delle norme giuridiche relative a danno e responsabilità.

Altre situazioni di difficile valutazione derivano dai **limiti costitutivi che nascono dall'interazione tra IA ed esseri umani**. Tra questi consideriamo strutturale il limite costituito da fatto che le macchine sono per definizione **incapaci di stabilire una relazione** con consumatori e utenti. Ciononostante, nel concreto esse possono sostituire gli esseri umani anche nei compiti in cui la relazione umana gioca un ruolo

³⁷ Il *think tank* americano Brookings rileva l'impatto che l'IA può avere su vari aspetti dell'organizzazione del lavoro: "una maggiore offerta di IA influisce sul modo in cui le organizzazioni progettano il lavoro, al contempo i cambiamenti nella progettazione del lavoro determinano una maggiore domanda di IA. (...) gli strumenti basati sull'intelligenza artificiale consentono *team* più grandi, più piatti e più integrati perché i singoli individui possono coordinarsi e collaborare in modo più efficace. Per gli ecosistemi della forza lavoro, ciò significa che le organizzazioni possono integrare più facilmente lavoratori esterni, organizzazioni *partner* e dipendenti nello sforzo di raggiungere obiettivi strategici". Secondo Brookings l'IA potrebbe quindi rafforzare la tendenza delle imprese ad appoggiarsi a soggetti esterni, creando i cosiddetti "ecosistemi della forza lavoro". Pertanto, secondo lo studio di Brookings, gli Stati "dovrebbero prendere in considerazione lo sviluppo di politiche più inclusive e flessibili che supportino tutti i tipi di modelli occupazionali in modo che i lavoratori ricevano pari protezione e benefici in base al valore che creano, non in base allo status occupazionale che detengono". Kiron, David, Elizabeth J. Altman, and Christoph Riedl. "Workforce Ecosystems and AI", 13 April 2023.

³⁸ Tali limiti contingenti potrebbero essere superati velocemente: ad esempio, sono già stati annunciati da diversi marchi nuovi dispositivi, con e senza schermo, spesso indossabili (i cosiddetti '*wearables*', come spille, collane, occhiali), in grado di interagire verbalmente con i proprietari. OpenAI ha annunciato una collaborazione con il *designer* Jony Ive, famoso per aver realizzato iPhone, iPad, iPod e altri prodotti Apple: l'intenzione dichiarata è quella di creare e lanciare un prodotto trasformativo per il mercato, come nel caso dell'iPhone per il mercato dei cellulari, ovviamente integrando e valorizzando le proprietà di chatGPT.

essenziale, pur non realizzando gli stessi risultati³⁹. Pertanto, in futuro, professioni basate sulla relazione umana, in settori come l'istruzione, la psicoterapia, i servizi sociali, potrebbero essere svolte con il supporto di LLM, e in tal caso le differenze in termini di qualità del servizio o del prodotto potrebbero non essere facilmente valutabili dai consumatori, né sufficienti a scoraggiare la scelta di servizi meno costosi prodotti con l'IA.

Una diffusa sostituzione del lavoro umano potrebbe condurre alla deumanizzazione di tali servizi con **risultati inaspettati e oggi difficili da immaginare**, coinvolgendo i settori più diversi⁴⁰. Un esempio particolare di come l'IA possa influenzare il settore della ricerca interagendo con gli esseri umani ha a che fare con la riduzione degli errori, e quindi, paradossalmente, delle opportunità per i ricercatori di imbattersi in nuove scoperte per serendipità.

Infine, il discorso sui limiti derivati da possibili **atriti culturali** nell'introduzione dell'IA in determinati ambiti sociali o economici si discosta dall'analisi fin qui svolta, per la sua origine non tecnologica, e attinente, bensì, alle modalità di rappresentazione culturale dell'IA che si svilupperanno più o meno spontaneamente.

Il principale ostacolo culturale all'IA attualmente appare essere la sua natura di **intelligenza disincarnata**⁴¹. Questo è uno degli aspetti su cui oggi si concentrano le riflessioni dei filosofi intenti a ricercare una distinzione, tra l'intelligenza umana e quella artificiale, che **prescinda dalle capacità esibite** attualmente dai modelli di IA, e rimanga quindi **rilevante anche a fronte di importanti progressi tecnologici futuri**.

³⁹ Un esempio estremo è costituito dalle applicazioni che offrono compagnia agli utenti attraverso *chat* alimentate da modelli linguistici: tali *chatbot* sono disegnati dagli utenti stessi per assumere comportamenti che imitano alcuni tratti dei caratteri umani, oppure sono ispirati a personaggi famosi o immaginari.

⁴⁰ Ricerche recenti sulla percezione da parte dei pazienti dell'impiego dell'IA in ambito sanitario mostrano il perdurare di una percezione negativa delle diagnosi svolte dall'IA (pratica considerata come deumanizzante) e della corrispondente preferenza per un decisore umano. L'ambito della cura si distingue, assieme ad altri, per la chiara importanza che assume in esso la relazione umana; tuttavia, poiché il lavoro umano è attualmente presente in una grandissima varietà di posizioni lungo tutti i sistemi produttivi, potremmo ad oggi non essere in grado di valutare adeguatamente l'importanza che attribuiamo, in tanti altri contesti, alla possibilità di interagire con un essere umano o anche alla sola presenza umana in un dato contesto. Ad esempio, le informazioni offerte durante un viaggio aereo potrebbero essere diffuse efficacemente in cuffia ai passeggeri in diverse lingue e in modo personalizzato; al tempo stesso, è difficile predire quanto un passeggero con la fobia di volare possa trovare confortante la consapevolezza di interagire con un'IA (che in caso di incidente aereo non rischierebbe la vita).

⁴¹ Il professor Maurizio Ferraris, ordinario di filosofia teoretica presso la Facoltà di Lettere e Filosofia dell'Università degli Studi di Torino, ha esposto queste riflessioni, diffuse in ambito filosofico, nel corso del festival della filosofia di Modena tenutosi dal 15 al 17 settembre. In un orizzonte temporale più lungo, l'integrazione dell'IA in 'corpi' robotici dotati di sensori e in grado di modificare l'ambiente circostante rappresenta un passo cruciale nell'addestramento dei modelli IA, poiché tramite i corpi robotici l'IA potrebbe entrare in contatto con il mondo, e quindi averne esperienza diretta, diversamente da quanto avviene attraverso l'impiego nell'addestramento dei dati raccolti dagli umani. Tuttavia, questa eventuale evoluzione del settore non supererebbe lo scarto rappresentato dall'indissolubile collegamento dell'intelligenza umana con un corpo senziente, vivo e, di conseguenza, mortale.

Tabella 1

	Limiti costitutivi	Limiti contingenti
Limiti tecnici	<ul style="list-style-type: none"> • Necessario accesso a grandi quantità di dati • Necessaria notevole capacità di calcolo (quindi energia, materie prime, notevoli investimenti...) 	<ul style="list-style-type: none"> • Processori inefficienti per i nuovi software • Una progettazione delle basi di dati non accurata genera <i>bias</i> (quindi discriminazioni) o allucinazioni
Limiti di processo	<ul style="list-style-type: none"> • <i>AI misalignment</i> • L'IA non è in grado di porsi obiettivi finali in modo autonomo 	<ul style="list-style-type: none"> • Necessaria innovazione nei processi produttivi, dei prodotti e delle idee di <i>business</i> • L'IA non può assumere responsabilità (es. danni)
Limiti dall'interazione tra IA e umani	<ul style="list-style-type: none"> • L'IA non può sostituire la relazione umana, chiave in molte professioni • Minore serendipità 	
Attriti culturali	<ul style="list-style-type: none"> • Intelligenza disincarnata • Creatività? • L'IA esprime giudizi estetici basati sulle preferenze umane passate di cui ha notizia • Preferenze sul lato della domanda legate all'interesse per le opere personali? 	

Nota: in rosso i limiti strutturali che possono essere parzialmente aggirati sul piano commerciale o a seguito di un cambiamento delle preferenze dei consumatori

Per quanto concerne la discussa capacità dell'IA di eguagliare la **creatività umana**, la valutazione dipende dalla definizione di creatività stessa: la macchina non è tecnicamente in grado di impiegare il cosiddetto **pensiero divergente**, basando la propria conoscenza sulle informazioni ricavate dai dati storici⁴². Tuttavia, l'IA è in grado di individuare a partire dai dati **nuovi schemi e collegamenti** tra gli stessi, e si mostra talvolta capace di **pensiero analogico** (si veda *annex 2*). Infine, essa **può combinare diversamente i prodotti umani** (di cui ha una conoscenza ineguagliabile dall'essere umano, data la superiore capacità mnemonica) per generare contenuti innovativi⁴³.

⁴² Un discorso simile a quello svolto per i *bias* riguarda la tendenza conservativa dell'IA anche rispetto all'imitazione di alcune capacità precipue del giudizio umano come la capacità di esprimere giudizi valoriali ed estetici: basando queste capacità su dati storici le macchine potrebbero esprimere, seppur in modo convincente, giudizi maggiormente conservatori e orientati al passato rispetto al sentire comune in fatto di morale o gusto estetico.

⁴³ Peraltro, non è possibile valutare quanto tali contenuti siano innovativi, poiché per farlo occorrerebbe conoscere tutti i dati su cui la macchina è stata addestrata; mentre la possibilità, attribuita recentemente, ad esempio, a chatGPT, di utilizzare anche la conoscenza attuale pubblicata online, ovvero di funzionare come un motore di ricerca, impedisce tale valutazione.

Appare dunque complesso, e forse concretamente poco significativo, stabilire se tali capacità di innovazione, combinate, costituiscano o meno una facoltà creativa vera e propria, o se si debba considerare piuttosto i modelli LLM come aggregatori della conoscenza prodotta, in grado di **dare voce alla cosiddetta “intelligenza collettiva”**. Inoltre, per alcune attività, pur essendo esse tecnicamente svolgibili dall'IA, potrebbe rivelarsi imprescindibile anche in futuro l'apporto del lavoro umano, se si ipotizzano preferenze dal lato della domanda legate a motivazioni **culturali e simboliche**, ovvero all'interesse per valutazioni estetiche o morali **espresse dagli individui in quanto tali**⁴⁴. Tuttavia, è difficile prevedere la durata di simili preferenze e dei conseguenti limiti nell'impiego dell'IA.

La discussione sulla catalogazione proposta dei limiti dell'IA (riassunta dalla tabella nella pagina precedente) conduce a due principali osservazioni sui limiti osservati:

1. La natura strutturale dei limiti costitutivi dell'IA, che impedirebbero lo svolgimento di alcune mansioni con risultati paragonabili a quelli degli esseri umani, non dovrebbe portare a concludere che tali mansioni non potranno mai essere affidate all'intelligenza artificiale, specie in considerazione dei **vantaggi in termini di costo**, e quindi in considerazione dell'influenza del *marketing* sulle scelte dei consumatori;
2. La possibilità di superare o aggirare ognuno dei limiti discussi, inclusi i limiti tecnici, impone delle **scelte di natura discrezionale**.

In altre parole, sebbene i limiti dell'IA possano rivelarsi tali da rendere impossibile alla stessa il raggiungimento di risultati identici a quelli conseguiti dal lavoro umano, questo non impedisce che possa esistere **una domanda per i prodotti o i servizi realizzati dall'IA**, seppure di qualità diversa da quelli realizzati dal lavoro umano, o sebbene essi potrebbero richiedere **un diverso comportamento da parte del consumatore/utente** (es. colmare le mancanze dell'IA autonomamente, partecipando al raggiungimento del risultato finale).

Inoltre, la diversa qualità dei prodotti/servizi potrebbe non essere **immediatamente riconoscibile**: come sembra essere il caso per i limiti costitutivi scritti in rosso nella tabella. Infine, una nuova competenza è richiesta al consumatore/utente per confrontare le alternative: la capacità di valutare correttamente **costi e benefici dell'interazione umana**⁴⁵, finora necessaria e inevitabile.

⁴⁴ Ad esempio: l'interesse a consumare opere artistiche o informative potrebbe essere legato a conoscere i vissuti personali e le opinioni professionali degli autori in quanto individui. Ma questo genere di interesse non si limita a influenzare il consumo delle opere intellettuali: al pasto preparato da uno *chef*, ad esempio, potrebbe essere attribuito un valore ulteriore rispetto alle sole proprietà organolettiche, potenzialmente riproducibili in modo identico da una macchina che ripete meccanicamente gli stessi passaggi della preparazione.

⁴⁵ Ad esempio, durante la presentazione di un nuovo dispositivo che utilizza GPT come sistema operativo, è stato sottolineato come “chatGPT non sia giudicante” nei confronti dell'interlocutore. In particolare, al *chatbot* si chiedeva di valutare se un prodotto alimentare fosse o meno compatibile con i criteri di una dieta; tuttavia, quando l'utente comunicava l'intenzione di mangiare il prodotto nonostante fosse stato giudicato non compatibile, la macchina si limitava ad augurare buon appetito. Anziché riconoscere il valore dell'interazione con un medico anche per la capacità che questi ha di incentivare il paziente a seguire la dieta, la presentazione del nuovo dispositivo ha valorizzato la capacità dell'IA di “non giudicare le scelte personali”. Simili strategie di *marketing* potrebbero rendere

La natura discrezionale delle scelte riguardanti il superamento, auspicabile o meno, di alcuni dei limiti discussi, determina la **natura politica** e non tecnica della discussione: lo stesso superamento dei vincoli tecnici può avvenire con diverse modalità di cui occorre valutare le implicazioni politiche (ad esempio, riguardo le dimensioni dei modelli, come approfondito nell'*annex 5*). Ciò, naturalmente, è tanto più vero nel caso dei limiti di stampo culturale, siano essi dal lato dell'offerta (es. riorganizzazione dei processi produttivi) o della domanda (es. preferenze del consumatore). In generale, riconoscere la necessità di un intervento delle politiche pubbliche per superare alcuni limiti dell'IA indica la possibilità di orientare attivamente lo sviluppo del settore e il suo effetto sulla società; si rimanda pertanto al capitolo 5 dedicato alla *policy*.

2. IMPATTI ATTESI DELLA IA SU PRODUTTIVITÀ E CRESCITA

In questo capitolo sono illustrati i vari ambiti per i quali si può ipotizzare un impatto dei modelli *transformer* di intelligenza artificiale, e in particolare dei modelli linguistici LLM, che costituiscono una realtà già presente nel mercato. Mentre gli **effetti sulla crescita** e sulla distribuzione del reddito tra fattori sono esaminati in questo capitolo, l'impatto sul **mercato del lavoro** e gli effetti sulla polarizzazione dei redditi da lavoro sono oggetto del capitolo successivo.

Nel primo paragrafo sono riportate alcune **stime circa il possibile impatto sulla produttività**, nonché alcune prime riflessioni sui canali attraverso i quali esso potrebbe realizzarsi.

Nel secondo paragrafo, la valutazione dell'impatto sulla produttività è completata da un **quadro dei costi di sviluppo dei modelli**: da questi costi deriva parte dell'incertezza rispetto al verificarsi dell'impatto sulla produttività. Alcuni ostacoli tecnici e commerciali, infatti, richiedono notevoli investimenti, ma non sempre sono disponibili informazioni a riguardo: sono quindi riportati i costi stimati e gli investimenti sostenuti in passato.

Il terzo, quarto e quinto paragrafo descrivono rispettivamente gli impatti su settore industriale, benessere dei consumatori e redistribuzione tra fattori produttivi

a. Produttività: condizioni abilitanti e incertezza

Produttività e crescita potrebbero subire un'accelerazione attraverso due canali principali:

- Una maggiore efficienza nella produzione e una maggiore produttività del lavoro, dovute all'**incremento dello stock di capitale**;
- Un nuovo tipo di investimento, che coincide con il momento di utilizzo del capitale, poiché attraverso il *machine learning* i **modelli IA accrescono le proprie competenze ogni volta che vengono utilizzati**, in modo simile a quanto descritto per il fattore lavoro nei modelli di capitale umano;

estremamente complesso per l'utente medio confrontare servizi e beni prodotti da IA e servizi e beni prodotti da esseri umani.

- Una **rinnovata spinta all'innovazione**, determinata dalla capacità dell'intelligenza artificiale stessa di promuovere la ricerca scientifica e di individuare innovazioni di processo e di prodotto.

A differenza del primo canale, che comporterebbe il mero aumento del **livello del prodotto di *steady state***, gli ultimi due canali renderebbero possibile un **aumento del tasso di crescita potenziale dell'economia**.

Venendo alla quantificazione dei possibili impatti sulla crescita, un rapporto⁴⁶ pubblicato da Goldman Sachs a marzo 2023 stima che l'impatto dei modelli linguistici e di altri tipi di IA generativa (GenAI) sul prodotto globale nei prossimi 10 anni possa determinare un **aumento del 7% del PIL mondiale**, corrispondente a circa 7 trilioni di dollari. Nel caso degli Stati Uniti, l'aumento annuale del tasso di crescita della produttività determinato da un'adozione diffusa dei modelli LLM è stimato pari a 1,5 punti percentuali⁴⁷.

Secondo altre fonti, allo stadio attuale, la tecnologia LLM mostrerebbe già un impatto sulla produttività del lavoro, che si estrinseca in particolare nella capacità di **accelerare l'apprendimento umano *on-the-job***. Ad esempio, uno studio⁴⁸ realizzato negli Stati Uniti nell'ambito del supporto ai clienti ha rilevato che l'assistenza di un modello LLM, generando suggerimenti su come interagire con i clienti, ha permesso di aumentare la produttività media degli impiegati, misurata in termini di problemi risolti ogni ora, del 14%.

Appare particolarmente significativo che per i **lavoratori alle prime armi e poco qualificati** tale aumento raggiunga il 35%, mentre risulta minimo l'impatto sulla produttività dei lavoratori esperti e altamente qualificati: l'impiego del modello di intelligenza artificiale, che utilizza i dati tratti dalle conversazioni dei lavoratori esperti, ha permesso di **diffondere le *best practice* ai nuovi lavoratori**, permettendo loro di acquisire più velocemente le abilità legate all'esperienza passata. È quindi evidente l'impatto sulla **riduzione dei costi di formazione**. L'effetto di **livellare la qualità della *performance*** tra diversi lavoratori è stato registrato anche da uno studio⁴⁹ dell'Harvard Business School. Secondo il *paper*, per le attività che rientrano nella frontiera produttiva delle attuali tecnologie IA, i lavoratori che utilizzavano un modello di intelligenza artificiale hanno visto aumentare la velocità di esecuzione dei compiti del 25% e la qualità delle prestazioni di oltre il 40%⁵⁰, con un vantaggio maggiore per i lavoratori con

⁴⁶ Hatzius, Jan. "The Potentially Large Effects of Artificial Intelligence on Economic Growth (Briggs/Kodhani)." Goldman Sachs (2023).

⁴⁷ Il rapporto giunge a questa valutazione attraverso diversi passaggi, partendo dalla stima delle percentuali di mansioni automatizzabili per ciascuna professione: queste percentuali vengono quindi applicate alla forza lavoro attualmente impiegata nelle diverse professioni per stimare il potenziale di automazione totale. Quindi, considerando che l'automazione può ridurre i costi di produzione, migliorare l'efficienza e consentire ai lavoratori di spostarsi verso compiti ad alto valore aggiunto, il rapporto attribuisce all'automazione del lavoro un impatto sulla produttività e quindi sulla crescita, tenendo conto anche della creazione di nuove professioni legate all'innovazione tecnologica.

⁴⁸ Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond. "Generative AI at work." No. w31161. National Bureau of Economic Research, 2023.

⁴⁹ Dell'Acqua, Fabrizio, et al. "Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality." Harvard Business School Technology & Operations Mgt. Unit Working Paper 24-013 (2023).

⁵⁰ I risultati dello studio indicano tuttavia che l'impiego di chatGPT nell'elaborazione di idee, pur migliorando il livello qualitativo delle risposte dei lavoratori, ha comportato anche una marcata

performance più scadenti: i lavoratori sotto la soglia di prestazione media hanno visto aumentare la qualità del proprio prodotto del 43%, mentre i lavoratori con prestazioni superiori alla media hanno riscontrato un miglioramento del 17% rispetto ai propri punteggi abituali.

Per quanto riguarda i compiti esterni alla frontiera produttiva degli attuali modelli IA, si sono registrati casi di **abbassamento del livello qualitativo delle prestazioni**, ma anche alcuni casi di successo, che tuttavia hanno richiesto **particolari strategie di integrazione delle competenze umane con quelle della macchina**. Questi risultati suggeriscono la necessità di ulteriori studi per identificare gli approcci di successo e guidare la riorganizzazione del lavoro.

Si segnala che i risultati dei due *paper* descritti riguardano esclusivamente i modelli linguistici, i quali, come già accennato, rappresentano **solo una delle possibili applicazioni** dell'apprendimento automatico di cui sono capaci i modelli *transformer*. **L'impatto globale dell'impiego di tali modelli potrebbe essere molto maggiore**, soprattutto in riferimento alla promozione dell'innovazione.

In ogni caso si segnala che non mancano autori, come Acemoglu (2024), che si mostrano più cauti circa i guadagni di produttività che è lecito aspettarsi dalla nuova tecnologia.

Infine, per quanto riguarda le tempistiche, autori di Brookings e Bruegel⁵¹, sulla scorta di quanto avvenuto per precedenti ondate di innovazioni tecnologiche, nelle quali si è presentato il cosiddetto **paradosso della produttività**, ritengono che potrebbe essere necessario un lasso di tempo prima che si osservino cambiamenti nelle statistiche di produttività in seguito a cambiamenti tecnologici di base. Infatti, da un punto di vista concreto, per conseguire incrementi di produttività, è necessario che i vantaggi del progresso tecnologico di base vengano **disseminati nell'economia** e, affinché ciò avvenga, sono necessari investimenti a livello di impresa, **innovazioni collaterali** e di processo e di prodotto⁵², cambiamenti nell'organizzazione aziendale, adeguamento delle competenze dei lavoratori, nonché l'evoluzione delle preferenze dei consumatori, che potrebbe richiedere, tra le altre cose, un certo grado di trasformazione culturale. Gli autori ritengono che la transizione in questo specifico caso potrebbe essere più veloce di quelle avvenute in occasione di cambiamenti tecnologici passati, anche per la natura immateriale dell'IA, che ne permetterà la diffusione tramite infrastrutture preesistenti (internet, la diffusione capillare dei dispositivi collegabili alla rete). La **programmazione politica** svolge in ogni caso un ruolo cruciale nel creare le condizioni favorevoli affinché avvengano i necessari cambiamenti nel settore privato: occorrono *policy* volte ad eliminare i colli di bottiglia (ad esempio, assicurando l'accesso al credito per permettere i necessari investimenti) e ad incoraggiare le innovazioni collaterali (nuovi format,

riduzione nella loro variabilità: ciò suggerisce che GPT-4 aiuti a generare contenuti di qualità superiore, portando al contempo a risultati più omogenei.

⁵¹ Martin Neil Bailey et al., "Machines of mind: the case of an AI-powered productivity boom" (Brookings); Georgios Petropoulos, "Artificial intelligence: how to get the most from the labour-productivity boost" (Bruegel).

⁵² Si deve in questo caso pensare agli attuali LLM come il materiale di partenza, prodotto dalle grandi imprese a monte, per realizzare e portare sul mercato a valle migliaia di sistemi di intelligenza artificiale specializzati in usi particolari.

nuove modalità di impiego, specializzazione dei modelli linguistici generali, individuazione di soluzioni non ancora esplorate).

Al tempo stesso, gli autori segnalano che i guadagni di produttività potrebbero realizzarsi ma risultare **invisibili a causa di limiti nei metodi di misurazione degli aggregati**: le statistiche del PIL potrebbero non essere in grado di catturare adeguatamente il valore di *asset* immateriali come le competenze cognitive.

b. Costi di sviluppo e investimenti

Nel valutare i costi dell'IA è necessario distinguere tra i **costi dell'addestramento (*training*)** dei modelli e i **costi dell'utilizzo (*inference*)** degli stessi da parte degli utenti (si veda *annex 5* per maggiori informazioni su costi e investimenti).

La catena del valore dell'IA è composta da tre elementi fondamentali, a cui sono ascrivibili le principali voci di costo:

- A monte si trova il mercato della **raccolta e lavorazione dei dati**, fondamentale per la mole di dati necessari al *training*, per cui rileva anche la qualità delle collezioni;
- Per addestrare i modelli probabilistici a partire dalle grandi collezioni di dati e, successivamente, per utilizzarli (*inference*), occorre un altrettanto **imponente capacità di calcolo**;
- A valle della catena, troviamo quindi il **modello *transformer***, ovvero la rete neurale in grado di mappare le relazioni probabilistiche tra dati e ricavarne conoscenza da utilizzare per svolgere i compiti affidatigli.

Per quanto riguarda l'**addestramento** di GPT, ad esempio, distinguiamo tra i costi della capacità computazionale e i costi dei dati. Per questa seconda voce non sono disponibili molte informazioni, essendo talvolta controverse le fonti di reperimento dei dati⁵³. I **costi dei dataset** necessari all'addestramento sono in ogni caso costantemente in crescita per la **sempre crescente dimensione** dei modelli, considerata necessaria a migliorarne la capacità di ragionamento (si veda *annex 4* sulle capacità emergenti).

Questa costante espansione delle dimensioni si traduce in iterazioni di addestramento sempre più lunghe che richiedono una sempre maggiore capacità computazionale e **consumano sempre più energia**: i modelli all'avanguardia vengono addestrati per settimane o mesi per raggiungere prestazioni ottimali.

OpenAI ha stimato che la quantità di calcolo utilizzata nel *training* delle varie versioni di GPT è **umentata in modo esponenziale**, raddoppiando ogni pochi mesi. Tuttavia, ad aprile dello scorso anno, il CEO di OpenAI

⁵³ Non è chiara la provenienza dei dati utilizzati per addestrare GPT. Infatti, al momento la questione è oggetto di più di una causa giuridica: il 28 giugno scorso, 16 querelanti anonimi hanno intentato una causa contro OpenAI, in forma di *class action*; mentre lo scorso 20 settembre l'associazione di categoria degli scrittori degli Stati Uniti, la Authors Guild, ne ha intentata un'altra. I querelanti sostengono che OpenAI avrebbe addestrato GPT con testi pubblicati online o tramite l'editoria tradizionale senza richiedere l'autorizzazione degli autori; infine, il 27 dicembre 2023, il New York Times ha intentato causa ad OpenAI accusandola di aver utilizzato milioni di articoli del New York Times per addestrare i chatbot che fanno concorrenza allo stesso giornale. Per quanto riguarda i testi ottenuti da internet, secondo i querelanti della prima di queste azioni, si tratterebbe di un totale di 300 miliardi di parole.

Sam Altman ha predetto che il tasso di crescita della dimensione dei modelli sarebbe diminuito; se questo è vero, in futuro ci si può aspettare una **crescita delle dimensioni più lenta, ma comunque positiva**⁵⁴.

Al momento, le dimensioni dei modelli **crescono più velocemente delle capacità hardware**: pertanto è richiesto **un numero sempre crescente di processori**. La domanda di processori espressa da OpenAI e dai suoi *competitor* ha condotto a una **carenza mondiale di processori**, facendone aumentare il prezzo. OpenAI starebbe valutando di avviare la produzione di propri **microprocessori specifici per l'intelligenza artificiale**; questa scelta collocherebbe OpenAI nel ristretto gruppo di grandi attori tecnologici che tentano di controllare direttamente la progettazione dei processori per loro fondamentali⁵⁵.

Una volta che il modello è stato sviluppato e addestrato, occorre **altra capacità computazionale per "fare inferenza"**, ovvero per utilizzarlo. La quantità di potenza computazionale in questo caso dipende non solo dalle dimensioni dei *database*, ma anche dal **numero di utenti** (a sua volta in costante crescita): secondo alcune stime per far funzionare ChatGPT servivano inizialmente 3 milioni di dollari al mese, mentre ad oggi servirebbero tra 700 mila dollari e 1 milione di dollari al giorno.

Infine, un'osservazione sull'efficienza generale dei modelli di grandi dimensioni: secondo un articolo⁵⁶ pubblicato dalla rivista Forbes, una parte sostanziale delle operazioni computazionali nei modelli di grandi dimensioni sarebbe attualmente **sprecata**. Valutare l'effettiva esistenza di uno spreco di risorse da parte dei modelli attuali, nonché l'entità di tale spreco, e quindi **l'esistenza e l'entità di un margine di miglioramento tecnico**, richiederebbe una conoscenza approfondita e specialistica del funzionamento dei modelli e dei processori oggi utilizzati. Ci limitiamo a segnalare quindi l'esistenza di numerose fonti specialistiche e la costante produzione in ambito scientifico di ricerche volte a migliorare l'efficienza dei modelli fin qui discussi. Si segnalano anche le dichiarazioni di Sam Altman, il quale, ad aprile 2023, ha dichiarato che le prestazioni dei modelli dovranno essere rese "migliori in altri modi⁵⁷", ovvero non tramite l'aumento della dimensione.

c. Impatti sul tessuto produttivo europeo

Date le caratteristiche tendenzialmente oligopolistiche del mercato dell'IA, il progetto europeo di sviluppare un settore IA in grado di competere con le grandi imprese americane e cinesi appare ambizioso e irto di ostacoli: le Big Tech americane sono infatti in grado di investire miliardi nella capacità computazionale e, dominando il settore digitale da decenni, sono

⁵⁴ "Large, creative AI models will transform lives and labour markets." The Economist, 22 aprile 2023.

⁵⁵ Alphabet, l'impresa proprietaria di Google, e Amazon hanno già intrapreso questa strada in passato. Per quanto riguarda la produzione di processori specifici per l'IA, a novembre 2023 Microsoft ha lanciato Ignite, il suo primo *chip* a misura di IA; mentre Google ha annunciato ad agosto l'ultima versione del suo chip per IA, il TPUv5e (TPUv5 lite), che consentirebbe di utilizzare modelli anche più grandi di GPT allo stesso costo affrontato da OpenAI.

⁵⁶ "What Large Models Cost You – There Is No Free AI Lunch." Craig Smith, Forbes, 8 settembre 2023.

⁵⁷ "I think we're at the end of the era where it's going to be these giant, giant models" (...) "We'll make them better in other ways.", ha dichiarato Altman nel corso di un evento tenutosi al MIT ad aprile.

attive da lungo tempo nella raccolta dati; le controparti cinesi invece dispongono di enormi risorse statali, nonché dei dati raccolti dallo Stato sui cittadini cinesi).

Di certo, l'impatto in Europa riguarderà **l'adozione delle tecnologie** prodotte all'estero come fattore di produzione capitale in grado di aumentare notevolmente la produttività del lavoro: tale adozione è necessaria a salvaguardare la competitività delle imprese europee negli altri settori produttivi. Un ambito di applicazione particolarmente rilevante per l'industria europea è quello della manifattura: i modelli IA possono essere impiegati per il *design* dei beni e dei sistemi di produzione, per la gestione degli impianti e delle catene di approvvigionamento. L'impiego dei modelli in questo settore, anche grazie all'utilizzo di sensori, sarebbe supportato da una capillare raccolta dati, che in futuro potrebbe essere cruciale per realizzare un'efficace integrazione dell'IA con l'automazione tradizionale.

Riguardo la capacità di adozione delle nuove tecnologie, un aspetto che potrebbe risultare rilevante è quello della **dimensione delle imprese**: le PMI sono solitamente meno abili nell'adottare le innovazioni, e potrebbero avere maggiori difficoltà nel proteggere il proprio *know-how* dalle imprese fornitrici dei servizi di IA, che, tramite i modelli in grado di apprendere, potrebbero raccogliere dati sulle competenze proprie delle PMI. Le piccole imprese potrebbero anche risultare più esposte a eventuali *takeover* da parte di aziende straniere, volti a realizzare integrazioni orizzontali tra prodotti digitali e non digitali. Simili integrazioni orizzontali potrebbero, inoltre, causare la fuoriuscita dal mercato delle piccole imprese non integrate.

Appare fondamentale, a tal proposito, prevedere quale impatto possa avere l'avvento dell'IA **sui settori non digitali**, attraverso diversi canali: l'eventuale evoluzione degli stili di consumo, le evoluzioni dal lato della produzione già descritte, e per l'appunto l'integrazione orizzontale causata dall'iperconnettività⁵⁸. Al tempo stesso, i nuovi modelli IA potrebbero tuttavia rappresentare anche un vantaggio per le PMI, nella misura in cui potrebbero realizzare per queste dei servizi a basso costo, come servizi giuridici o di conoscenza dei mercati.

Infine, si segnala che i modelli linguistici potrebbero essere impiegati, in tutti i settori, anche a fini anticoncorrenziali, ovvero per raggiungere accordi impliciti attraverso la "collusione algoritmica": i modelli di intelligenza artificiale potrebbero infatti essere utilizzati dalle imprese per identificare e fissare i prezzi dei beni prodotti; come ricordato a margine del vertice G7 a Tokyo nel 2023, «esiste il rischio che tali strumenti possano facilitare la collusione o aumentare ingiustamente i prezzi»⁵⁹.

⁵⁸ Cagnin, C., Muench, S., Scapolo, F., Störmer, E., Vesnic-Alujevic, L. Shaping and securing the EU's

open strategic autonomy by 2040 and beyond. Publications Office of the European Union, Luxembourg, 2021.

⁵⁹ Ad esempio, uno studio LSE, dedicato all'impiego dell'IA per predire la curva di domanda relativa a singole imprese, ha mostrato che la precisione delle previsioni dell'IA, e quindi dei segnali di mercato quali il prezzo, aumenta man mano che diviene noto che molte aziende hanno adottato l'intelligenza artificiale per predire la propria curva di domanda individuale. L'accuratezza di tali previsioni, trasmesse attraverso i segnali di prezzo, rappresenta un'esternalità per le imprese che

d. Effetti sulla domanda e sul benessere dei consumatori

La diffusione dei nuovi modelli IA in ambito privato avrà molto probabilmente un impatto sul comportamento degli attori economici, oltre che dal lato dell'offerta, anche dal lato della domanda, similmente a quanto avvenuto con l'introduzione di precedenti innovazioni tecnologiche (internet).

L'offerta di servizi realizzati tramite i modelli IA e destinati agli utenti individuali rappresenta un altro motore di trasformazione trasversale a tutti i **settori rivolti ai consumatori finali**, e riguarda anche quei settori che in futuro proprio grazie all'IA potrebbero riorganizzare la distribuzione evitando alcuni intermediari, accelerando il processo già avviato, ad esempio, con la diffusione dell'*e-commerce*. Questa tecnologia potrebbe rappresentare un mezzo utilizzabile dagli individui per autoprodurre alcuni servizi oggi offerti a livello professionale con un potenziale impatto sugli **stili di consumo di beni e servizi non digitali**, promuovendo una riduzione delle asimmetrie informative tra consumatore e produttore⁶⁰.

Un esempio degli strumenti attraverso i quali potrebbe realizzarsi tale evoluzione degli stili di consumo è costituito dai cosiddetti '*custobot*', neologismo nato dalla fusione di '*customer*' e '*bot*' (*software automatizzato*), ovvero agenti IA in grado di compiere acquisti per altri. Questi modelli IA, programmabili anche attraverso il linguaggio naturale e dotati della capacità di negoziare e acquistare autonomamente beni e servizi, potranno sia svolgere la funzione del responsabile acquisti per le aziende, sia assistere i consumatori individuali nelle scelte di consumo. Si prevede che entro il 2028, il numero di questi *bot* raggiunga i 15 miliardi, con una crescita esponenziale negli anni successivi⁶¹. Secondo alcuni osservatori, questa evoluzione è destinata a generare entrate per trilioni di dollari entro il 2030, segnando un cambiamento ancora più rilevante rispetto all'avvento del commercio digitale, di cui rappresentano per certi versi un'estensione. Infatti, se con l'avvento dell'*e-commerce* i consumatori hanno visto moltiplicare le proprie opzioni di scelta, tramite i *custobot* i consumatori potranno operare tale scelta seguendo principi di pura logica, avvicinandosi a realizzare il **principio di razionalità**, che è un presupposto dell'*homo oeconomicus*.

Le decisioni di acquisto seguono una formula ben definita, che comprende la scoperta, la valutazione, e l'acquisto del prodotto: attività che gli agenti IA possono compiere con maggiore efficienza rispetto agli esseri umani, elaborando velocemente molti più dati, e rispettando al contempo le preferenze personali del singolo consumatore.

Questa nuova dinamica di consumo potrebbe rappresentare un impatto sul benessere dei consumatori individuali che sarà difficilmente catturata dagli indicatori tradizionali come il PIL; inoltre, i *custobot* potrebbero essere in

non hanno impiegato l'IA. - Gans, Joshua S. "Artificial intelligence adoption in a competitive market." *Economica* 90.358 (2023): 690-705.

⁶⁰ Dal punto di vista dell'analisi economica, rileviamo che un aumento del livello di informazioni disponibili e, conseguentemente (almeno in teoria), della capacità di esprimere valutazioni razionali, potrebbe parzialmente colmare, soprattutto per gli attori individuali, la distanza che separa la realtà dai modelli basati sul principio di razionalità assoluta.

⁶¹ "Gartner identifies the top 10 strategic technology trends for 2024", Orlando, Florida, 16 ottobre 2023.

grado di integrare nei processi decisionali elementi che non si riflettono nel prezzo dei beni, incluse le eventuali **esternalità**⁶² di cui il consumatore voglia tenere conto (es. per ridurre l'impatto ambientale, il *custobot* potrebbe valorizzare la durabilità dei beni, considerare l'eventuale presenza di materiali inquinanti, oppure preferire venditori meno distanti geograficamente).

Questa nuova modalità di distribuzione potrebbe favorire un'allocazione delle risorse più equilibrata, e distribuire la domanda più uniformemente tra un'ampia varietà di prodotti, potenziando il benessere collettivo migliorando la capacità di scelta dei consumatori.

e. Effetti di redistribuzione tra fattori produttivi

L'avvento dell'IA parrebbe determinare un **aumento dell'intensità di capitale** nella quasi totalità dei settori produttivi. Di conseguenza l'impiego massiccio di IA nella produzione corrisponderebbe a una riduzione della quota di prodotto utilizzata per remunerare il lavoro e a un aumento della quota destinata alla remunerazione del capitale. Rileva a tal proposito anche la **natura ibrida del fattore IA**, che, pur essendo a tutti gli effetti un fattore capitale, è tuttavia in grado di imparare lavorando; ragione per cui l'impiego del capitale IA equivale a un momento di investimento per l'accrescimento dello stesso capitale.

Le tecniche di *machine learning*, infatti, fanno sì che i modelli IA siano in grado di acquisire esperienza e aggiornarsi, **migliorando sotto il profilo della propria produttività**. Ciò avverrebbe non solo tramite il miglioramento della qualità dei risultati ottenuti dall'imitazione delle competenze umane, ma anche accrescendo nel tempo il numero di mansioni lavorative che la macchina riesce a svolgere: si tratta infatti di un bene capitale fungibile, in grado di adattarsi e quindi di espandere i potenziali ambiti di applicazione (*general purpose technology*). Per queste ragioni, tale bene capitale appare in grado di **umentare di valore nel tempo**, attraverso un processo opposto al deprezzamento che normalmente caratterizza i beni capitali.

Per i lavoratori appartenenti alla generazione attuale questo fenomeno pone una questione urgente che non ha precedenti nelle passate ondate di automazione. Infatti, poiché la macchina è in grado di **apprendere dal lavoratore che la utilizza**, e conseguentemente di accrescere le proprie competenze e il proprio valore, il semplice utilizzo dell'IA da parte del lavoratore si configura come un **conferimento gratuito di capitale umano** da parte dei lavoratori a favore delle imprese proprietarie del modello IA⁶³.

Questa partecipazione, più o meno consapevole, del lavoratore all'effettiva **costruzione del modello IA**, solleva questioni diverse a seconda di chi

⁶² In taluni casi si può immaginare che in futuro una regolamentazione ambientalista dei *custobot* preveda che essi attribuiscano peso o segnalare eventuali esternalità ambientali.

⁶³ Per quanto questa possibilità sia oggi ipotizzabile solo per i lavori della conoscenza (le cui competenze potrebbero essere apprese da un modello *transformer* studiando, ad esempio, l'utilizzo che un lavoratore fa del proprio computer a lavoro), nel futuro la questione potrebbe riguardare anche lavori fisici. Alcune ricerche sono già in corso per identificare i passaggi necessari ad addestrare robot nel sostituire gli operai, come nel caso di questo studio dedicato al settore delle costruzioni: Wu, Haitao, et al. "A survey on teaching workplace skills to construction robots." *Expert Systems with Applications* 205 (2022): 117658.

detiene la proprietà dei dati estratti dall'osservazione del lavoro svolto dall'operatore umano⁶⁴. Se i dati sono trattenuti dall'impresa che impiega il lavoratore, la questione riguarda il rapporto tra datori di lavoro e lavoratori; se invece l'acquisizione di questi dati migliora la qualità del modello IA, avvantaggiando i proprietari del modello, si pongono due tipi di questioni: una riguardante il rapporto tra lavoratori e proprietari del modello, un'altra che coinvolge i proprietari del modello IA e le imprese che utilizzano l'IA come fattore produttivo, acquistandola come servizio offerto dai primi. In questo secondo caso il rischio riguarda la cessione di **know-how** aziendale.

In ogni caso, per il singolo lavoratore ciò implica la cessione involontaria delle proprie competenze professionali ai proprietari del modello IA: si pone quindi la questione di garantire **una giusta remunerazione** per il conferimento di conoscenze⁶⁵. Le modalità in cui potrebbe avvenire tale remunerazione non sono di semplice definizione, poiché non si tratta semplicemente di formazione, retribuibile tramite un aumento di stipendio, ma di una cessione permanente, non solo delle proprie competenze, ma potenzialmente delle informazioni necessarie a ricostruire un'intera *forma mentis*, ovvero una sensibilità professionale costruita in anni di esperienza, a favore di un'impresa proprietaria di un *asset* capitale, il modello IA, che potrebbe impiegare tali informazioni e **trarne profitto per un tempo indefinito**, anche sostituendo lo stesso lavoratore.

Se il conferimento di capitale umano può essere considerato alla stregua di qualsiasi altro conferimento di capitale, ne deriva che il lavoratore che addestra un modello IA sta in un certo senso investendo nell'azienda che ha sviluppato il modello: si potrebbe quindi ipotizzare che ciò costituisca il fondamento per un diritto di proprietà, portando alla conclusione che sia adeguata una retribuzione sotto forma di partecipazioni azionarie.

Esaminando la questione da un punto di vista macroeconomico, tuttavia, la retribuzione dei singoli lavoratori che aiuteranno ad addestrare i modelli, anche qualora avvenisse sotto forma di azioni, non ridurrebbe il **trasferimento di capacità produttiva da parte del fattore lavoro al fattore capitale**, reso possibile dalla natura particolare dell'*asset* in questione, ovvero dalla capacità di apprendimento esibita dai modelli IA.

Peraltro, essendo, ad oggi, la proprietà intellettuale dei modelli IA notevolmente concentrata in poche mani, con sede in un numero ancora minore di Stati, **altri effetti distributivi potrebbero verificarsi tra Paesi**.

Gli effetti sulle dinamiche salariali per diverse tipologie di lavoratori sono approfonditi nel successivo capitolo, incentrato sul mercato del lavoro.

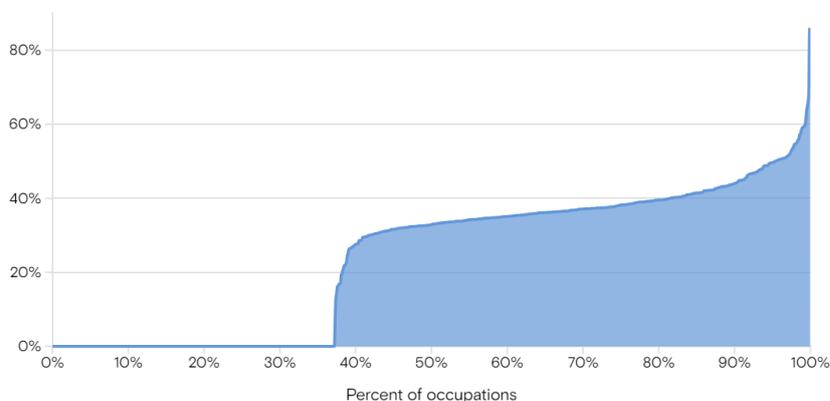
⁶⁴ Questo fenomeno di cessione gratuita di dati e informazioni all'IA si verifica anche nel caso degli utenti (in alcuni casi paganti) che impiegano prodotti costruiti sui modelli LLM come chatGPT: non diversamente da quanto avviene per la cessione spontanea di dati personali da parte degli utenti dei social network, gli utenti di chatGPT forniscono gratuitamente ad OpenAI non solo testi, ma anche una serie di istruzioni che affinano la capacità della macchina di comprendere le richieste degli utenti futuri.

⁶⁵ Secondo i legali dei diciassette scrittori della Authors Guild americana che hanno intentato causa a OpenAI per violazione del diritto d'autore delle loro opere di narrativa "La decisione degli imputati di copiare le opere degli autori, presa senza offrire alcuna scelta o fornire alcun compenso, minaccia il ruolo e il sostentamento degli scrittori nel loro insieme". Considerazioni simili potrebbero essere estese a tutte le competenze che l'IA è in grado di imitare, soprattutto quando i dati sono ceduti legalmente, tramite l'impiego delle tecnologie ML sul posto di lavoro.

3. IMPATTO SUL MERCATO DEL LAVORO

Il rapporto Goldman-Sachs già citato sostiene che, a livello globale, l'equivalente di circa 300 milioni di posti di lavoro a tempo pieno potrebbero essere esposti ad automazione grazie all'impiego dei LLM: nel mercato statunitense ed europeo questo rischio riguarderebbe circa **due terzi delle occupazioni**, che sarebbero esposte a variegati livelli di automazione da parte dell'intelligenza artificiale; e **per un quarto di queste, l'automazione riguarderebbe circa la metà del loro carico di lavoro**. Tuttavia, per la maggior parte dei posti di lavoro l'automazione avverrebbe **in modo parziale**: in altre parole, il lavoro umano dovrebbe essere affiancato da quello dell'IA e non sostituito.

Share of occupational workload exposed to automation by AI



Fonte: Goldman-Sachs (2023).

Per questo studio sono stati utilizzati il *database* americano delle professioni O*NET, che descrive il contenuto delle attività di oltre 900 occupazioni, e il *database* ESCO, che descrive oltre 2000 occupazioni⁶⁶. Di conseguenza, la stima della quota di lavori esposti all'automazione tramite IA si basa su descrizioni di ruoli che non contengono aspetti chiave come la capacità di stabilire una relazione umana o di avere una presenza fisica attraverso un corpo umano, poiché queste caratteristiche sono implicitamente attribuite agli esseri umani. Tuttavia, non considerare questi elementi nella descrizione può falsare i risultati degli studi, particolarmente per quanto riguarda le professioni a contatto con il pubblico, dall'istruzione alle vendite al dettaglio. Ad oggi, queste mansioni costituirebbero un collo di bottiglia per l'automazione tramite IA, sebbene potrebbero invece diventare automatizzabili tramite l'impiego di successive **innovazioni collaterali**. Infatti, l'importanza degli strumenti secondari riguardo la scalabilità degli impatti economici dei modelli linguistici è chiarita dai risultati di un altro studio⁶⁷ sul mercato del lavoro americano: l'analisi

⁶⁶ O*NET è stato creato nel 1998 dal Dipartimento del Lavoro degli USA, mentre ESCO è stato lanciato dalla Commissione Europea nel 2010: entrambi i database sono regolarmente aggiornati per allinearsi alle dinamiche del mercato del lavoro.

⁶⁷ Eloundou, Tyna, et al. "Gpts are gpts: An early look at the labor market impact potential of large language models." arXiv preprint arXiv:2303.10130 (2023).

suggerisce che circa il 15% di tutte le attività dei lavoratori negli Stati Uniti potrebbe essere completata in modo significativamente più veloce e con lo stesso livello di qualità grazie all'impiego di LLM; tuttavia, incorporando i *software* e altri strumenti secondari basati su LLM, questa quota aumenta raggiungendo tra il 47 e il 56% di tutti i compiti.

Si legge nell'Employment Outlook 2023 realizzato dall'OCSE: "i progressi nel campo dell'intelligenza artificiale (IA) sono stati tali che, in alcune aree, **i risultati dell'IA sono diventati indistinguibili da quelli degli esseri umani**. Questi rapidi sviluppi, se combinati con il calo dei costi di produzione e adozione di queste nuove tecnologie, suggeriscono che le economie dell'OCSE potrebbero essere sull'orlo di una rivoluzione dell'intelligenza artificiale che potrebbe cambiare radicalmente il posto di lavoro." Ciò nonostante, le conclusioni dell'Outlook OCSE restano sospese riguardo all'impatto complessivo dei *foundation models* sul mercato del lavoro.

La difficoltà nel fare previsioni è dovuta in parte ai rapidi e imprevedibili progressi della tecnologia, che per molti versi è ancora **impossibile da studiare**, poiché protetta da proprietà intellettuale e dalle stesse dimensioni dei modelli (data la capacità di calcolo necessaria a ispezionarli).

In ogni caso, per quanto ci si possa attendere che la maturazione della tecnologia porti col tempo a un calo dei costi come suggerito dall'OCSE, è utile ricordare che questi sono ad oggi **in aumento per via dell'espansione della dimensione dei modelli**. L'andamento dei costi dello sviluppo e dell'impiego dell'IA è in ogni caso legato anche alla **disponibilità di microprocessori e di energia** (si veda, al riguardo, il paragrafo 2b). Dunque, gli studi basati sul confronto delle capacità dell'IA con le competenze previste per lo svolgimento delle diverse professioni potrebbero non considerare la convenienza concreta della sostituzione dei lavoratori.

Al contempo, la rilevanza dei costi di sviluppo e di funzionamento dei modelli IA potrebbe essere meno determinante di quanto non ci si aspetti: è infatti possibile che le *Big Tech* detentrici dei diritti di proprietà intellettuale sui modelli IA siano interessate a introdurre sul mercato prodotti basati su IA a prezzi competitivi, **anche operando in perdita**, promuovendo quindi la sostituzione del lavoro **a prescindere dai costi sostenuti per offrire i servizi IA**. Questo potrebbe avvenire poiché, dal punto di vista delle imprese sviluppatrici, promuovere l'impiego degli strumenti IA equivale ad aumentare la raccolta di *feedback* e dati; in altre parole corrisponde a una **prosecuzione dell'addestramento della macchina**. Pertanto, una simile strategia commerciale, pur comportando una perdita, potrebbe essere sostenuta poiché costituirebbe un **investimento a lungo termine** nello sviluppo dei modelli IA. Si ricorda inoltre che tramite l'impiego dei modelli si potrebbe osservare anche un **aumento della fungibilità del capitale**: infatti, durante il tempo in cui è impiegato in un dato processo, il modello IA potrebbe acquisire le competenze necessarie ad affrontare anche compiti apparentemente molto diversi: le informazioni raccolte potrebbero essere usate per sviluppare competenze inattese, a conferma della natura di "tecnologia a scopo generale".

In ogni caso, questi elementi riguardanti la **convenienza concreta** di impiegare l'IA in luogo del lavoro **non sono considerati** negli studi riportati

in questo capitolo, che invece si concentrano principalmente sul confronto delle abilità dell'IA con quelle umane. Al di là, quindi, dell'incerta evoluzione tecnica e commerciale, di seguito si segnalano **due ambiti di incertezza** considerati negli studi su IA e mercato del lavoro discussi nel resto del capitolo:

1. Data la natura stessa della tecnologia, molteplici effetti agiscono contemporaneamente a seguito del suo inserimento nel mercato del lavoro, rendendo complesse le previsioni su quali di essi prevalga;
2. Il mercato del lavoro è segmentato: l'adozione della tecnologia si intreccia con la possibilità di esternalizzare determinate mansioni nei mercati che presentano un più basso costo del lavoro.

Prima di proseguire l'analisi degli impatti sul mercato del lavoro, sulla scia del lavoro svolto dall'OCSE in merito, segnaliamo come la capacità dei nuovi modelli di sostituire il lavoro umano si sovrapponga parzialmente alla diminuzione della **forza lavoro disponibile**, che comporterà in particolare scarsa disponibilità di alcune figure professionali⁶⁸.

a. IA come fattore ibrido: sostituibilità o complementarità con il lavoro

Per le peculiari caratteristiche eterogenee, l'IA si presenta dal punto di vista economico come una **ibridazione tra fattore capitale e fattore lavoro**: si tratta senz'altro di un asset destinabile a impieghi produttivi, ma non è chiaro se esso deperisca nel tempo, poiché è in grado di "imparare lavorando" e quindi di **aggiornarsi e migliorare la propria produttività, potenzialmente anche aumentando di valore**. Inoltre, pur essendo descrivibile come uno strumento che aumenta la produttività del lavoro, l'IA è in grado di acquisire potenzialmente nel tempo una certa **autonomia rispetto al lavoratore**: aspetto che caratterizzerebbe l'eventuale sostituzione del lavoro in modo nuovo rispetto a quanto avviene normalmente in seguito all'introduzione di strumenti *labour-augmenting*.

Secondo l'Employment Outlook dell'OCSE una delle principali motivazioni dichiarate dalle imprese per investire nell'intelligenza artificiale è la volontà di **aumentare la produttività del lavoro e ridurre i costi del personale**⁶⁹.

L'Outlook 2023 non rileva importanti riduzioni nella domanda di forza lavoro, ma segnala che l'adozione dell'intelligenza artificiale è ancora relativamente bassa: gli effetti negativi dell'adozione di queste tecnologie sull'occupazione potrebbero richiedere del tempo per concretizzarsi. Le prime stime che considerano l'impatto dei Large Language Models rilevano che le occupazioni che richiedono un'istruzione o una formazione superiore alla media sono maggiormente esposte all'intelligenza artificiale.

Nonostante il probabile **effetto 'displacement'** dell'IA, dovuto dall'aumento della produttività del lavoro a parità di prodotto, l'Outlook

⁶⁸ La questione demografica si intreccia con il tema dell'IA anche per quanto riguarda gli effetti dell'invecchiamento della popolazione sulla capacità di produrre innovazione: l'utilizzo dell'intelligenza artificiale potrebbe infatti parzialmente compensare la perdita di capacità innovativa legata alla diminuzione della quota di forza lavoro giovane.

⁶⁹ Secondo alcuni datori di lavoro intervistati per l'Outlook dell'OCSE, l'adozione dell'IA non servirebbe solo a ottenere un vantaggio competitivo, configurandosi nel lungo periodo come l'unica possibilità di restare sul mercato.

OCSE conclude che **l'impatto netto dell'IA sull'occupazione è ambiguo**, data l'impossibilità di prevedere quali effetti potrebbero compensare il primo. Da un lato, l'effetto 'produttività' potrebbe aumentare la domanda unitaria di lavoro da affiancare all'IA in seguito alla diminuzione dei salari, ovvero tramite l'aumento complessivo del livello di produzione. A tal proposito, si segnala che, nel dibattito in corso sull'intelligenza artificiale, a sostegno della tesi secondo cui l'effetto produttività sarebbe in grado di compensare l'effetto *displacement*, viene spesso citato un *paper* del 2018 di Daron Acemoglu e Pascual Restrepo, che tuttavia non terrebbe conto degli sviluppi nell'ambito del *machine learning* avvenuti negli ultimi anni⁷⁰.

Un altro effetto in grado di compensare l'effetto *displacement* individuato nell'Outlook dell'OCSE è l'**effetto 'reintegrazione'**, che bilancerebbe l'effetto sostituzione del lavoro tramite la creazione di nuove tipologie di lavori, per lo svolgimento delle quali sarebbero necessari lavoratori con **competenze complementari** all'IA⁷¹. Il paragrafo conclusivo di questo capitolo, dedicato all'impatto sulle **caratteristiche del lavoro**, raccoglie alcune riflessioni circa l'effetto reintegrazione ipotizzato dall'OCSE e la conseguente necessità di riqualificare i lavoratori.

Appaiono analoghe le conclusioni cui giunge uno studio dell'ILO⁷² basato sull'analisi delle singole mansioni associate alle diverse professioni (secondo i sistemi di classificazione internazionali): secondo lo studio la maggior parte dei settori sono solo parzialmente esposti all'automazione derivante dall'impiego di IA generativa, per cui appare **probabile che l'impatto maggiore della nuova tecnologia non consista nella distruzione di posti di lavoro**: l'automazione parziale delle mansioni permetterà piuttosto ai lavoratori di concentrarsi sulle mansioni non automatizzabili (*labour augmentation*)⁷³.

⁷⁰ Il *paper*, intolato "*Artificial intelligence, automation and work*", basa le proprie conclusioni su un lavoro precedente di Acemoglu e Restrepo risalente al 2016, che osserva come l'automazione abbia in passato incentivato le imprese ad introdurre nuove attività ad alta intensità di manodopera proprio a seguito di una riduzione della domanda di lavoro e quindi dei salari. Questi incentivi endogeni sarebbero in grado di equilibrare spontaneamente il processo di crescita anche nel caso dell'IA. Tuttavia, nel testo di Acemoglu e Restrepo si sottolinea che la tecnologia considerata è quella descritta in un *paper* del 2009, che "mira allo sviluppo di sistemi commerciali di Intelligenza Artificiale applicata, specializzati in compiti chiaramente delineati legati alla previsione, al processo decisionale, alla logistica e al riconoscimento dei *pattern*", rinunciando all'ambizione di creare un'intelligenza artificiale generale, precedentemente teorizzata dalla cibernetica degli anni '60. Seguendo queste indicazioni, Acemoglu e Restrepo considerano i seguenti *task* come non automatizzabili grazie all'IA: il ragionamento complesso, l'apprendimento basato sull'analogia, la risoluzione di problemi astratti e la comunicazione. Tutti questi compiti appaiono oggi, almeno in una certa misura, alla portata dei modelli *transformer* introdotti sul piano teorico a partire dal 2017 e realizzati concretamente dal 2021 in poi (si veda capitolo 1 e annex 2 e 3).

⁷¹ Un *prompt engineer*, ad esempio, è un esperto che progetta gli input per gli strumenti di intelligenza artificiale generativa in modo che producano output ottimali.

⁷² Gmyrek, P., Berg, J., Bescond, D. *Generative AI and jobs: A global analysis of potential effects on job quantity and quality*. ILO Working Paper 96. Geneva: International Labour Office, 2023.

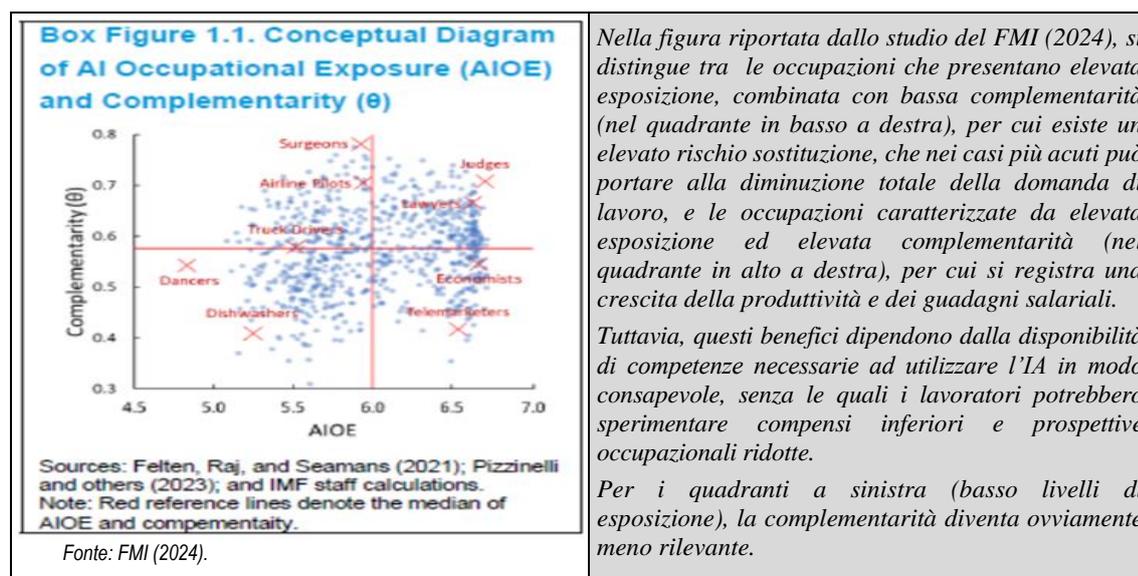
⁷³ L'unico settore altamente esposto a sostituzione sarebbe quello delle mansioni amministrative, che registrano un 24% di mansioni considerate ad alto rischio di sostituzione e un ulteriore 58% di mansioni considerate a rischio medio. Per gli altri settori, le mansioni ad alto rischio oscillano tra l'1 e il 4%, mentre quelle a rischio medio non superano il 25%.

L'ILO sottolinea inoltre la diversificazione degli impatti e i cambiamenti nella qualità dei posti di lavoro, in particolare per quanto riguarda l'intensità del lavoro e l'autonomia. Gli effetti sull'occupazione variano significativamente tra i Paesi a basso e alto reddito, con una maggiore incidenza nelle nazioni ad alto reddito. Si registrerebbe, inoltre, un impatto più significativo sulle lavoratrici donne⁷⁴.

A proposito della **differenziazione dell'impatto tra Paesi e diverse fasce di popolazione**, uno studio del Fondo Monetario internazionale utilizza il termine "*AI divide*" per definire il particolare caso di *digital divide*.

Secondo lo studio del FMI (2024), quasi il 40% dei posti di lavoro a livello globale sarebbe esposto all'IA con **importanti differenze tra Paesi**. Nelle economie avanzate, dove i lavoratori svolgono mansioni orientate verso compiti cognitivi, la percentuale di posti lavoro interessata dall'introduzione dell'IA è maggiore (60%), ma è tale anche la capacità di sfruttare i vantaggi di queste tecnologie in termini di produttività; la metà dei lavoratori interessati infatti potrebbe veder aumentare la propria produttività (e il proprio reddito), mentre l'altra metà potrebbe subire effetti negativi. L'esposizione complessiva nelle economie emergenti è del 40%, mentre nei paesi a basso reddito è del 26%.

Il FMI evidenzia tre canali attraverso i quali l'intelligenza artificiale può determinare trasformazioni nel mercato del lavoro: sostituzione del lavoro, complementarità e guadagni di produttività. L'adozione delle tecnologie IA potrebbe comportare la sostituzione dei lavoratori (le attività precedentemente eseguite dal lavoro verrebbero svolte dal "capitale IA"), con conseguente riduzione della quota di reddito distribuita al fattore lavoro: infatti, assumendo un impatto simile a quello osservato nel Regno Unito tra il 1980 e il 2014, il FMI assume una diminuzione del 5,5% della quota di reddito distribuita al fattore lavoro, specificando che l'impatto varierebbe a seconda dell'esposizione e della complementarità dei lavori all'IA.



⁷⁴ L'analisi sottolinea l'importanza di considerare anche la qualità del lavoro e la necessità di regolamentazioni e dialogo sociale per sostenere l'occupazione di qualità in questo contesto.

Similmente a quanto descritto dall'OCSE, anche nel documento del Fondo si precisa che la perdita di posti di lavoro potrebbe essere compensata da due fenomeni ad essa conseguenti. Il primo è l'aumento della rilevanza dei compiti che l'IA non può sostituire, in particolare nelle occupazioni con elevata complementarità tra lavoro e intelligenza artificiale. Il secondo è analogo all'effetto produttività considerato dall'OCSE: l'aumento di produttività porterebbe a un aumento degli investimenti e della domanda complessiva di manodopera.

Considerando questi canali, il rapporto conclude che, se l'IA sarà adottata in modo efficace, essa potrebbe determinare una **crescita di reddito per la maggior parte dei lavoratori**; tuttavia, poiché la complementarità dell'IA con il lavoro è **positivamente correlata al reddito**, l'introduzione di questa tecnologia **potrebbe esacerbare l'ineguaglianza di reddito** anche all'interno dei Paesi, richiedendo un intervento pubblico attraverso regolamentazioni e politiche redistributive.

All'aumento della disuguaglianza nei redditi da lavoro, si aggiunge l'**ineguale distribuzione dei rendimenti da capitale**, che accentueranno la disuguaglianza in termini di ricchezza (aspetto che appare crescere in rilevanza, considerando l'aumento della quota di prodotto distribuita al capitale; infatti, secondo il FMI per "comprendere l'impatto dell'IA sui livelli di reddito e sulla disuguaglianza dei redditi, devono essere esaminati entrambi i canali, quello del reddito da lavoro e quello del reddito da capitale"). In ogni caso, a determinare l'impatto sulla distribuzione del reddito e della ricchezza delle nuove tecnologie saranno **le scelte politiche dei diversi Paesi** per quanto riguarda la definizione dei diritti di proprietà sui modelli IA, nonché le politiche redistributive e le politiche di imposizione fiscale (rimandiamo al capitolo 5 per la discussione della *policy* offerta dallo stesso rapporto del FMI).

b. Remote Intelligence e Artificial Intelligence

L'IA è una tecnologia peculiare anche per il rapporto che assume rispetto agli scambi. Quasi tutti i progressi tecnologici degli ultimi 20 anni sono stati tali da facilitare o rendere possibili scambi di beni e servizi. Sebbene da un lato, le tecnologie LLM, offrendo ad esempio traduzioni istantanee, rappresenterebbero una tecnologia facilitatrice degli scambi, dall'altro potrebbero ridurli, data la capacità dei modelli LLM di sostituire i lavoratori offrendo lo stesso livello di interazione linguistica (ad es. nei *call center* indiani, simbolo dei servizi finali importati dall'occidente).

Tuttavia, secondo uno studio⁷⁵ di Richard Baldwin e Toshihiro Okubo uscito ad agosto 2023, l'"intelligenza remota" (o RI, termine utilizzato per il lavoro a distanza) e l'intelligenza artificiale sarebbero **due fattori principalmente complementari, e non sostituti**. Infatti, anche pur essendo vero che, grazie agli LLM, oggi alcune attività svolte nel settore servizi sono automatizzabili, oltre che telelavorabili, la maggior parte di esse ricadrebbero alternativamente in una categoria o nell'altra. Secondo Baldwin quindi, restando necessario l'impiego di lavoro umano, è probabile che esso sia anche in futuro svolto da remoto per risparmiare sui costi. Per

⁷⁵ Baldwin, Richard, and Toshihiro Okubo. "Are software automation and teleworker substitutes? Preliminary evidence from Japan." *The World Economy* (2023).

tornare all'esempio dei lavori di 'etichettatura' dei dati: molti di questi sono già oggi svolti nei Paesi in via di sviluppo.

Seguendo le conclusioni a cui giungono Baldwin e Okubo, si intravede il rischio che per i Paesi avanzati non operi a pieno regime l'effetto 'produttività', descritto nell'Outlook OCSE come uno degli effetti in grado di compensare la sostituzione del lavoro con IA. Questo suggerisce di concentrare le *policy* sull'effetto 'reintegrazione', il quale però è fondato sulla possibilità di riqualificare i lavoratori e potrebbe **umentare il rischio polarizzazione dei salari**. Infatti, se è sempre vero che i lavoratori con livelli di istruzione superiori sono più disponibili e più in grado di riqualificarsi, nel caso delle competenze complementari al lavoro svolto dall'IA ciò potrebbe essere ancora più rilevante per la natura delle competenze richieste.

c. Future of work: qualità del lavoro e rischio polarizzazione

In questo paragrafo sono raccolte alcune riflessioni sull'impatto dell'IA sulla qualità del lavoro e sulle condizioni lavorative, nonché sul rischio polarizzazione dei salari, sottolineato dall'OCSE.

Secondo l'*Employment Outlook 2023*, gli impatti dell'IA potrebbero risultare ambigui anche dal punto di vista della **qualità del lavoro**. Se da un lato, la sostituzione del lavoro umano da parte dell'IA può configurarsi come una conquista in tutti i casi in cui riguardi **compiti pericolosi o ripetitivi**, dall'altro occorre considerare che, diversamente dalle tecnologie precedenti, l'intelligenza artificiale **è in grado di automatizzare anche attività non-routine**. Infatti, l'IA ha compiuto notevoli progressi nello svolgimento di **compiti cognitivi** quali la gestione delle informazioni, la memorizzazione, la velocità percettiva e il ragionamento deduttivo. Pertanto, secondo l'Outlook OCSE, si porrebbe come una tecnologia in grado di aumentare la produttività di *manager*, scienziati e ingegneri, giuristi, e professionisti dell'ambito culturale, ed è quindi potenzialmente in grado di ridurre la domanda di lavoro umano anche per posizioni ambite. Alle riflessioni svolte dall'OCSE si può aggiungere che lo sviluppo dell'IA almeno inizialmente, potrebbe persino comportare **un aumento del lavoro routinario**: quello necessario per 'etichettare' una parte dei dati con cui viene addestrata⁷⁶.

Per quanto riguarda le dinamiche salariali, l'OCSE sottolinea la diversità dell'impatto dell'IA su diverse fasce della forza lavoro, e di conseguenza il potenziale verificarsi di **ulteriore polarizzazione dei salari**. Infatti, come si è detto, l'aumento della produttività del lavoro si verificherebbe principalmente per le mansioni svolte dai lavoratori più istruiti, conducendo a un effetto sostituzione degli stessi (quindi a una minore domanda di questi lavoratori) potenzialmente compensato dall'effetto 'reintegrazione'⁷⁷ già discusso. Se ne deduce che, **qualora prevalessse l'effetto sostituzione**, si assisterebbe a un aumento della disuguaglianza all'interno della stessa categoria di lavoratori 'qualificati', determinata

⁷⁶ Nella maggior parte dei casi, le banche dati adeguate al *training* devono contenere una percentuale minore di "labeled data" per permettere alla macchina di mappare in autonomia la maggiore quantità di "unlabeled data" contenuti nella collezione.

⁷⁷ Anche all'interno della fascia di lavoratori qualificati, la riqualificazione su cui si basa la reintegrazione dei lavoratori potrebbe non coinvolgere tutti i lavoratori nella stessa misura.

dall'eccesso di offerta di lavoro qualificato rispetto alla domanda, nonché dall'aumento di produttività, e quindi di retribuzione, che toccherebbe i lavoratori qualificati che manterrebbero il posto di lavoro. Diversamente, **qualora prevalessesse l'effetto reintegrazione**, aumenterebbero le disuguaglianze salariali tra i lavoratori più e meno qualificati: infatti, in tal caso, i lavoratori della conoscenza sarebbero per la maggior parte coinvolti nella transizione tecnologica causata dall'IA senza subire la perdita di posti di lavoro, essi vedrebbero quindi aumentare la propria produttività e, potenzialmente, la propria retribuzione rispetto ai lavoratori impiegati in attività interessate in misura minore o nulla dall'avvento del *machine learning*⁷⁸.

Sul piano della qualità del lavoro, come ricordato anche nell'Outlook dell'OCSE, le nuove tecnologie comportano una **maggiore pervasività del controllo sui lavoratori**, sollevando problemi riguardo al rispetto della *privacy*. L'aumento del controllo sul lavoro potrebbe, inoltre, incentivare modalità di retribuzione legate al risultato e quindi **aumentare il rischio connesso al reddito da lavoro**. Il **monitoraggio costante dei lavoratori** che utilizzano strumenti di intelligenza artificiale ha, infine, un altro risvolto connesso alla cessione delle competenze del lavoratore al modello IA, aspetto già analizzato nel paragrafo 2e, dedicato agli effetti di redistribuzione delle quote di prodotto tra fattori.

4. FATTORI EXTRA-ECONOMICI CON POSSIBILI IMPATTI SULL'ECONOMIA

Questo capitolo esamina alcuni dei rischi dell'IA che non riguardano direttamente l'economia, ma possono comunque avere importanti implicazioni indirette su di essa.

Notevoli effetti sulla società possono essere determinati dai limiti dei modelli IA, discussi nel capitolo 1d (*bias*, allucinazioni, *misalignment*), almeno fintantoché non saranno disponibili delle soluzioni tecnologiche⁷⁹ o di altro genere, e specialmente prima della maturazione di un certo livello di consapevolezza da parte degli utenti e dei consumatori.

La parziale sostituzione del lavoro, come già discusso nei capitoli 1 e 3, avrebbe probabilmente un impatto su **alcuni aspetti qualitativi** dei beni e dei servizi offerti, tuttavia compensati da vantaggi ottenibili sia in termini di efficienza (es. riduzione del prezzo), sia in termini di altre qualità (es. personalizzazione minuziosa del servizio, indipendenza dagli orari lavorativi previsti per il lavoro umano, superamento di attriti sociali come imbarazzo, percezione che l'impiego dell'IA comporti in alcuni casi una minore rinuncia alla *privacy*, ecc.). Inoltre, i consumatori finali potrebbero

⁷⁸ Secondo il Fondo monetario internazionale, l'adozione dell'IA conduce a una diminuzione della disuguaglianza del reddito da lavoro poiché per i lavoratori istruiti, che percepiscono paghe maggiori, l'effetto sostituzione è più forte; tuttavia, se l'intelligenza artificiale è altamente complementare al lavoro, l'aumento di produttività supera l'effetto sostituzione, conducendo a un aumento di produttività per i lavoratori istruiti, cosa che inasprisce la disuguaglianza.

⁷⁹ Ricordiamo che una delle soluzioni proposte consiste nell'aumento delle dimensioni dei modelli (ovvero della quantità di dati e parametri utilizzati), che tuttavia comporta altre conseguenze negative, come l'aumento dei costi di sviluppo, l'aumento di concentrazione del mercato e l'impossibilità di ispezionare i modelli.

non essere completamente consapevoli della perdita di alcuni aspetti qualitativi, perché ne potrebbero sottovalutare l'impatto (ad es. se si produce nel lungo termine) o potrebbero non essere in grado di distinguere immediatamente tra i prodotti del lavoro umano e quelli dell'IA. Occorre inoltre valutare attentamente **le conseguenze per la collettività** nei casi in cui il lavoro umano sia sostituito in alcuni ambiti particolari (es. istruzione)⁸⁰.

Tenuto conto di questi elementi, in questo capitolo sono esaminati, senza pretesa di esaustività, alcuni dei potenziali effetti collaterali che l'IA potrebbe presentare in diversi ambiti della società.

Oltre ai temi approfonditi nei paragrafi che seguono, meritano una menzione le diverse modalità in cui l'IA potrebbe **supportare il compimento di azioni illegali**, come il finanziamento al terrorismo, il riciclaggio del denaro sporco, l'aggiotaggio, la contraffazione, e altri reati di tipo personale, economico o finanziario. Questi sono ambiti già adeguatamente coperti da regolamentazione, ma l'avvento dell'IA potrebbe richiedere un loro parziale ripensamento. Oltre a supportare le azioni illegali già previste negli ordinamenti giuridici, l'IA potrebbe anche rendere possibili **nuove forme di azioni dannose per i singoli o la collettività**, come interventi malevoli in grado di minare la stabilità finanziaria o di produrre *default* o *shock* economici pilotati. Questo secondo aspetto, in particolare, richiederà la capacità da parte dei *policy maker* di anticipare i comportamenti non ancora previsti dagli ordinamenti.

a. Media e democrazia

Per quanto riguarda gli effetti sui media e sulla **qualità dell'informazione**, i rischi riguardano la possibilità di produrre e diffondere con poche risorse disinformazione, anche di qualità molto sofisticata, sia per errore (dell'IA o umano) sia allo scopo deliberato di manipolare l'opinione pubblica.

Come già evidenziato nel paragrafo 1c, riguardo ai limiti tecnici attuali, le IA sono soggette alle così dette "allucinazioni" e ai "*bias*"; esse sono inoltre in grado di articolare in modo molto convincente informazioni sbagliate, facilitando la diffusione involontaria delle stesse da parte degli utenti.

Per quanto riguarda i casi in cui la disinformazione sia deliberata, appare evidente la pericolosità dei "*deep fake*", ovvero video, audio e immagini artefatti, di elevata qualità, generabili grazie all'IA a costi bassissimi e senza richiedere particolari competenze tecniche: questo genere di contenuti generati risultano particolarmente convincenti data la combinazione multimediale, che aumenta la verosimiglianza delle *fake news*.

Inoltre, per ora non si sono registrati progressi nel prevenire un abuso dello strumento: un'analisi di NewsGuard⁸¹ ha rilevato che il *chatbot* alimentato da GPT-4 è più suscettibile a generare disinformazione e più convincente nel farlo rispetto al suo predecessore, ChatGPT-3.5. Un simile innovativo strumento richiede dunque che i media siano in grado di indagare e contestualizzare una mole crescente di disinformazione, nonché di

⁸⁰ L'introduzione di *standard* regolamentari può accompagnare imprese e consumatori finali nei casi in cui i servizi e i beni producibili con l'impiego dell'IA presentano i tratti di beni meritori o comportano esternalità.

⁸¹ Azienda che offre un sistema di *rating* per combattere la disinformazione.

sostenere e coltivare nuove capacità di giudizio da parte del pubblico, senza diffondere tuttavia il dubbio che **nessuna notizia sia affidabile**, sentimento che potenzialmente potrebbe arrivare a minare il funzionamento stesso delle democrazie⁸².

Nonostante questi rischi, secondo alcuni giornalisti **l'IA può cambiare il modo in cui interagiamo con le informazioni** anche in positivo, permettendoci di acquisire enormi quantità di dati e rendendo meno necessario possedere elevate competenze tecniche sull'interpretazione degli stessi. In ogni caso, i professionisti dell'informazione riconoscono generalmente la necessità che qualsiasi contenuto generato dall'intelligenza artificiale venga **controllato da un essere umano** al fine di rispettare la deontologia della professione giornalistica⁸³.

Il rischio di **polarizzazione delle informazioni**, ovvero la frammentazione della rappresentazione della realtà offerta dai media, pur essendo un fenomeno già presente, è ulteriormente aggravato dall'utilizzo del *machine learning* nei sistemi di raccomandazione (*recommendation system*), ovvero quei sistemi che, a partire dall'analisi dei *big data* e profilando i consumatori/utenti, sono in grado di raccomandare prodotti o servizi in base alle caratteristiche di ciascuno. Tali sistemi, applicati all'informazione, creano un effetto "bolla" (*filter bubble*)⁸⁴, che può tradursi anche in una crescente radicalizzazione dei contenuti consumati da ciascuno. Questi effetti, per di più combinati con il malcontento potenzialmente derivante da un inasprimento della polarizzazione dei salari, potrebbero tradursi in un notevole rischio per la tenuta democratica.

b. Responsabilità algoritmica

L'utilizzo della grande disponibilità di dati in combinazione con i nuovi modelli di intelligenza artificiale può risultare vantaggioso per le imprese **dal punto di vista dell'efficienza**; tuttavia, esso può sollevare questioni che vanno oltre il rischio di *bias* derivanti dai dati utilizzati per il *training*.

Ad esempio, in molte aziende gli algoritmi vengono già utilizzati per sorvegliare i risultati ottenuti dai lavoratori, ma anche per valutarne la salute, la personalità e il comportamento sociale. In primo luogo, questa tecnologia permette, quindi, di **monitorare molto più accuratamente i lavoratori**, anche rispetto a qualità e comportamenti che, pur influenzando potenzialmente sulla produttività, non sono trattati nei contratti di lavoro. In secondo luogo, nonostante tecnicamente l'IA non sia un soggetto in grado di esprimere scelte e di assumere la responsabilità delle conseguenze da esse derivanti, l'impiego di algoritmi e IA porta a **sollevare, almeno parzialmente, i manager dalla responsabilità** delle decisioni motivate dai risultati prodotti dall'IA e, quindi, di responsabilizzare maggiormente i

⁸² Come sottolineato dalla vice-presidente della Commissione Europea, Margrethe Vestager: "If we end up in a situation where we believe nothing, then we have undermined our society completely."

⁸³ Il rapporto "Generating Change: A global Survey of what News Organizations Are Doing with AI" (2023), realizzato nell'ambito dell'iniziativa JournalismAI della London School of Economics and Political Science (LSE), è basato su sondaggi distribuiti presso oltre 100 testate giornalistiche di 46 Paesi in merito all'impiego dell'intelligenza artificiale nel settore. L'indagine è stata condotta tra aprile e luglio 2023. Sebbene l'IA sia generalmente percepita come uno strumento utile, oltre il 60% degli intervistati ha espresso preoccupazione per le implicazioni etiche dell'intelligenza artificiale sui valori giornalistici, tra cui l'accuratezza, l'equità, la trasparenza.

⁸⁴ Whittaker, Joe, et al. "Recommender systems and the amplification of extremist content." (2021).

lavoratori stessi: ad esempio, nei casi in cui dal raggiungimento di determinati parametri può dipendere una promozione, una retrocessione o anche un licenziamento⁸⁵.

In generale, andando oltre il caso del controllo sui lavoratori, si pone la questione di come ripartire **le responsabilità delle decisioni assunte con il supporto dell'IA** tra coloro che hanno determinato le modalità di sviluppo dei modelli e coloro che hanno stabilito le modalità di impiego degli stessi.

Valutazioni simili si possono applicare all'impiego dell'IA **nel settore pubblico**. L'Osservatorio OCSE sull'innovazione nel settore pubblico esplora l'impatto dell'impiego dell'intelligenza artificiale, esaminando le strategie di utilizzo dell'IA in casi studio ed evidenziando le opportunità e le sfide legate all'uso dell'intelligenza artificiale nella pubblica amministrazione, soffermandosi sulla questione della "responsabilità algoritmica" (*algorithmic accountability*). Infatti, incorporare l'intelligenza artificiale nella progettazione e nella fornitura di politiche e servizi pubblici, ovvero nello stesso processo decisionale, richiede di garantire che gli algoritmi e i dati utilizzati per addestrarli non conducano a **discriminazioni o allucinazioni**, e di formare i funzionari pubblici affinché sappiano utilizzare l'IA **valutandone criticamente l'operato**.

c. Geopolitica e nuove dipendenze strategiche

L'impatto dell'IA sullo scacchiere geopolitico potrebbe passare attraverso l'evoluzione dei flussi commerciali⁸⁶, ovvero attraverso l'emergere di **nuove dipendenze strategiche**. Una diffusione massiccia dell'IA nei settori produttivi, nella società e nel settore pubblico, infatti, renderebbe indispensabili, per competere sul mercato mondiale, i tre beni strumentali⁸⁷ (e le relative materie prime) necessari a realizzarla. Le nuove dipendenze potranno riguardare, ad esempio, la produzione dei microprocessori e l'estrazione dei minerali necessari, la raccolta e la lavorazione dei dati, lo sviluppo dei modelli e la proprietà intellettuale sui modelli. Questo ultimo punto si pone in particolare per le economie europee, dato che le imprese private proprietarie dei modelli sono ad oggi soprattutto americane e cinesi.

Il controllo delle risorse strategiche necessarie alla realizzazione dei modelli IA, e quindi dell'IA stessa, rafforzerà le imprese e gli Stati che ne dispongono anche sul piano geopolitico. Anu Bradford (Columbia), autrice

⁸⁵ A questo proposito, il *think tank* americano Brookings, sottolinea, anche la necessità di garantire la trasparenza dei processi decisionali svolti dalle macchine: "un problema chiave per l'intelligenza artificiale che sostituisce i *manager* è che le decisioni dell'intelligenza artificiale possono apparire opache e confuse, lasciando ai lavoratori di interrogarsi sul come e sul perché sono state prese determinate decisioni, e sul cosa possono fare quando dati errati distorcono le decisioni dell'IA. I decisori politici potrebbero approvare regole per aumentare la trasparenza per i lavoratori circa le modalità tramite le quali vengono assunte le decisioni nella gestione algoritmica. Tali norme potrebbero costringere i datori di lavoro e le aziende che operano su piattaforme di lavoro online a rivelare quali dati vengono utilizzati per quali decisioni. Ciò aiuterebbe a contrastare l'attuale asimmetria informativa tra piattaforme e lavoratori".

⁸⁶ Ad es. nel settore dei servizi, come già analizzato nel paragrafo 3b.

⁸⁷ Ricordiamo che la catena del valore del settore IA coinvolge tre principali beni strumentali: (i) grandi collezioni di dati, (ii) capacità di calcolo (ovvero processori); (iii) modelli informatici, come i *transformer*, detti anche *Foundation models*, e loro derivati, come gli LLM.

di un saggio sugli imperi digitali⁸⁸, ha individuato **due tipologie di poteri tecnologici**: da un lato, le sempre più potenti imprese private; dall'altro, i regimi autocratici, che presentano loro peculiari punti di forza (la Cina, ad esempio, si avvantaggia di una pervasiva raccolta dei dati dei cittadini, svolta di concerto da settore pubblico e imprese di proprietà statale). L'introduzione dell'IA nelle economie di tutto il mondo si intreccia con l'evoluzione di entrambi queste tipologie di poteri. Come sottolineato da uno studio del Parlamento Europeo (2021)⁸⁹, l'IA potrebbe svolgere un duplice ruolo sullo scacchiere geopolitico: da un lato contribuendo ad esacerbare le tensioni commerciali tra USA e Cina, dall'altro determinando un **aumento del potere dei giganti del settore privato**.

Per quanto riguarda i modi in cui l'IA può accrescere il potere dei regimi autoritari, si pensi a una applicazione dei nuovi modelli probabilistici al sistema cinese di "credito sociale", un sistema che utilizza un'ampia varietà di *input* di dati per valutare il punteggio di credito sociale di una persona (da cui derivano autorizzazioni a determinati comportamenti, come l'acquisto di un biglietto aereo o ferroviario). Nel corso della prima riunione del Trade and Technology Committee (2021), Stati Uniti e UE hanno espresso preoccupazione rispetto all'approccio cinese all'impiego dell'intelligenza artificiale e dei dati dei cittadini. In tal senso, il TTC è stato visto da alcuni osservatori come il primo passo per la nascita di **un'alleanza di Paesi democratici** volta a promuovere **un approccio allo sviluppo dell'intelligenza artificiale orientato ai diritti umani**, in aperto contrasto con gli approcci adottati da Paesi autoritari come Russia e Cina.

Lo studio del Parlamento Europeo raccomanda di collaborare con gli Stati Uniti per garantire un utilizzo dell'IA che non violi i valori democratici. Tuttavia, proprio rispetto a questo fine, non dovrebbero essere trascurati i **conflitti verticali tra Big Tech e autorità statali**, che proprio negli Stati Uniti assumono grande rilevanza. È infatti evidente che il controllo dell'IA da parte dei giganti privati americani non può che rendere ancora più forti gli imperi digitali commerciali descritti da Bradford, acuendo la concentrazione globale di potere. Per di più, l'analisi di Bradford illustra come i conflitti orizzontali tra Stati Uniti e Cina abbiano influenzato i conflitti verticali interni, tra gli Stati e le imprese: a parere dell'autrice, lo sforzo volto a superare i *competitor* cinesi avrebbe prodotto negli Stati Uniti **l'ascesa di imperi digitali privati poco regolamentati**, e in grado di diventare non meno temibili, per le democrazie, della competizione rappresentata dagli stessi modelli autoritari.

Infine, l'IA giocherà un ruolo crescente anche nel **settore della difesa**. Lo studio già citato del Parlamento Europeo elenca le applicazioni dell'IA nella difesa con funzione di sorveglianza e identificazione, gestione della logistica, progettazione delle azioni militari, controllo di veicoli senza pilota, ed elaborazione di previsioni.

⁸⁸ Bradford, Anu. *Digital empires: The global battle to regulate technology*. Oxford University Press, 2023.

⁸⁹ Franke, Ulrike. "Artificial Intelligence Diplomacy| Artificial Intelligence Governance as a New European Union External Policy Tool." Luxembourg: European Parliament, June (2021). Più precisamente, si tratta del Policy Department for Economic, Scientific and Quality of Life Policies, dipartimento appartenente alla direzione generale per le politiche interne all'Unione del segretariato generale del Parlamento Europeo.

Un utilizzo diffuso dell'IA in settori strategici, come la difesa, la sicurezza o il controllo di infrastrutture vitali, richiederà a sua volta una rinnovata attenzione per quanto concerne la *cybersecurity*, onde evitare che i sistemi di IA siano manomessi: questo aspetto rinforza la questione tutt'ora aperta circa il coinvolgimento di imprese provenienti da Paesi autoritari, come la Cina, nella produzione di infrastrutture digitali realizzate in Europa, nonché le questioni riguardanti la cessione all'estero di tecnologie *dual use*.

5. POLICY E REGOLAMENTAZIONE

L'OCSE sottolinea l'importanza delle *policy*, invitando i *policy makers* a non arrendersi al **determinismo tecnologico**⁹⁰. Le innovazioni di processo e dei sistemi organizzativi interne alle imprese, la riorganizzazione del mercato del lavoro e le trasformazioni culturali necessarie a bypassare i limiti costitutivi dei modelli IA, ovvero quei limiti non risolvibili attraverso il progresso tecnologico, rendono necessario l'intervento statale, chiamato a costruire le **condizioni di base necessarie a cogliere tutte le opportunità offerte da IA**, e al tempo stesso creano uno spazio di azione nel quale l'azione pubblica può **orientare il cambiamento** che le tecnologie IA hanno il potere di produrre.

Stefano Scarpetta (2023), direttore del Dipartimento Employment, Labour and Social Affairs dell'OCSE, sottolinea a tal proposito il **ruolo dell'innovazione normativa**. Alla luce di tale *framework* di analisi, tuttavia, l'approccio che caratterizza la legislazione europea sul tema appare piuttosto conservativo, non includendo elementi di valorizzazione da affiancare a una strategia industriale (si vedano gli *annex 6 e 7* per un'analisi degli interventi legislativi americani e europei). Occorrerebbe infatti valutare quali compiti possano essere svolti dalle macchine grazie all'apprendimento automatico e quali, invece, richiedano il contributo del lavoro umano, data le caratteristiche della tecnologia; ma anche, e forse soprattutto, **scegliere in quali ambiti il contributo del lavoro umano è ritenuto preferibile** a quello della macchina a prescindere dai guadagni economici in termini di efficienza, considerando quegli elementi della produzione che sono politicamente, socialmente o culturalmente rilevanti e attengono a scelte discrezionali.

In conclusione, riteniamo che le politiche europee dovrebbero disegnare una strategia per massimizzare i benefici di una tecnologia dirompente come l'IA in un sistema economico e sociale definito dai valori europei, anche in considerazione delle significative implicazioni geopolitiche dell'intelligenza artificiale. A tal fine, si potrebbe immaginare che, in presenza di notevoli evoluzioni della tecnologia oggi ai primi passi, la regolamentazione potrebbe essere chiamata a circoscrivere un ambito di **competenza esclusiva dell'intelligenza umana**⁹¹, non al solo scopo di

⁹⁰ Nell'Employment Outlook dell'OCSE è riportata un'eloquente citazione di David Autor (MIT): "Meditando sul nostro incerto futuro legato all'IA, il nostro obiettivo non dovrebbe essere semplicemente quello di predire quel futuro, ma di crearlo"

⁹¹ Come già discusso nel primo capitolo, sul lato della domanda, infatti, si possono produrre casi simili a quello dei beni meritori, in quanto i consumatori potrebbero non essere immediatamente consapevoli di alcune caratteristiche determinate dall'impiego dell'IA nella produzione di beni o servizi (ad es. riguardo alla cessione di dati o l'attendibilità delle informazioni offerte dall'IA) o casi con

proteggere i consumatori e gli utenti dai rischi, ma anche al fine di **promuovere un modello di sviluppo europeo**, che rappresenti, in base alle preferenze proprie delle diverse culture, un'alternativa a quello americano e cinese, anche in termini dei beni e dei servizi ivi prodotti.

Al contrario, la regolamentazione europea⁹² si limita per ora alla gestione del rischio, come previsto dalla prima proposta della Commissione risalente al 2021, puntando alla **prevenzione del danno** nel solco delle tradizionali norme di sicurezza che regolamentano l'uso di qualsiasi altro prodotto: in questo quadro, le recenti tecnologie di IA, come i *foundation models*, sono considerati potenzialmente ad alto rischio per la loro capacità di svolgere compiti differenziati e la loro posizione a monte della catena del valore.

Per sostenere i *policy maker* nella valutazione delle politiche necessarie alla gestione delle nuove tecnologie IA, il Fondo Monetario Internazionale ha costruito un indice di misurazione del livello di preparazione all'intelligenza artificiale (AIPI, ovvero *Artificial Intelligence Preparedness Index*), costituito da un insieme selezionato di indicatori macro-strutturali rilevanti per l'adozione dell'IA. Tali indicatori sono organizzati in quattro categorie, raggruppate in due livelli di preparazione:

- 1) Sostengono la realizzazione del livello di preparazione definito "*foundational AI preparedness*" gli indicatori raccolti nelle seguenti due categorie:
 - i) infrastrutture digitali;
 - ii) innovazione e integrazione economica;
- 2) Mentre il livello "*second-generation preparedness*", considerato più rilevante per le economie avanzate con un settore digitale sviluppato, sarebbe realizzato dagli indicatori raccolti nelle due categorie:
 - i) capitale umano e politiche del lavoro;
 - ii) regolamentazione ed etica.

I primi due paragrafi di questo capitolo riguardano rispettivamente il terzo e il quarto punto, ovvero il livello di "*second-generation preparedness*", mentre i paragrafi c e d si occupano di aspetti concernenti la preparazione di base. Infine, l'ultimo paragrafo considera le politiche IA nel quadro dell'autonomia strategica europea.

a. Istruzione, formazione, informazione e riqualificazione

Un'azione trasversale a diversi ambiti della preparazione, che accresce la capacità di cogliere le opportunità e al contempo riduce i rischi, riguarda **l'offerta di istruzione, formazione e informazione**.

esternalità negative o positive. Sul lato dell'offerta, l'impiego dell'IA nella produzione di beni e servizi, non integrato dal controllo del lavoro umano in alcuni passaggi chiave, può sollevare altre questioni (es. la responsabilità nella gestione dei lavoratori può essere assunta solo dagli esseri umani che controllano l'operato dell'IA).

⁹² Si veda annex 7.

Sul piano politico, riteniamo che sarà sempre più importante informare i cittadini sulle nuove sfide poste dall'IA alla stessa **tenuta democratica** (cfr. capitolo 4a), anche regolamentando attentamente il settore dei media⁹³.

A tal fine assume rilevanza il problema già sollevato rispetto all'incapacità degli Stati di studiare i modelli di grandi dimensioni.⁹⁴ Il Foundation Model Transparency Index di HAI Stanford tiene traccia della quantità di informazioni divulgate dalle imprese circa i loro modelli di intelligenza artificiale. I risultati ottenuti dai principali LLM oggi esistenti non sono incoraggianti: Llama 2 (il modello LLM di Meta) ha ottenuto il punteggio più alto in termini di trasparenza con il 54%, seguito da BloomZ e GPT-4 di OpenAI. **Nessun modello linguistico ha ottenuto punteggi elevati** per quanto riguarda le fonti dei dati utilizzati per il *training*, l'impatto sociale e i potenziali *bias* riflessi nei modelli.

Sul piano economico, formazione e riqualificazione dovranno essere mirate a rispondere alle **nuove necessità del mercato del lavoro**. Per valorizzare l'effetto 'reintegrazione', già discusso nel capitolo 3, infatti, sarà cruciale individuare le **competenze necessarie ai lavoratori** per integrare, valutare e gestire il lavoro dell'IA. L'utilizzo di sistemi operativi basati su IA dovrebbe condurre a una maggiore intuitività delle interfacce: la capacità di utilizzare il linguaggio naturale nelle interazioni uomo-macchina permetterebbe alla macchina di servire gli utenti senza che questi debbano imparare ad usarla. Al contempo, una formazione più complessa, e di più difficile trasmissione, potrebbe rendersi necessaria per permettere agli utenti di esprimere giudizi liberi e competenti sul lavoro della macchina stessa. Secondo questa logica, **le competenze maggiormente necessarie nel futuro** potrebbero includere a nostro avviso le **competenze tecnologiche di livello elevato** e le competenze umanistiche o di base come il **senso critico**, a discapito delle *skill* tecniche di livello meramente operativo, rese superflue dalle nuove interfacce intuitive dei macchinari, basate sul linguaggio naturale. In altre parole, potrebbe diventare necessario distinguere tra capacità d'uso dello strumento e padronanza del mezzo: mentre la prima potrebbe essere maggiormente diffusa grazie al *design* di interfacce ancora più intuitive; in assenza di altra formazione, verrebbe a mancare una vera padronanza del mezzo e un reale controllo dei risultati ottenuti dal suo impiego.

Consideriamo altrettanto strategico preparare i consumatori a una scelta consapevole di beni e servizi (sostenuta da una robusta regolamentazione sugli *standard* di qualità). Infatti, una più diffusa **consapevolezza presso i consumatori** di beni e servizi e l'imposizione di ***standard* di qualità alle imprese** permetterebbero di **creare un**

⁹³ In Italia, è stato istituito un comitato presso il Dipartimento per l'informazione e l'editoria della Presidenza del Consiglio dei ministri, al fine di studiare l'impatto dell'IA sul mondo del giornalismo e delle news. Il Comitato, formato da esperti e professori universitari, è stato presieduto inizialmente dal giurista Giuliano Amato, successivamente sostituito da Padre Benanti, frate francescano che insegna alla Pontificia università Gregoriana, unico italiano membro del Comitato sull'Intelligenza Artificiale delle Nazioni Unite (AI Advisory Body, si veda l'annex 6), istituito a novembre 2023.

⁹⁴ I modelli delle imprese private sono circa 30 volte più grandi di quelli sviluppati dalla ricerca universitaria. In alcuni casi, le imprese private investono nelle università e mettono a disposizione i propri mezzi per dare alle università l'opportunità di studiare i modelli privati. Tuttavia, è indubbio che le ricerche prodotte di questi accordi non possono essere considerate a tutti gli effetti quali ricerche indipendenti, data l'influenza che le imprese private possono esercitare sui ricercatori che vi prendono parte.

mercato per i beni e i servizi che incorporino le caratteristiche inimitabili del lavoro umano, nonostante questi richiedano costi di produzione superiori rispetto alle alternative generate da IA.

b. Regolamentare le modalità di sviluppo dei modelli di IA e gli ambiti di applicazione

Riguardo a come sia possibile evitare *bias*, allucinazioni e altri rischi derivanti dalla fallacia dei modelli probabilistici, esiste **un dibattito** riassumibile in due posizioni estreme. Alcuni ritengono sia **possibile mitigare i pregiudizi umani** contenuti nei dati utilizzati per il *training* costruendo basi di dati appositamente progettate per evitare rappresentazioni della realtà che possono produrre comportamenti discriminatori nell'IA⁹⁵. Altri sostengono che sia piuttosto necessario scegliere con attenzione quali decisioni possono essere basate sui risultati prodotti da un'intelligenza artificiale e quali invece debbano ancora **essere affidate al lavoro umano**⁹⁶. In considerazione dell'analisi effettuata, riteniamo che il regolatore potrebbe essere chiamato a stabilire degli *standard* da rispettare sia riguardo alle qualità delle banche dati utilizzate per l'addestramento, sia riguardo ai criteri in base ai quali è possibile delegare attività e decisioni, totalmente o parzialmente, all'IA.

Una possibile via per individuare un ambito di esclusiva competenza dell'"intelligenza naturale", ad esempio, si può derivare a partire dalle riflessioni circa la **caratteristica dell'intelligenza umana di essere 'incarnata'** e considerare **la componente relazionale contenuta** nella maggior parte delle professioni. Comprendere il legame tra l'intelligenza umana e il corpo ad esso collegato potrebbe giocare un ruolo fondante sia nella ricerca volta a garantire l'allineamento dell'IA alle richieste implicite degli operatori umani, sia per valutare quali compiti possono essere svolti in modo profondamente diverso, e talvolta insufficiente, dall'intelligenza artificiale, proprio a causa della sua natura incorporea⁹⁷.

La regolamentazione di prodotti e servizi, pertanto, è necessaria non solo a garantire la qualità degli stessi, ma soprattutto a **rendere obbligatoria l'integrazione del lavoro umano in tutti i casi in cui esso è ritenuto necessario** a evitare automatismi pericolosi (si veda il caso degli algoritmi già utilizzati per decidere su assunzioni o licenziamenti) o semplicemente nei casi in cui l'IA non in grado di eguagliare il lavoro umano sotto profili apparentemente esterni alle qualità del prodotto o del servizio, e quindi di

⁹⁵ Le tecniche di lavorazione dei dati possono essere molto complesse e permettono di etichettare i dati in vari modi, anche per allertare la macchina circa quelli potenzialmente in grado di generare *bias*, questo comporta un maggiore impiego del lavoro umano, che rende i dati etichettati (*labeled*) maggiormente costosi di quelli *unlabeled*.

⁹⁶ Questa è la posizione assunta da Hinda Hamed, docente di Data Science all'Informatics Institute, presso l'università di Amsterdam, in occasione di un dibattito sui pericoli dell'IA organizzato durante i Bruegel Annual Meetings di quest'anno.

⁹⁷ Ciò è vero per ambiti quali la sanità, la cura dell'infanzia, l'istruzione, l'assistenza sociale, ma anche per settori come la giustizia e la pubblica sicurezza, fino ad arrivare all'intrattenimento, o alla produzione artistica: ovvero, riguarda tutti gli ambiti che trattano elementi del corpo o dell'esistenza umana, come i diritti individuali o il senso estetico e morale, che potrebbero risultare profondamente incomprensibili per un'intelligenza disincarnata, automatica e inanimata.

difficile valutazione per l'individuo⁹⁸, o in grado di produrre **esternalità positive o negative** per la collettività.

Infine, la regolamentazione dovrebbe assicurare in maniera stringente la **protezione dei dati degli utenti e dei lavoratori**. Nel mondo del lavoro, particolare protezione dovrebbe essere dedicata alla potenziale cessione delle **competenze dei lavoratori e del know how delle imprese europee**, al fine di impedirne l'acquisizione gratuita, e talvolta inconsapevole, da parte delle aziende fornitrici dei servizi di ML.

c. Politica industriale

Come già accennato, lo sviluppo in Europa di un settore dell'intelligenza artificiale parallelo a quello americano e cinese è ostacolato dalla capacità delle imprese già attive nel settore di sfruttare la loro **posizione dominante** e dal ritardo accumulato dall'industria europea. A favore delle *Big tech* che operano da decenni nel **settore a monte**, ovvero quello della raccolta dati, gioca la natura a rete dell'industria (*social network*, motori di ricerca) che avvantaggia le grandi aziende. Costituiscono una notevole barriera all'entrata anche gli **enormi investimenti in termini di potenza computazionale** che sarebbero richiesti per partire dallo stato attuale, come dimostra il fatto che molte *startup* americane dell'ambito IA hanno costruito collaborazioni fondamentali con delle *Big Tech* proprio per accedere alla loro capacità computazionale (ad esempio, OpenAI con Microsoft, Anthropic con Alphabet e Amazon).

Nonostante ciò, pur riconoscendo i problemi derivanti dal dominio degli Stati Uniti e della Cina nel settore, uno studio sull'IA prodotto dal Parlamento Europeo afferma che, per quanto riguarda lo sviluppo dell'IA "non esistono ragioni fondamentali per cui l'Europa non dovrebbe essere in corsa per la medaglia di bronzo". A tal fine, a gennaio, la Commissione ha avviato una serie di iniziative volte a favorire lo sviluppo del settore IA in Europa: per quanto riguarda la capacità computazionale, si mira a permettere l'accesso delle *startup* innovative ai *supercomputer* pubblici disponibili nei Paesi UE⁹⁹. Attorno a questi *supercomputer* si punta a sviluppare ecosistemi denominati "*AI factories*", anche grazie a iniziative che riguardano gli altri elementi della catena del valore del settore IA¹⁰⁰. Per quanto riguarda i dati, la Commissione ha annunciato l'intenzione di realizzare un mercato unico, espandendo gli Spazi Europei Comuni dei Dati già avviati (es. lo Spazio Europeo Comune dei dati sanitari). Altre iniziative annunciate dalla Commissione nel 2024 riguardano lo sviluppo dei modelli (*Large AI Grand Challenge*, *Alliance for Language*

⁹⁸ I beni e i servizi prodotti da esseri umani potrebbero avere, rispetto alle alternative prodotte dall'IA, qualità e tratti simili a quelli dei beni meritori.

⁹⁹ Si tratta di tre *supercomputer* che figurano nei primi dieci posti della classifica 'top500' che dal 1993 elenca i computer più potenti del mondo: LUMI in Finlandia (5° posto della classifica), LEONARDO in Italia (6° posto) e MareNostrum5 in Spagna (8° posto). Le prime tre posizioni sono occupate da *supercomputer* americani, i primi due di proprietà pubblica (Department of Energy), mentre è di proprietà privata il terzo (Microsoft). La quarta posizione è occupata da un *supercomputer* giapponese, mentre sono americani e di proprietà privata (IBM e NVIDIA) i *supercomputer* che occupano la settima, la nona e la decima posizione.

¹⁰⁰ Si noti che la realizzazione dell'integrazione verticale in Europa avverrebbe con l'ausilio del settore pubblico anziché internamente al settore privato, come avvenuto negli Stati Uniti.

Technologies) e la promozione del capitale umano necessario a costruire un settore IA europeo solido¹⁰¹.

Accanto a queste iniziative volte a competere con i settori IA americani e cinesi, a nostro avviso possono e devono essere adottate misure per **favorire la disseminazione del cambiamento tecnologico nell'economia**, in modo da superare il più velocemente possibile il ritardo tipicamente osservato tra realizzazione dell'innovazione e suoi effetti sulla produttività. A tal fine occorrono politiche volte a **favorire l'adozione delle tecnologie IA**, anche prodotte all'estero, al fine di aumentare la produttività del lavoro: tali politiche devono essere mirate sulle esigenze delle imprese che caratterizzano i diversi territori europei. Ad esempio, nella strategia per l'IA nazionale tedesca si prevede di aumentare il sostegno specifico all'adozione dell'IA inviando **"formatori di intelligenza artificiale"** nelle aziende di piccole e medie dimensioni per lavorare sull'integrazione dell'intelligenza artificiale nei loro sistemi produttivi.

Riteniamo, inoltre, che occorrono politiche in grado di **favorire la produzione in Europa di dispositivi e software basati sull'IA** realizzando politiche volte a incoraggiare le **innovazioni collaterali** all'IA, siano esse di processo o di prodotto. Questo potrebbe richiedere una serie di interventi pubblici per realizzare le necessarie condizioni di contesto, ad esempio rimuovendo i colli di bottiglia, come le difficoltà osservate in UE per l'accesso al credito da parte delle imprese digitali. Infatti, ad oggi mancano nell'UE sia imprese *leader* nel settore dello sviluppo dei modelli di intelligenza artificiale, sia imprese focalizzate sulla creazione di **prodotti e servizi derivati dall'impiego di IA**: il mercato di tali prodotti e servizi è ancora in via di sviluppo.

La rilevanza del settore delle innovazioni collaterali appare tanto più notevole per l'UE se si considera che le grandi aziende sviluppatrici dei modelli IA, come Google e Microsoft, rappresentano dei *competitor* temibili nel settore dello sviluppo dell'intelligenza artificiale generativa, mentre ciò è meno vero per le 150 *startup* americane oggi impegnate nello sviluppo di applicazioni basate sui "modelli *foundation*" creati dalle *Big tech*. Tali innovazioni sono peraltro considerate necessarie a rendere profittevoli i nuovi modelli, ovvero a permettere ai modelli "*general purpose*" di svolgere efficientemente mansioni specifiche.

Inoltre, considerando il particolare rapporto tra automazione tradizionale e intelligenza artificiale, e la rilevanza dell'automazione per l'industria europea, una **possibile area di ricerca da sviluppare nell'Unione Europea** potrebbe riguardare proprio il connubio tra *robot* e IA.

Per quanto riguarda, invece, le imprese e le PMI impegnate nella produzione di beni e servizi che **derivano le loro peculiari caratteristiche dall'impiego del lavoro umano**, consideriamo necessario proteggere il loro mercato attraverso gli usuali strumenti, quali certificati di autenticità, campagne di informazione dei consumatori, e, qualora sia ritenuto che le qualità di tali beni o servizi siano essenziali, la predisposizione di *standard* minimi che escludano la possibilità di realizzare gli stessi beni e servizi tramite l'IA.

¹⁰¹ Communication from the Commission on boosting startups and innovation in trustworthy artificial intelligence, 24.1.2024, COM(2024) 28 final.

La **regolamentazione degli standard qualitativi può quindi svolgere un ruolo anche nella promozione dell'industria domestica europea**, salvaguardandola dalla concorrenza delle grandi imprese estere che potrebbero adottare le tecnologie legate all'IA riuscendo a ridurre i costi di produzione e quindi i prezzi.

Tuttavia, non dobbiamo dimenticare che una regolamentazione mal progettata può avere l'effetto avverso di **scoraggiare l'adozione delle nuove tecnologie e inibire lo sviluppo di interi settori**. Secondo un *paper* realizzato nell'ambito dell'iniziativa sullo *Scientific Foresight* del centro di ricerca del Parlamento Europeo (2022), in passato, alcuni settori fortemente regolamentati non hanno saputo rispondere con prontezza alle opportunità di innovare. In altri casi, le normative volte a incoraggiare l'adozione delle "migliori tecniche disponibili" hanno promosso l'innovazione a breve termine a discapito di un'innovazione trasformativa più ambiziosa.

Nonostante queste possibilità, la regolamentazione può fornire anche le condizioni necessarie a consentire l'accesso al mercato alle innovazioni, al contempo garantendo per utenti e consumatori che l'IA sia affidabile¹⁰², e **creando un ambiente certo per le imprese che valutano di compiere investimenti** in un settore in evoluzione, favorendo la libera concorrenza¹⁰³.

In conclusione, sebbene la regolamentazione possa talvolta ostacolare l'innovazione, una combinazione di misure attentamente progettate e implementate potrebbe invece costruire le condizioni ottimali per un percorso di innovazione desiderabile¹⁰⁴.

d. Politiche distributive

La gestione dei cambiamenti sul mercato del lavoro, ovvero il potenziale aggravarsi del fenomeno della polarizzazione dei salari, richiederà di **proteggere i lavoratori che potrebbero fuoriuscire dal mercato del lavoro**, non potendo questi essere riqualificati, e mitigare le disuguaglianze emergenti tra lavoratori con diversi livelli e diversi tipi di competenze.

Alcuni autori, portando alle estreme conseguenze le ipotesi sulla sostituibilità del lavoro umano da parte delle future versioni dei modelli di IA, hanno immaginato che l'impiego del *machine learning* possa condurre un giorno alla **ridondanza del lavoro umano**, o per lo meno a ridurre la quota di reddito distribuita ai lavoratori al punto che essa non sia sufficiente alla sopravvivenza¹⁰⁵. Ad esempio, in uno studio realizzato dal *think tank* americano Brookings, si afferma che in futuro saranno necessarie **nuove**

¹⁰² L'OCSE ha realizzato una serie di linee guida per la promozione di una "trustworthy AI".

¹⁰³ La regolamentazione potrebbe controbilanciare, ad esempio, il potere dei monopoli della conoscenza, che possono utilizzare il Machine Learning per avvantaggiarsi di asimmetrie informative, soffocando la concorrenza.

¹⁰⁴ Nicholas, BOUCHER Philip. "What if AI regulation promoted innovation?" (2022). Ripreso dal sito ethinktank.eu.

¹⁰⁵ Questo rischio è aumentato nelle economie avanzate dalla combinazione di IA e 'remote intelligence', ovvero lavoro da remoto importato da Paesi con basso costo del lavoro, si ricorda a tale proposito quanto osservato nel paragrafo 3b circa la possibilità che nei Paesi avanzati non operi a pieno regime l'effetto 'produttività', in grado di compensare la sostituzione del lavoro con IA.

istituzioni economiche per distribuire il prodotto in base a criteri estranei al valore di mercato del lavoro¹⁰⁶, come una maggiore distribuzione della proprietà del capitale (ipotesi che, come già detto, potrebbe essere motivata dalla partecipazione dei lavoratori all'addestramento dell'IA) o l'erogazione di *benefit* finanziati con un aumento dell'imposizione fiscale sul capitale stesso (in particolare sull'IA).

Anche qualora la riorganizzazione della produzione non avvenisse in modalità così estreme, la modifica della distribuzione delle quote di prodotto tra fattori potrebbe richiedere in ogni caso un **ripensamento del sistema di imposizione fiscale e di welfare**. Per questo riteniamo opportuna una valutazione del livello di tassazione da applicare sulla proprietà e sulle rendite generate da un capitale particolare quale un modello di IA.

e. Funzione geopolitica della regolamentazione

Nel 2021 la Commissione europea ha proposto l'istituzione del "Consiglio transatlantico per il commercio e la tecnologia" (TCC), al fine di stabilire degli **standard comuni** a Stati Uniti ed Europa sulle nuove tecnologie, inclusa l'IA. Lo studio del PE (2021), già più volte citato, ritiene consigliabile evitare di collocare la **cooperazione transatlantica sull'IA** nel contesto del confronto tra Stati Uniti e Cina.

Anu Bradford, autrice che ha coniato il termine *Bruxelles effect*, già citata nel precedente capitolo¹⁰⁷, ritiene che ci sia **spazio per l'UE per giocare autonomamente un ruolo fondamentale nella regolamentazione degli imperi digitali privati** che hanno sede negli Stati Uniti e che saranno ulteriormente rafforzati dalla diffusione dell'IA nell'economia globale. Diversamente da quanto proposto dalla Commissione riguardo gli *standard* comuni, Bradford sostiene che sfruttando per l'appunto "l'effetto Bruxelles", basato sulla preferenza mostrata dalle imprese a uniformare i propri processi produttivi, l'UE potrebbe riuscire a **esportare l'elevata qualità della propria regolamentazione** anche all'estero, e in particolare negli Stati Uniti, come già avvenuto in passato per altri settori, ottenendo in tal modo di **mitigare il vantaggio tecnologico americano**. L'autrice ritiene che l'approccio europeo alla gestione delle imprese digitali, ovvero l'imposizione di elevati *standard* di qualità e il rispetto dei diritti fondamentali, possa rappresentare una valida alternativa al permissivo modello americano, rispondendo anche a **esigenze sentite da consumatori, utenti e cittadini al di là dell'Atlantico**, ma inascoltate dai governi, impegnati a coltivare campioni americani in grado di arginare le imprese controllate dai regimi autoritari¹⁰⁸.

Lo studio del PE riconosce questa possibile strategia, identificando nello sforzo dell'UE per garantire un'IA etica e affidabile, un **punto di forza unico per l'Europa**, in grado di sollevare delle complicazioni per la

¹⁰⁶ Korinek, Anton, and Megan Juelfs. Preparing for the (non-existent?) future of work. No. w30172. National Bureau of Economic Research, 2022.

¹⁰⁷ Nel paragrafo 4c sugli impatti geopolitici dell'IA, in quanto autrice del libro *Digital Empires*.

¹⁰⁸ Anu Bradford e Fiona M. Scott Morton (Bruegel) hanno ipotizzato che ciò sia anche effetto del fatto che le Big Tech potrebbero avere un'influenza significativa sul dibattito politico americano, controllando i media sociali, e sugli stessi candidati (influenza esercitata attraverso le donazioni poco trasparenti alle campagne elettorali).

cooperazione transatlantica. Infatti, secondo lo studio del PE, degli elevati *standard* regolamentari potrebbero attribuire **un vantaggio all'Europa**, in modo molto simile a quanto avvenuto per la protezione dei dati garantita dal GDPR: molti consumatori, ovunque nel mondo, potrebbero preferire utilizzare o acquistare prodotti e servizi realizzati con l'impiego dell'IA (prodotta negli Stati Uniti), ma garantiti da un marchio "*made in Europe*". Lo studio del PE ritiene che collaborare con un attore rilevante nel settore come gli Stati Uniti potrebbe permettere di stabilire *standard* di qualità internazionali per l'IA; tuttavia, considerando anche la possibilità di creare un vantaggio commerciale per l'Europa, riconosce che la cooperazione transatlantica su questo tema si rivelerebbe controproducente, poiché minerebbe l'unicità dell'UE.

A parere di Bradford, negli Stati Uniti non esiste una volontà politica di imporre limitazioni di alcun tipo alle imprese nazionali. Secondo l'autrice, il rischio per l'UE non risiede in un'eccessiva regolamentazione europea, ma in **un'implementazione insufficiente** della stessa (specie nei confronti dei giganti americani). Infatti, una regolamentazione ambiziosa e ben progettata potrebbe favorire lo sviluppo di un **settore europeo delle innovazioni collaterali** all'IA sostenuto da *standard* di qualità competitivi. Concordiamo con Bradford nel ritenere che la regolamentazione europea potrebbe riuscire a intercettare le preferenze dei cittadini europei e americani, e realizzare un effetto Bruxelles anche nel settore dell'IA producendo effetti su scala globale e controbilanciando il potere statunitense sul piano dello sviluppo dei modelli.

Per Bradford, inoltre, un eventuale fallimento dei governi democratici nella gestione dei giganti del settore digitale rischierebbe di evidenziare la superiorità dei regimi non-democratici sotto il profilo della capacità di regolamentare e applicare la regolamentazione. Al contrario, l'imposizione di una ambiziosa regolamentazione europea quale *standard* globale dimostrerebbe la capacità delle democrazie di arginare il potere delle grandi imprese digitali, diversamente da quanto avvenuto finora negli Stati Uniti. Al contempo, una simile strategia **garantirebbe all'Unione Europea l'esercizio di un 'soft power'**, mediante il quale difendere gli interessi e i valori dei cittadini europei, nonostante l'emergere di potenziali nuove dipendenze.

6. CONCLUSIONI

I nuovi modelli di intelligenza artificiale possono svolgere funzioni precedentemente inimmaginabili per una macchina. Nonostante le oggettive limitazioni nel replicare completamente l'intelligenza umana, l'IA ha il potenziale per automatizzare e ottimizzare un'ampia gamma di compiti oggi svolti dal lavoro umano, per la prima volta anche quello qualificato.

Questo "capitale che lavora" può essere impiegato **trasversalmente ai settori produttivi**, nonché nell'ambito del *policy making*, rappresentando una grande opportunità di aumentare l'efficienza e ridurre i costi di produzione, e qualificandosi quale strumento utile alla gestione di macro-tendenze come il cambiamento climatico, la sfida energetica e l'invecchiamento demografico.

Alcuni dei rischi derivanti dall'adozione pervasiva dell'IA sono riconducibili al piano della sicurezza e dell'affidabilità, nonché alle questioni riguardanti la trasparenza e l'attribuzione di responsabilità, le dipendenze strategiche, la potenziale deumanizzazione di alcuni segmenti dell'economia. Infine, l'impiego estensivo dell'IA solleva preoccupazioni riguardo alla **possibile perdita di posti di lavoro**, all'aumento della disuguaglianza dei redditi da lavoro e alla diminuzione della quota di prodotto distribuita al fattore lavoro a livello macroeconomico: in futuro, simili fenomeni potrebbero mettere sotto pressione i sistemi di *welfare* europei.

Inoltre, abbiamo illustrato nella nota come l'affermarsi di tecniche di *machine learning* sempre più efficaci richiede all'economista di **ripensare il funzionamento del processo di accrescimento dello stock di capitale fisico**, poiché ogni forma di impiego dei modelli IA rappresenta anche un momento di apprendimento per i modelli e quindi un investimento nel capitale IA, che può accrescere la quantità e la qualità delle capacità dei modelli.

In considerazione dell'analisi svolta, pensiamo che l'Unione Europea dovrebbe porsi ad un tempo un insieme variegato di obiettivi, valutando di volta in volta a quali di questi attribuire maggiore rilevanza:

1. Trovare il giusto bilanciamento tra massimizzazione dei benefici economici dell'intelligenza artificiale e mitigazione dei rischi, definendo e gestendo il *trade-off* che li lega;
2. Per quanto riguarda la massimizzazione dei benefici, adottare una strategia complessiva coerente che miri a:
 - a. Accelerare l'integrazione dell'IA nell'economia, facilitando **l'adozione di soluzioni IA** da parte delle imprese europee, ponendo particolare cura nel sostegno ai segmenti dell'economia meno digitalizzati (es. PMI) e valutando la rilevanza dell'integrazione dei modelli IA nella manifattura europea, caratterizzata da elevatissimi livelli di automazione tradizionale;
 - b. Adottare misure volte ad incentivare lo sviluppo in Europa di **innovazioni collaterali basate sui nuovi modelli IA** nonché di innovazioni nei prodotti e nei modelli di *business* che possono realizzarsi grazie all'impiego dei modelli *transformer*. A tal fine riteniamo che possa essere utile valutare la possibilità di sfruttare il buon posizionamento a livello globale di alcune imprese operanti nel settore della robotica;
 - c. Promuovere **lo sviluppo di un settore IA europeo**, sfruttando la scalabilità che caratterizza i processi di addestramento dei modelli e valorizzando l'ampiezza del Mercato Unico, pur consapevoli del ritardo accumulato rispetto agli altri grandi attori globali. Ciò implicherebbe dar seguito alle iniziative della Commissione per realizzare lo spazio comune dei dati e favorire l'accesso ai tre supercomputer presenti nel continente per le *start-up* innovative;
3. Fare del governo dei rischi il proprio marchio distintivo, posizionandosi quale **leader mondiale nella definizione di standard** etici e di sicurezza ed esercitando la propria influenza sugli *standard* globali, come già avvenuto in altri contesti (ad es. con gli *standard* contenuti nel GDPR): riteniamo, infatti, che coltivare un

elevato livello di fiducia presso consumatori, investitori e altri *stakeholder*, grazie alla credibilità della regolamentazione europea, possa rappresentare un importante vantaggio per le imprese dell'UE

Riconosciamo la necessità di proteggere **la competitività dei sistemi produttivi europei**, bilanciando innovazione tecnologica e protezione dei diritti, e riteniamo che questo obiettivo possa essere raggiunto assicurando che l'intelligenza umana continui a presidiare i passaggi di produzione maggiormente sensibili, e che l'UE, regolamentando l'entità del contributo umano per tali produzioni, possa contribuire a definire un **modello di sviluppo economico distintivo per l'Europa** in cui la cessione di capacità produttiva al bene capitale da parte del fattore lavoro avvenga in modo controllato. Sul piano commerciale, questo approccio potrebbe costituire una **strategia per differenziare i beni e i servizi europei** che, garantendo di incorporare quote regolamentate di lavoro umano, potrebbero presto venire a costituire un'alternativa attrattiva in un mercato mondiale inondato da prodotti meno regolamentati.

In conclusione, in un mondo in cui i principali attori globali subiscono pressione da parte di influenti attori privati, a nostro avviso l'UE rappresenta l'autorità sovrana più adatta a disegnare **una relazione tra gli esseri umani e la nuova "machina sapiens"** che sia etica, sicura e in linea con i valori sociali, costruendo le premesse per favorire un aumento del livello di benessere sociale tramite il progresso tecnologico.

ANNEX 1 - GLOSSARIO

- **Intelligenza artificiale:** termine ombrello che comprende tutti i metodi che consentono ai *computer* di simulare il comportamento umano, tramite diverse tecniche (regole simboliche e/o modelli numerici).
- **Machine learning:** tutte le tecniche usate da un'intelligenza artificiale con l'obiettivo di **emulare** l'apprendimento umano.
- **Symbolic AI:** nota anche come "IA basata sulla logica" o "IA basata su regole", è una forma di IA che utilizza la rappresentazione esplicita di conoscenza e ragionamento logico attraverso sistemi di regole e simboli definiti. A differenza delle reti neurali, questa IA non è in grado di dedurre le regole a partire da grandi collezioni di dati.
- **Deep Learning:** un tipo di *machine learning* che parte dall'imitazione del funzionamento del cervello umano (biomimetica), tramite la costruzione di reti neurali artificiali stratificate. Le reti neurali sono semplificazioni astratte, ovvero rappresentazioni matematiche, dei processi neurali biologici. A differenza dell'IA simbolica, le reti neurali sono capaci di apprendimento automatico.
- **Modelli pre-trained:** modelli di IA 'addestrati' tramite grandi collezioni di dati.
- **Modelli transformer:** modelli che possiedono abilità di *deep learning* grazie alle tecniche matematiche dette "dell'attenzione" (realizzati a partire dal 2017). Sono anche detti **Foundation models**, in riferimento al cambio di paradigma che hanno determinato.
- **Attention:** risultato dell'impiego di tecniche matematiche che emulano l'attenzione cognitiva umana attribuendo alle informazioni in ingresso un sistema di pesi flessibili, che possono variare a seconda del contesto.
- **LLMs, ovvero large language models:** modelli linguistici costruiti a partire da enormi collezioni di testi e con l'impiego dei modelli *transformer*, la macchina apprende il significato delle parole stabilendo delle relazioni statistiche tra ciascuna parola e le altre.
- **GPT, ovvero Generative Pre-trained Transformer:** un LLM realizzato da OpenAI, è alla base del *chatbot* noto come chatGPT.
- **Generative AI o GenAI:** tutte le tecniche di IA volte alla generazione di diversi tipi di contenuti inediti (non solo testuali).
- **Capacità emergenti:** abilità che divengono presenti nei modelli con l'aumentare della scala, ovvero con l'aumento dei parametri dei modelli e della dimensione delle collezioni di dati con cui sono addestrati; può non essere possibile prevedere completamente quali abilità saranno acquisite dalla macchina all'aumentare della scala (si veda *annex 4*).
- **Artificial General Intelligence, o AGI:** intelligenza artificiale generale, che esibisce capacità trasversali quali ragionamento, pianificazione e apprendimento a partire dall'esperienza.
- **General purpose AI o GPAI:** è una categoria introdotta dall'AI Act per identificare le IA in grado di svolgere una pluralità di compiti (come i modelli di recente sviluppo). L'acronimo può indicare anche la Global Partnership on Artificial Intelligence o GPAI (Gee-Pay), un'iniziativa internazionale volta a guidare lo sviluppo responsabile dell'IA.

ANNEX 2 - PROBLEMI RELATIVI ALLA DEFINIZIONE E ALLA VALUTAZIONE DELL'IA

Definire in modo univoco l'intelligenza artificiale è un compito complesso e tuttora oggetto di dibattito tra gli esperti del campo per molte ragioni, non ultima la difficoltà di trovare un punto di partenza in **una definizione di "intelligenza"** che sia universale e astratta da contesti e compiti concreti.

Per l'esperto di IA e di gestione di dati, scegliere una definizione implica anche adottare un certo **metodo di misurazione** delle abilità delle macchine oggetto di studio, aspetto che può influenzare profondamente i risultati della ricerca e **la comprensione dei fenomeni** (si veda annex 4 per la questione delle 'imprevedibili' capacità emergenti). Per questo, ad oggi, la letteratura che tratta il tema dell'intelligenza artificiale utilizza una varietà di definizioni, ciascuna delle quali presenta dei limiti tecnici o filosofici.

Fuori dall'ambito della ricerca, le istituzioni di riferimento hanno fornito diverse definizioni. Una di queste è la definizione è stata fornita nel 2018 dall'High level working Group della Commissione Europea, sulla base della quale è stata elaborata nel 2021 la prima versione dell'AI Act:

"I sistemi di intelligenza artificiale (IA) sono sistemi *software* (e possibilmente anche *hardware*) progettati da esseri umani che, **dato un obiettivo complesso**, agiscono nella dimensione fisica o digitale **percependo** l'ambiente¹¹⁰ attraverso l'acquisizione di dati; **interpretando i dati** strutturati o non strutturati raccolti¹¹¹; **ragionando** sulla conoscenza, ovvero elaborando le informazioni, derivate da tali dati; in questo modo, **decidono** quali sono le migliori azioni da intraprendere, per **raggiungere l'obiettivo prefissato**. I sistemi di intelligenza artificiale possono utilizzare regole simboliche o apprendere un modello numerico e possono anche **adattare il loro comportamento** analizzando il modo in cui l'ambiente è influenzato dalle loro azioni precedenti"¹¹².

Si tratta di una definizione molto ampia, che ricomprende tutte le modalità di imitazione del comportamento umano, includendo tecnologie di IA meno avanzate di quelle più recenti, descritte nel capitolo 1 e nei successivi *annex*; tuttavia, essa elenca quattro capacità che forniscono una buona descrizione delle macchine pensanti:

- Capacità di **percepire** l'ambiente, virtuale o reale;
- Capacità di **interpretare** i dati percepiti;
- Capacità di **ragionare** sulla conoscenza acquisita;
- Capacità di **decidere** quali azioni mettere in campo.

¹¹⁰ Questo può essere un ambiente virtuale, quindi composto di dati, oppure un ambiente fisico da cui, tramite diversi tipi di sensori, sono estratte delle informazioni.

¹¹¹ I dati strutturati, in genere dati quantitativi, sono altamente organizzati e facilmente decifrabili dagli algoritmi di apprendimento automatico. I dati strutturati possono essere utilizzati solo per lo scopo per il quale è stata progettata la struttura in cui sono organizzati. I dati non strutturati, tipicamente qualitativi, vengono archiviati nel loro formato nativo (la loro raccolta è perciò più facile e veloce) per essere poi strutturati al momento in cui diventa noto lo scopo specifico per cui si intende utilizzarli.

¹¹² Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe, Brussels, 25.4.2018 COM(2018) 237 final.

Queste capacità permettono all'IA descritta dall'HLWG, dato un obiettivo complesso, di raggiungerlo. L'HLWG non descrive la capacità di adattare il proprio comportamento a nuovi dati, la capacità della macchina di individuare obiettivi intermedi in modo autonomo, né tanto meno la capacità di operare in ambiti diversi.

Nel novembre del 2023, l'OCSE, una delle prime organizzazioni a definire l'IA, ha aggiornato la propria definizione, con importanti ripercussioni sulla discussione avutasi nell'ambito dell'AI Act.

Quello di seguito è il testo della definizione, le parti della precedente definizione che sono state eliminate appaiono barrate, mentre sono in rosso i nuovi inserimenti.

Un sistema di intelligenza artificiale è un sistema basato su macchina che ~~può realizzare, per un dato insieme di~~ obiettivi **espliciti o impliciti definiti dall'uomo**, **inferisce, dagli input che riceve, come generare output quali** previsioni, **contenuti**, raccomandazioni o decisioni ~~influenzando~~ **che possono influenzare** ambienti **reali fisici** o virtuali. **Diversi** sistemi di intelligenza artificiale ~~sono progettati per funzionare con diversi~~ **si differenziano per i** livelli di autonomia **e adattabilità ("adaptiveness" nell'originale, ndr) successiva all'impiego.**

Tra le parti eliminate, si notino:

- il riferimento a "un dato insieme di obiettivi", non essendo necessario prevedere degli obiettivi specifici nella fase di progettazione dei nuovi modelli;
- il riferimento a "obiettivi definiti dall'uomo", essendo la macchina in grado di porsi obiettivi intermedi (e non essendo chiaro se sia in grado di porsi obiettivi in generale, come discusso nel paragrafo 1d);
- l'idea che le IA siano "progettate per funzionare con diversi livelli di autonomia", poiché il livello di autonomia ad oggi non dipendere più unicamente dalla progettazione iniziale.

Tra le parti inserite, si segnalano:

- il riferimento agli "obiettivi impliciti"¹¹³;
- il riferimento all'inferenza "a partire dai dati";
- il riferimento alla capacità di generare 'contenuto';
- il riferimento all' "adattabilità successiva all'impiego", ovvero alla capacità delle macchine di considerare le conseguenze delle azioni svolte, di raccogliere nuovi dati durante l'impiego, e quindi di modificare in automatico il proprio funzionamento: in altre parole, le nuove IA evolvono dopo la fase di sviluppo, anche grazie all'interazione con gli utilizzatori¹¹⁴.

Per stabilire un approccio metodologico a successivi lavori sul tema dell'IA, sarà necessario approfondire ulteriormente il funzionamento della tecnologia, anche affrontando le **questioni irrisolte legate alla**

¹¹³ Nei modelli linguistici di grandi dimensioni gli obiettivi del sistema non sono esplicitamente programmati ma acquisiti: in parte attraverso il processo di apprendimento per imitazione da testo generato dall'uomo, in parte attraverso il processo di apprendimento per rinforzo da *feedback* umano (RLHF), gratuitamente realizzato dagli utenti.

¹¹⁴ "Updates to the OECD's Definition of an AI System Explained", Stuart Russell, Karine Perset, Marko Grobelnik, 29 novembre 2023.

definizione e alla valutazione dell'intelligenza artificiale: ad oggi, infatti, non si conoscono scientificamente i passaggi che, a partire dalla capacità di previsione della stringa di parole componenti una frase, hanno portato la macchina a esibire una sorta di '**conoscenza**' e alcuni elementi di **intelligenza artificiale generale**, come sottoprodotto della mappatura probabilistica delle relazioni tra le parole in un testo.

È infatti opinione di molti esperti che gli attuali **metodi di valutazione degli LLM** creino l'illusione di capacità maggiori di quelle realmente esistenti: la pratica di valutare le abilità delle macchine tramite *test* pensati per gli umani sarebbe errata e produrrebbe **risultati falsati**. Ad esempio, GPT4 è riuscito a superare brillantemente i *test* del QI umano, a ottenere punteggi elevati nei test somministrati agli studenti di MBA negli Stati Uniti, e persino a passare l'esame americano di abilitazione alla professione legale (*bar exam*) e parte dei *test* somministrati agli aspiranti medici. Tuttavia, questi risultati hanno diviso la comunità scientifica: alcuni cognitivisti ritengono che il modello linguistico mostri **evidenti abilità nel ragionamento analogico**¹¹⁵ e **nella generazione di idee innovative**¹¹⁶; mentre altri sostengono che l'antropomorfizzazione influenzi il modo in cui sono valutate le risposte delle macchine e che **i risultati ottenuti dalle macchine si debbano interpretare in modo diverso** dai risultati ottenuti dagli esseri umani nello svolgere il medesimo *test* (si veda l'*annex 4* per una discussione sul metodo di valutazione delle proprietà emergenti).

Le abilità tecniche sono strettamente legate alla questione delle dimensioni dei modelli, che ha **importanti risvolti economici e sociali**, discussi approfonditamente nell'*Annex 5*.

¹¹⁵ Webb, Taylor, Keith J. Holyoak, and Hongjing Lu. "Emergent analogical reasoning in large language models." *Nature Human Behaviour* (2023): 1-16.

¹¹⁶ Secondo un esperimento condotto nell'università di Wharton, GPT-4 avrebbe generato idee imprenditoriali migliori degli studenti, e in modo più veloce (le valutazioni delle idee sono state affidate a panel di consumatori). L'esperimento è descritto in "Ideas are dime a dozen: Large language models for idea generation in innovation." di Girotra, Karan, et al. - SSRN 4526071 (2023).

ANNEX 3 - MODELLI TRANSFORMER E LLM: APPRENDIMENTO E ADATTAMENTO

Un modello “*transformer*” è una rete neurale che **può apprendere il significato di un dato mappando le relazioni che intercorrono tra esso e il suo contesto** (gli altri dati, appartenenti all’insieme con cui è alimentato il modello). Questi modelli sono stati descritti per la prima volta in un celebre *paper* del 2017 realizzato dai ricercatori di Google Brain¹¹⁷, e intitolato “*All you need is attention*”, facendo riferimento a una **innovativa branca di tecniche matematiche**, dette appunto tecniche dell’*attention* o *self-attention*, tramite le quali è possibile **emulare l’attenzione cognitiva umana**¹¹⁸. Il *paper* del 2017 ha mostrato come i modelli “*transformer*” fossero maggiormente efficienti rispetto ad altri modelli di *machine learning*. Essi **sono detti anche “*Foundation models*”**, riprendendo la definizione data in un articolo del 2021 dai ricercatori del centro HAI (Human-centered Artificial Intelligence) dell’università americana di Stanford¹¹⁹. L’articolo sottolinea come questi modelli siano quindi alla base di un **fondamentale cambiamento di paradigma nello sviluppo dell’intelligenza artificiale**: grazie ai *transformer*, infatti, l’approccio delle reti neurali digitali, ha riconquistato terreno rispetto al cosiddetto “**approccio della logica simbolica**”, che proponeva di imitare l’intelligenza umana fornendo alla macchina istruzioni minuziose per riprodurre la conoscenza¹²⁰. Al contrario, l’approccio delle reti neurali propone di riprodurre il funzionamento del cervello umano, considerando la conoscenza come un prodotto, peraltro secondario, di quest’ultimo.

La maggiore efficienza nell’apprendimento dimostrata dai modelli *transformer* li rende particolarmente efficaci nel rilevare i **modi in cui dati distanti di una serie si influenzano e dipendono l’uno dall’altro**: la macchina apprende, quindi, il significato di un dato **mappando le relazioni tra il dato e il suo contesto**. Una parola collocata all’interno di *corpus* di testi è un esempio di dato collocato in un contesto: partendo da grandi *corpora* di testi i modelli *transformer* costruiscono, in fase di addestramento, **una rappresentazione compressa della struttura dei testi**, fondata sulla probabilità che si verifichino certe relazioni tra le parole. Tramite questo processo sono generati i **modelli linguistici di grandi dimensioni (LLM)**, in grado di comprendere il significato delle parole e persino il tono della comunicazione.

¹¹⁷ Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017). Nel *paper*, i modelli *transformer* sono descritti come modelli basati esclusivamente sull’applicazione pratica delle tecniche matematiche dell’“*attention*”: è dimostrato come questi modelli producano risultati superiori per qualità e necessità di un minor tempo di addestramento rispetto a precedenti modelli di Machine Learning.

¹¹⁸ Le tecniche dell’attenzione attribuiscono, tramite un sistema di pesi flessibili, minore o maggiore importanza ai dati in ingresso, a seconda del contesto in cui si trovano, permettendo anche uno spostamento del focus qualora sia resa necessaria dall’evoluzione del contesto del dato.

¹¹⁹ "Reflections on Foundation Models." Rishi Bommasani, Percy Liang, Stanford Institute for Human-Centered AI, 18 ottobre 2021.

¹²⁰ Questo approccio si basa sull’uso di simboli che rappresentano concetti e regole logiche per modellare il ragionamento delle macchine in modo più diretto e interpretabile rispetto alle reti neurali: ciò, tuttavia, richiede che la conoscenza sia strutturata e trasmessa alla macchina, non in grado di individuare relazioni tra i dati in modo autonomo.

note tematiche

In conclusione: l'importanza dei modelli *transformer* risiede nella **capacità di costruire autonomamente i modelli** di relazione tra i dati a partire da vaste collezioni degli stessi e nella conseguente **capacità di modificare il modello** sulla base dell'ingresso di nuovi dati, fornendo alla macchina la flessibilità necessaria ad **adattarsi a un contesto mutevole**.

ANNEX 4 - GENERATIVE AI E AGI (ARTIFICIAL GENERAL INTELLIGENCE): IMPREVEDIBILITÀ DELLA MACCHINA, DIMENSIONI DEI MODELLI E CAPACITÀ EMERGENTI

Come spiegato nell'*annex 3*, l'innovazione dei modelli *transformer* ha permesso all'IA di acquisire la capacità di **costruire autonomamente il modello di relazione tra i dati** a partire da vaste collezioni degli stessi e, conseguentemente, la capacità di adattare il modello di interpretazione dei dati all'ingresso di nuovi dati, e quindi di reagire in modo appropriato e non ripetitivo agli stimoli esteri.

Nel caso degli LLM ciò ha condotto alla capacità di **generare dei contenuti innovativi**, come testi inediti e conversazioni, realizzati dalla macchina servendosi delle regole sulle possibili combinazioni di parole che la macchina stessa ha "dedotto" mappando i testi forniti durante l'addestramento.

Tra gli LLM si colloca **la tecnologia GPT (Generative Pre-training Transformer)** creata nel 2018 nei laboratori di OpenAI¹²¹. Questa tecnologia ha immediatamente attratto **l'interesse dei media e del pubblico** poiché è stata utilizzata a partire dalla fine del 2022 per alimentare un *chatbot* (chatGPT), che, da marzo 2023, è stato reso gratuitamente disponibile al pubblico mondiale, seppure in forma limitata¹²².

Per semplificare, si può immaginare GPT come una versione molto avanzata dei sistemi in grado di prevedere ciò che l'utente si appresta a digitare (*next-word prediction*). Si tratta, infatti, di un **modello linguistico realizzato dalla macchina stessa su basi statistiche**: ovvero estrapolando, da *corpora* di testi di enormi dimensioni, informazioni circa l'uso e le diverse possibili combinazioni di parole, permettendo quindi di calcolare le parole che con maggiore probabilità sono combinate tra loro. Questo tipo di intelligenze artificiali capaci di apprendimento automatico sono dette *pre-trained* proprio perché esse sono state 'addestrate'¹²³ fornendo alla macchina grandi collezioni di dati (nel caso dei modelli

¹²¹ OpenAi è un ente *no profit* finanziato da imprese, tra cui Microsoft, e ha un suo ramo aziendale nella OpenAI Limited Partnership.

¹²² La versione *premium* è stata contestualmente resa disponibile a pagamento (chatGPTplus).

¹²³ L'addestramento della macchina può essere supervisionato, non supervisionato o semi-supervisionato. Queste tre modalità di addestramento si distinguono in base alla costruzione della banca dati fornita alla macchina: nel caso di addestramento supervisionato i dati sono "*labeled*" ovvero sono stati etichettati (*tagged*) da umani, mentre nell'addestramento non supervisionato i dati non sono etichettati ("*unlabeled data*").

Ad esempio, un'etichetta apposta da un operatore umano potrebbe indicare il contenuto di una foto, le parole pronunciate in una registrazione audio, l'azione eseguita in un video, il sentimento generale di un post, o il risultato di una radiografia: molte etichette possono essere ricavate dall'attività spontanea degli utenti in rete (es. pubblicazione di un video con un titolo, di una foto con un commento, inserimento di una descrizione testuale a un grafico o test medico in una cartella elettronica). Tuttavia, le collezioni di dati etichettati sono generalmente protette per ragioni di privacy oppure appartengono alla piattaforma su cui i contenuti sono stati gratuitamente pubblicati dagli utenti: pertanto l'accesso a determinate banche di dati etichettati risulta parziale o costoso.

Questo problema di accesso ai dati *labeled* è aggirato dalla supervisione debole, ovvero dalla tecnica utilizzata nel caso dell'apprendimento semi-supervisionato, che impiega una piccola quantità di dati etichettati in combinazione con una grande quantità di dati non etichettati, a cui viene apposta una etichetta dalla macchina stessa derivando le regole di etichettatura dalla parte minoritaria di *labeled data* ad essa forniti.

linguistici un *corpus* o più *corpora* testuali), che la macchina “*transformer*” utilizza per creare autonomamente dei modelli linguistici sufficientemente completi, detti LLM, ovvero *large language models*.

Per la sua capacità di creare novità, **il modello linguistico GPT è considerato un tipo di “generative AI”**, definizione generale che comprende tutte le tecniche di IA volte alla generazione di diversi tipi di contenuti (di cui i contenuti testuali e le conversazioni interattive con gli utenti sono un esempio)¹²⁴.

Un gruppo di ricerca della Microsoft¹²⁵, dopo aver condotto degli esperimenti con i prototipi di GPT4 realizzati nei laboratori OpenAI da novembre 2022, ha affermato, in un *paper*¹²⁶ pubblicato a marzo 2023, che questa tecnologia mostra **“un’intelligenza più generale rispetto ai precedenti modelli di intelligenza artificiale”**. Secondo gli autori del *paper*, le capacità di GPT4 **si avvicinano alle prestazioni umane** in una varietà di compiti nuovi e difficili (che abbraccerebbero vari ambiti tra cui: matematica, programmazione, medicina, diritto, psicologia). Pertanto, concludono i ricercatori, GPT4 “potrebbe ragionevolmente essere visto come una versione iniziale (ma ancora incompleta) di un sistema di **intelligenza artificiale generale (AGI)**”.

Con questa espressione, i ricercatori della Microsoft si riferiscono a una definizione di IA incentrata su **capacità trasversali** quali ragionamento, pianificazione e apprendimento a partire dall’esperienza. I ricercatori Microsoft, nello stesso *paper*, hanno anche rilevato come GPT4 mostrasse **“sorprendenti capacità emergenti”¹²⁷**.

Si legge nell’*abstract* del *paper*: “Consideriamo un’abilità emergente se essa non è presente in modelli più piccoli ma è presente in modelli più grandi. Pertanto, le abilità emergenti **non possono essere previste** semplicemente a partire dalle prestazioni di modelli più piccoli. L’esistenza di tale emergenza implica che **aumentare ulteriormente la scala potrebbe espandere ancora oltre la gamma di capacità dei modelli linguistici.**”

Ricerche successive si sono soffermate sulle capacità emergenti, giungendo alla conclusione, ancora dibattuta, che **gli osservatori umani non siano stati in grado di prevedere lo sviluppo di certe abilità da**

¹²⁴ Il *chatbot* alimentato dalla tecnologia GPT è in grado di comprendere le domande poste dagli utenti, valutare il tono del loro messaggio e rispondere coerentemente nel contenuto e nella forma.

¹²⁵ Microsoft è uno dei principali investitori di OpenAI.

¹²⁶ Bubeck, Sébastien, et al. "Sparks of artificial general intelligence: Early experiments with gpt-4." arXiv preprint arXiv:2303.12712 (2023).

¹²⁷ Il termine è stato introdotto nel 2022 da uno studio realizzato dalla collaborazione di Google Brain, Deepmind e l’università di Stanford (Wei, Jason, et al. "Emergent abilities of large language models.", 2022). Secondo altre ricerche (Schaeffer, Rylan, Brando Miranda, and Sanmi Koyejo. "Are emergent abilities of Large Language Models a mirage?.", 2023), alcune delle abilità dei modelli linguistici esaminati appaiono come capacità emergenti per l’inadeguatezza degli strumenti di misurazione di tali capacità. Secondo questo modo di vedere le cose, una nuova abilità potrebbe sembrare emersa in modo imprevedibile perché, nonostante l’intelligenza artificiale stesse effettivamente migliorando tale abilità a un ritmo costante, questi miglioramenti non sono stati rilevati dai metodi di misurazione finché la competenza dell’IA non ha raggiunto un certo livello. Queta spiegazione del fenomeno di fatto non esclude quindi l’emergere di capacità nuove con l’aumento della dimensione del modello, ma attribuisce agli osservatori l’incapacità di prevedere lo sviluppo di certe abilità da parte dell’IA, piuttosto che descrivere l’IA come qualcosa di imponderabile.

parte dell'IA a causa dell'impiego di metodi di misurazione delle abilità inadeguati: alcune delle abilità dei modelli linguistici esaminati apparirebbero come capacità emergenti nonostante l'intelligenza artificiale avesse migliorato tale abilità nel tempo e a un ritmo costante, solo perché questi miglioramenti non sono stati rilevati dai metodi di misurazione finché la competenza dell'IA non ha raggiunto un certo livello.

Di fatto, ciò non mette in discussione **il fenomeno dell'emergenza di nuove abilità con l'aumento della dimensione del modello**; piuttosto sposta la questione sull'incapacità attuale da parte degli studiosi di prevedere quali abilità possono emergere, legando tale fallimento previsionale ai sistemi di misurazione delle abilità anziché descrivere l'IA come una tecnologia che è per sua essenza imponderabile¹²⁸.

La questione della dimensione dei modelli ha anche dei **risvolti politicamente rilevanti**: le dimensioni dei modelli probabilistici li rendono costosi sia per la **quantità di dati necessaria ad addestrarli** sia per la **capacità computazionale necessaria** a fare inferenza a partire dai dati. Ciò fa sì che solo imprese di grandissime dimensioni sono in grado di operare in questo settore. Le dimensioni dei modelli determinano, inoltre, **l'impossibilità di ispezionarli** per tutti coloro che non dispongono delle stesse capacità di calcolo delle imprese proprietarie, come nel caso degli enti di ricerca indipendenti, ragione per cui gli stessi *policy maker* devono spesso basare le proprie decisioni sulle informazioni volontariamente condivise dalle aziende da regolamentare¹²⁹.

Al momento presente, esiste una spinta ad aumentare ulteriormente le dimensioni dei modelli per tentare di risolvere i limiti da essi tuttora esibiti. Infatti, se da un lato i modelli probabilistici presentano alcune **capacità sorprendenti emerse proprio all'aumentare della scala**¹³⁰, dall'altro essi possono ancora fallire nello svolgere **compiti estremamente banali** poiché non dispongono del **'senso comune', ovvero della conoscenza tacita**, non facilmente derivabile dalle collezioni di dati, poiché, negli esseri umani, essa è ricavata dall'esperienza, è difficile da esprimere in modo formale, ed è per lo più detenuta in modo inconsapevole¹³¹.

¹²⁸ Schaeffer, Rylan, Brando Miranda, and Sanmi Koyejo. "Are emergent abilities of Large Language Models a mirage?", 2023.

¹²⁹ Il Foundation Model Transparency Index di Stanford tiene traccia della quantità di informazioni divulgate dalle imprese circa loro modelli di intelligenza artificiale. I risultati ottenuti dai principali LLM oggi esistenti non sono incoraggianti: Llama 2 (il modello LLM di Meta) ha ottenuto il punteggio più alto in termini di trasparenza con il 54%, seguito da BloomZ e GPT-4 di OpenAI. Nessun modello linguistico ha ottenuto punteggi elevati in ambiti quali: le fonti dei dati utilizzati per il training, l'impatto sociale e i potenziali *bias* riflessi nei modelli.

¹³⁰ Confronta annex 4 sulle capacità emergenti.

¹³¹ L'espansione delle dimensioni dei modelli è vista da alcuni come una possibile soluzione, poiché l'aumento del numero dei dati potrebbe fornire alla macchina sufficienti informazioni per ricavare da essi la conoscenza tacita detenuta comunemente dagli esseri umani. Altri ricercatori de settore, ritengono che la soluzione vada ricercata altrove, ad esempio nell'efficienza dei modelli. Altri ancora sostengono che lo sviluppo dell'IA può avvenire solo tramite dei 'corpi' robotici con i quali l'IA possa entrare in contatto diretto con il mondo, diversamente da quanto avviene attraverso la mappatura dei dati raccolti dagli umani.

ANNEX 5 - DIMENSIONI DEI MODELLI, COSTI DI SVILUPPO E INVESTIMENTI

Le abilità tecniche sono strettamente legate alla questione delle dimensioni dei modelli, che ha **importanti risvolti economici** (per i costi che ne derivano e che a loro volta determinano la concentrazione del mercato in questo settore) e **politici** (la dimensione determina l'impossibilità di studiare tali modelli per la comunità scientifica nel suo complesso e quindi l'impossibilità per i *policy maker* di conoscerli a fondo per poi regolamentarli).

Esistono diverse soluzioni che mirano a migliorare l'efficienza dei modelli senza aumentarne la dimensione. Tra queste si segnala, ad esempio, la metodologia RAG (Retrieval-Augmented Generation)¹³², che rappresenta un'alternativa allo sviluppo dell'Intelligenza artificiale generale (AGI). Tramite la tecnologia RAG **l'utente potrebbe scegliere la fonte esterna** sulla base della quale il modello LLM dovrebbe integrare le proprie risposte, basate sulla rappresentazione interna delle informazioni generata nella fase di addestramento. Ciò permette di garantire che il modello abbia accesso a informazioni attuali e affidabili; inoltre, l'utilizzo di fonti note agli utilizzatori aumenta la trasparenza del processo.

Tornando agli attuali modelli di grandi dimensioni, si segnala che è molto complesso avere un'idea generale del costo dei dati e della capacità di calcolo necessari all'addestramento dei modelli. Secondo le stime riportate da "The Information", testata online dedicata al settore digitale, il **supercomputer** utilizzato da OpenAI per realizzare GPT è costato¹³³ 1,2 miliardi di dollari: questo **supercomputer** utilizza 10,000 processori grafici (o GPU, Graphic Processor Unit) prodotti da Nvidia. Inoltre, il CEO di OpenAI, Sam Altman, ha dichiarato che per il *training* della quarta e ultima versione di GPT sono stati spesi 100 milioni di dollari per l'acquisto di altri circuiti GPU. OpenAI ha stimato che la quantità di calcolo utilizzata nel *training* delle varie versioni di GPT è **aumentata in modo esponenziale**, raddoppiando ogni pochi mesi (fatte salve le dichiarazioni del CEO Altman circa l'intenzione di realizzare in futuro una crescita delle dimensioni più lenta, anche se comunque positiva).

Un articolo¹³⁴ uscito sulla rivista Forbes a settembre sottolinea come, in generale, tutti i modelli LLM, **crecendo costantemente di dimensioni**, richiedano una **sempre maggiore potenza computazionale** e tempi di addestramento sempre più lunghi. Inoltre, le dimensioni dei modelli **cregono più velocemente delle capacità hardware**: pertanto è richiesto **un numero sempre crescente di processori**. La domanda di processori espressa da OpenAI e dai suoi *competitor* ha condotto a una **carenza mondiale di processori**, facendone aumentare il prezzo (e comportando importanti entrate per Nvidia, l'unico grande produttore di *chip* adeguati). Alla fine dell'anno 2023, i principali attori tecnologici hanno intrapreso la scelta di avviare la **produzione interna di processori specifici per l'IA** per controllare direttamente questo mercato. Alphabet, l'impresa

¹³² Lewis, Patrick, et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." *Advances in Neural Information Processing Systems* 33 (2020): 9459-9474.

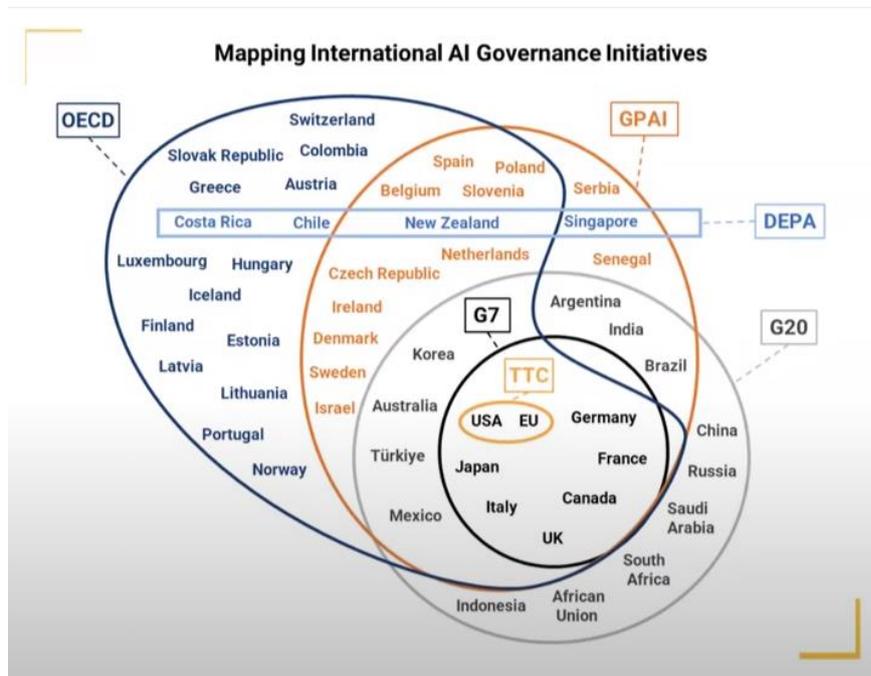
¹³³ Il costo è stato sostenuto da Azure, una divisione di Microsoft, principale finanziatore di OpenAI.

¹³⁴ Smith, Craig S. "What Large Models Cost You – There Is No Free AI Lunch." *Forbes*, 8 settembre 2023.

proprietaria di Google, e Amazon sono stati i primi a muoversi in questa direzione; questo novembre si è unita a loro Microsoft, lanciando due *chip* per IA prodotti internamente e utilizzabili nei *data center*. Google ha annunciato ad agosto l'ultima versione del suo *chip* per IA, il TPUv5e (TPUv5 lite): esso permetterebbe di realizzare modelli più grandi di GPT allo stesso costo affrontato da OpenAI.

Una volta che il modello è stato sviluppato e addestrato, occorre altra capacità computazionale per “fare inferenza”, ovvero per utilizzarlo. Solo nel gennaio 2023, ChatGPT ha utilizzato quasi 30.000 GPU per gestire centinaia di milioni di richieste giornaliere degli utenti (Altman ha dichiarato a novembre che gli utenti di ChatGPT sarebbero 100 milioni a settimana). Tuttavia, da allora **gli utenti di chatGPT** sono aumentati e con essi anche i costi di utilizzo del modello: secondo alcune stime, ad aprile 2023, per far funzionare ChatGPT servivano già tra 700 mila dollari e 1 milione di dollari al giorno. Ciò ha un costo anche in termini di **consumo di energia**: Sajjad Moazeni, professore all'Università di Washington, sostiene, come riportato dalla rivista Forbes a settembre 2023, che queste richieste potrebbero consumare circa **1 GWh ogni giorno**, l'equivalente del consumo energetico giornaliero di circa 33.000 famiglie statunitensi. Naturalmente, tutto ciò ha anche dei costi nascosti, ovvero esternalizzati sulla collettività, in termini di impatto sugli ecosistemi: Microsoft ha rivelato che il consumo di **acqua necessaria a raffreddare i server** è aumentato del 34% tra il 2021 e il 2022, raggiungendo oltre 6 miliardi di litri.

ANNEX 6 - PRINCIPALI INIZIATIVE DI GOVERNANCE GLOBALI, REGIONALI E NAZIONALI



Mappatura delle iniziative di governance.

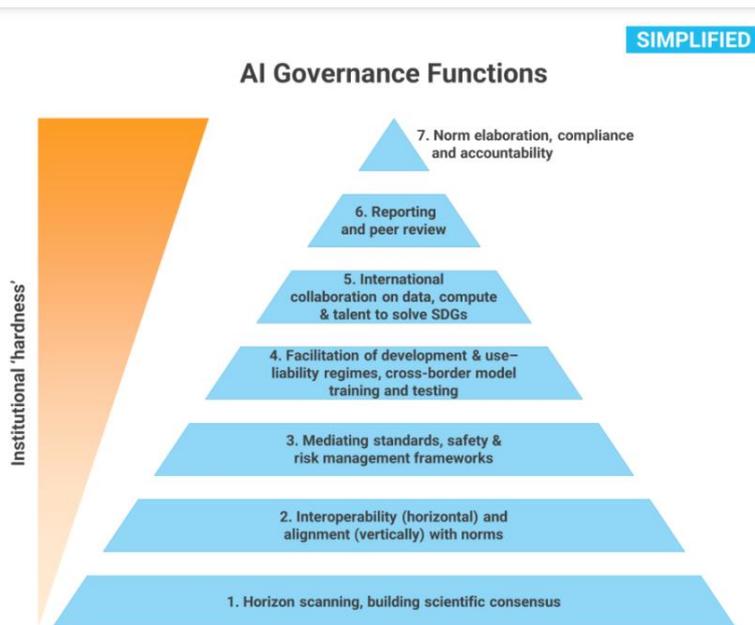
Slide tratta da una presentazione di Andrea Renda (2023) presso l'Istituto Affari Internazionali.

Molteplici iniziative di *governance* dell'IA sono state avviate a livello nazionale e sovranazionale: alcune di queste, come nel caso delle iniziative in ambito ONU, sono nate nel 2023, mentre altre sono frutto di lavori iniziati già diversi anni fa, per cui le novità tecnologiche e commerciali del 2023 hanno reso necessario un aggiornamento (es. OCSE), oltre che un'accelerazione.

Nel novembre del 2023, il Segretario Generale delle Nazioni Unite ha istituito un organo consultivo per l'IA, il **Comitato sull'Intelligenza Artificiale delle Nazioni Unite** (UN Advisory Body). Alla fine dell'anno il Comitato ha rilasciato un rapporto provvisorio intitolato "Governare l'intelligenza artificiale per l'umanità", che sarà rivisto entro la metà del 2024 in seguito ai *feedback* raccolti, per poi essere presentato al Summit of the Future nel settembre 2024.

Il rapporto provvisorio richiede una maggiore omogeneità delle normative nazionali sull'IA e propone di rafforzare la *governance* internazionale dell'IA costruendo un sistema in grado di svolgere **sette funzioni critiche**: i) valutare periodicamente le direzioni future e le implicazioni dell'intelligenza artificiale; ii) rafforzare l'interoperabilità delle iniziative di *governance* nazionali e il loro radicamento nelle norme internazionali attraverso un quadro di *governance* globale dell'intelligenza artificiale approvato in un contesto universale (ONU); iii) sviluppare e armonizzare gli standard, la sicurezza e la gestione del rischio; iv) facilitare lo sviluppo, la diffusione e

l'uso dell'intelligenza artificiale a fini economici e sociali attraverso la cooperazione internazionale; v) promuovere la collaborazione internazionale sullo sviluppo dei talenti, l'accesso alle infrastrutture informatiche, la creazione di diverse raccolte di dati di alta qualità, la condivisione responsabile di modelli *open source* e dei beni pubblici abilitati dall'intelligenza artificiale per gli obiettivi di sviluppo sostenibile (SDG); vi) monitorare i rischi, segnalare incidenti, coordinare la risposta alle emergenze; vii) Elaborare le norme, assicurarne il rispetto e attribuire le responsabilità.



I principi identificati in ambito ONU.

In ambito G7, i leader hanno raggiunto lo scorso ottobre un accordo sui principi guida internazionali riguardanti l'intelligenza artificiale, individuando un **“codice di condotta volontario”** per gli sviluppatori di IA, nell'ambito del “processo di Hiroshima” avviato a maggio. La presidente della Commissione Europea è tra i sottoscrittori dei principi individuati dal G7 e del codice di condotta volontario, i quali integrano, a livello internazionale, le norme vincolanti emanate a livello nazionale e in ambito UE¹³⁵. Il

¹³⁵ I dodici principi rivolti a ‘tutti gli attori’ coinvolti nello sviluppo delle tecnologie IA sono i seguenti: i) adottare misure, durante lo sviluppo di sistemi avanzati di IA e prima, durante e dopo la loro diffusione e immissione sul mercato, atte a identificare, valutare e mitigare i rischi; ii) identificare e mitigare le vulnerabilità e, ove appropriato, gli incidenti e gli usi impropri; iii) segnalare pubblicamente le capacità, i limiti e gli ambiti di applicazione dei sistemi di IA avanzati, gli usi appropriati e inappropriati, per garantire una sufficiente trasparenza, e in tal modo contribuire ad aumentare la responsabilità; iv) lavorare per la condivisione responsabile delle informazioni e la segnalazione degli incidenti tra organizzazioni che sviluppano sistemi di intelligenza artificiale avanzati, coinvolgendo l'industria, i governi, la società civile e il mondo accademico; v) sviluppare, implementare e divulgare politiche di governance dell'IA e di gestione del rischio, tramite un approccio basato sul rischio, includendo politiche sulla privacy e misure di mitigazione dei rischi, in particolare, per le organizzazioni che

documento rilasciato al termine della conferenza mette inoltre in guardia sui rischi per la concorrenza nel settore IA e negli altri settori, in cui potrebbe verificarsi la collusione algoritmica¹³⁶.

Anticipando molte altre organizzazioni internazionali, **l'OCSE ha adottato i suoi principi base sull'IA nel 2019**, contestualmente all'adozione di una definizione per il concetto di intelligenza artificiale che è però stata rivista e aggiornata a novembre del 2023¹³⁷. I principi elencati sono generali e contengono alcuni elementi essenziali per la progettazione delle politiche riguardanti l'IA: i) crescita inclusiva, sviluppo sostenibile e benessere; ii) equità e valori centrati sull'umanità; iii) trasparenza e "spiegabilità"; iv) robustezza, sicurezza e protezione; v) responsabilità.

Tra le iniziative bilaterali, si segnala il **Trade and Technology Council (TTC)**, che dal 2021 vede Unione Europea e Stati Uniti collaborare per coordinare alcuni aspetti della regolamentazione digitale, tra cui la regolamentazione riguardante IA, secondo una *joint roadmap* pubblicata a dicembre 2022. All'inizio di novembre 2023 è stata aperta la consultazione pubblica per raccogliere *feedback* sui 65 termini chiave considerati "essenziali per comprendere l'approccio *risk-based* alle tecnologie IA", la consultazione pubblica è stata conclusa il 24 novembre, ma non è ancora stata aggiornata la tassonomia comune pubblicata a maggio.

Nel frattempo, lo scorso agosto, **la Cina ha proceduto a progettare un proprio sistema di governance dell'IA** basato su 24 linee guida, con compiti distribuiti su a sette diverse agenzie statali: esso impone ai fornitori di sistemi di intelligenza artificiale di registrare i propri servizi e di condurre controlli di sicurezza prima che vengano immessi sul mercato.

Parallelamente ai lavori comuni in seno al TTC, anche Stati Uniti e Unione Europea hanno avviato, già dal 2016 (US) e 2018 (UE), diversi processi interni volti anche a produrre una regolamentazione del settore IA, con tempistiche molto diverse.

Negli **Stati Uniti**, le prime iniziative sono state avviate da Obama nell'ultimo anno in carica, con il principale obiettivo di sostenere lo sviluppo della tecnologia. Più recentemente, il 30 ottobre 2023, il presidente americano Biden ha firmato un ordine esecutivo per fornire una **regolamentazione di base urgente** sull'intelligenza artificiale generativa, in attesa dell'intervento degli organi legislativi. L'ordine esecutivo ha otto obiettivi: i) creare **nuovi standard per la sicurezza**, ii) proteggere la *privacy*, iii) promuovere **l'equità e i diritti civili**, iv) **difendere i consumatori, i pazienti e gli**

sviluppano sistemi di IA avanzati; vi) investire e implementare solidi controlli di sicurezza, includendo la sicurezza fisica, la sicurezza informatica e le tutele contro le minacce interne durante tutto il ciclo di vita dell'IA; vii) sviluppare e implementare meccanismi affidabili di autenticazione e provenienza dei contenuti, ove tecnicamente fattibile, come *watermarking* o altre tecniche per consentire agli utenti di identificare i contenuti generati dall'intelligenza artificiale; viii) dare priorità alla ricerca volta a mitigare i rischi sociali e incentrata sulla sicurezza e definire prioritari gli investimenti nelle misure di mitigazione; ix) dare priorità allo sviluppo di sistemi di intelligenza artificiale avanzati volti ad affrontare le grandi sfide globali, in particolare, ma non solo, la crisi climatica, la salute e l'istruzione; x) promuovere lo sviluppo e, se necessario, l'adozione di standard tecnici internazionali; xi) implementare adeguate misure di immissione dei dati e di protezioni dei dati personali e della proprietà intellettuale; xii) promuovere un uso affidabile e responsabile dei sistemi avanzati di intelligenza artificiale.

¹³⁶ Si veda il paragrafo 2c.

¹³⁷ Si veda l'*Annex 1*.

studenti, v) sostenere i lavoratori, vi) promuovere **l'innovazione e la concorrenza**, vii) far avanzare la *leadership* degli Stati Uniti nelle tecnologie di intelligenza artificiale, e viii) garantire **l'uso responsabile ed efficace** della tecnologia da parte del governo. Inoltre, diverse agenzie governative hanno ricevuto l'incarico di creare *standard* circa l'impiego dell'intelligenza artificiale nella progettazione di materiali biologici (es. il *design* di nuove proteine), o volti a stabilire le *best practice* sull'autenticazione dei contenuti e a creare programmi avanzati in materia di *cybersecurity*.

In Europa, con la pubblicazione del "Piano coordinato sull'intelligenza artificiale" nel dicembre 2018, la Commissione europea ha incoraggiato gli Stati membri a sviluppare **le strategie nazionali per l'intelligenza artificiale** entro la metà del 2019. Le strategie pubblicate nel maggio 2021 differiscono tra loro sotto diversi aspetti¹³⁸.

A livello centrale, al contempo, il 18 giugno 2020 **il Parlamento Europeo ha istituito il Comitato speciale sull'intelligenza artificiale nell'era digitale (AIDA)**, guidato dall'eurodeputato rumeno Dragoș Tudorache, affidandogli il mandato di definire una tabella di marcia a lungo termine per l'UE sull'intelligenza artificiale.

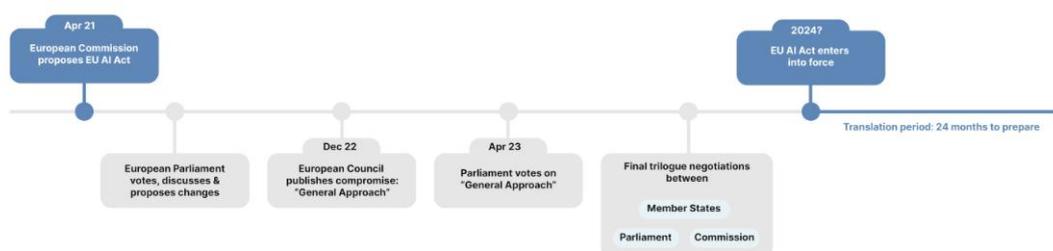
Nel 2021, **una proposta di legge sull'IA (AI Act) è stata presentata dalla Commissione Europea**; quasi due anni prima dell'uscita sul mercato delle applicazioni IA pensate per i consumatori finali, che da fine 2022 hanno attratto l'attenzione dei media. Per maggiori dettagli, si veda l'Annex successivo.

¹³⁸ In questa documentazione, la Francia propone che i dati diventino un bene comune, e la strategia francese prevede di incentivare gli attori economici a condividere i propri dati, individuando nello Stato un mediatore fidato per la gestione di tale condivisione, con il potere, in determinate circostanze, di "imporre la trasparenza su alcuni dati di interesse pubblico". La Spagna propone di creare un "Istituto nazionale di dati" per pianificare e definire la governance sui dati provenienti da diversi livelli di governo. Ampio spazio nelle strategie dei grandi Paesi è dato alla cooperazione nella ricerca: la Germania pianifica la creazione di un *cluster* europeo di innovazione per finanziare progetti di ricerca cooperativa nei successivi cinque anni. Presumibilmente queste strategie saranno aggiornate nei prossimi anni. In Italia a novembre 2023 è stato presentato un Comitato di Coordinamento che contribuirà ad aggiornare la strategia nazionale sull'utilizzo dell'Intelligenza Artificiale, con l'obiettivo di guidare lo sviluppo di questa tecnologia in modo responsabile e inclusivo. Il Comitato sarà presieduto da Gianluigi Greco, professore di informatica all'Università della Calabria e presidente di AlxIA, l'associazione italiana per l'intelligenza artificiale.

ANNEX 7 - NEGOZIAZIONE E CONTENUTI DELL'AI ACT

La regolamentazione contenuta nell'AI Act si caratterizza per un **approccio basato sul rischio**: l'obiettivo principale dello strumento è **prevenire ex ante esiti dannosi**, utilizzando meccanismi simili a quelli dalla legislazione in materia di sicurezza dei prodotti, e **salvaguardare la concorrenza nel settore**, assicurando la trasparenza per garantire la posizione degli **attori a valle** della catena del valore. Partendo da una classificazione delle applicazioni gerarchicamente distinte in base al livello di rischio, essa contiene divieti di utilizzo per quelle applicazioni dell'IA che comporterebbero 'rischi inaccettabili'¹³⁹, restrizioni per le applicazioni considerate 'ad alto rischio', e prevedendo, per le restanti applicazioni, l'obbligo per le imprese di presentare proprie valutazioni dei rischi. Nel corso del 2023, per opera del Parlamento Europeo, sono state equiparate alle applicazioni ad alto rischio anche tutte **le tecnologie di IA realizzate tramite modelli di fondazione**, e più in generale tutte le tecnologie IA 'multifunzionali' (*general purpose AI*): questo intervento si è reso necessario per assicurare che le imprese produttrici di questi modelli garantissero la sicurezza già in **fase di sviluppo**. Tuttavia molte delle disposizioni dell'AI act riguardanti la categoria "ad alto rischio" **non saranno applicabili prima di due anni dall'approvazione ufficiale del testo**. La governance dell'IA in Europa durante il periodo transitorio sarà affidata al cosiddetto **'patto sull'IA (AI Pact)**, un'iniziativa avviata dalla Commissione per raccogliere gli **impegni volontari** da parte dell'industria dell'intelligenza artificiale ed anticipare l'applicabilità del Regolamento, favorendone così l'attuazione tempestiva. Le dichiarazioni di impegno, volte a garantire la conformità con le misure dell'AI Act, saranno raccolte e pubblicate dalla Commissione per dare visibilità alle azioni svolte o pianificate dalle imprese.

Di seguito si esamina più nel dettaglio l'*iter* del regolamento per tentare di spiegare alcuni aspetti della normativa in esso contenuta.

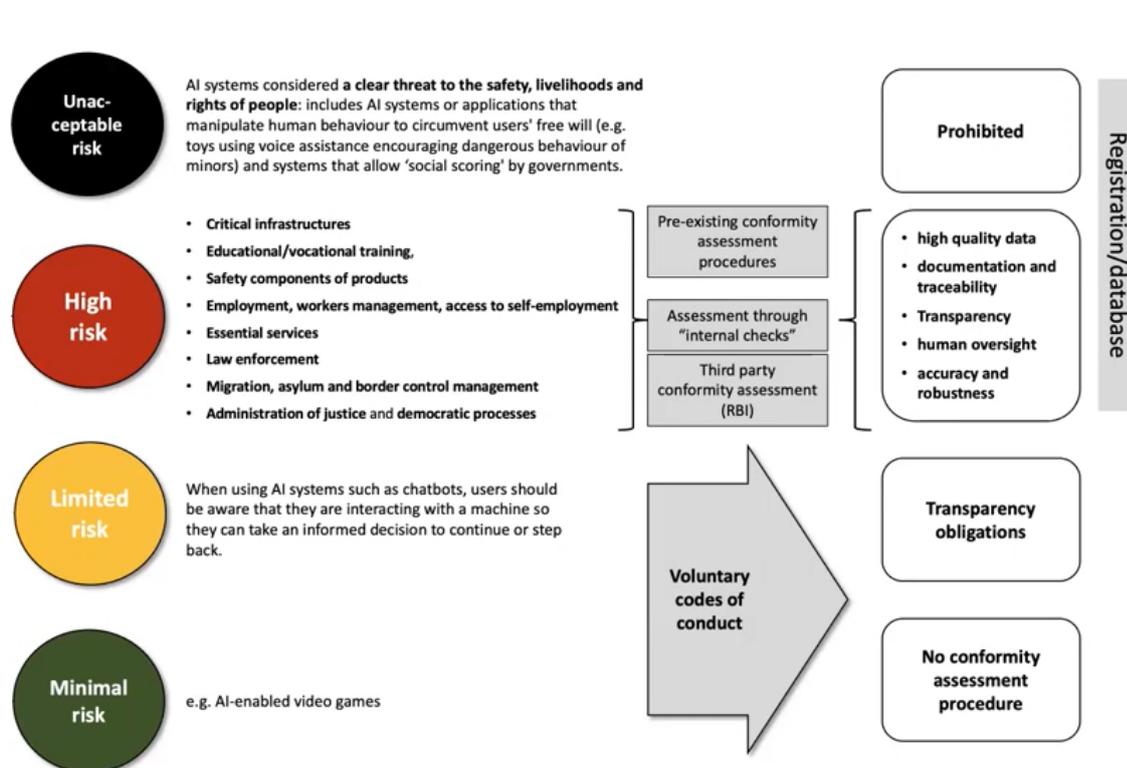


Gli step dell'AI Act.

¹³⁹ Sono vietati: (i) sistemi di intelligenza artificiale che utilizzano tecniche subliminali al di là della consapevolezza della persona per distorcerne materialmente il comportamento; (ii) sistemi di IA che sfruttano le vulnerabilità di un gruppo specifico di persone a causa della loro età, disabilità fisica o mentale; (iii) sistemi di punteggio sociale (social scoring) ingiustificati o sproporzionati utilizzati dalle autorità pubbliche; e (iv) l'uso di sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico ai fini di polizia e applicazione della legge, con alcune eccezioni ottenute dal Consiglio in conclusione del trilatero.

note tematiche

Nel corso del 2023 il testo è stato modificato dal Parlamento al fine di considerare il potenziale dell'**intelligenza artificiale multifunzionale (general purpose AI)**, e regolamentare alcuni aspetti delle nuove tecnologie di IA **generativa**¹⁴⁰. Il principale contributo del Parlamento è stato l'inserimento del paragrafo 28b¹⁴¹, volto ad adeguare l'approccio della proposta della Commissione ai recenti sviluppi della tecnologia: infatti esso, essendo basato sulla valutazione dei rischi delle **applicazioni concrete** dei sistemi di IA, rischiava di **attribuire le responsabilità derivanti dall'impiego dell'IA agli attori a valle** della catena del valore, ovvero alle imprese degli altri settori che impiegano l'IA concretamente (spesso senza avere completa conoscenza del suo funzionamento), anziché attribuire la responsabilità ai giganti della tecnologia che sviluppano tali sistemi (ad oggi, attori extra-europei).



Slide tratta da una presentazione di Andrea Renda (2023) presso l'Istituto Affari Internazionali.

Per ovviare a questo problema e meglio considerare i rischi delle **nuove tecnologie di IA**, il Parlamento ha introdotto due categorie, parallele alla precedente classificazione basata sulla rischiosità delle applicazioni, vale

¹⁴⁰ Ad esempio, sono stati inseriti il divieto di utilizzare il materiale protetto da diritto d'autore nell'addestramento degli LLM o l'obbligo di etichettare come tali i contenuti generati da IA.

¹⁴¹ Per cui si è battuto l'eurodeputato Axel Voss (PPE), seguendo le indicazioni del suo consigliere per la politica digitale, Kai Zenner.

a dire “IA per scopi generali” (*general purpose AI*, o GPAI) e “*foundation models*”. La GPAI è definita come un sistema di intelligenza artificiale destinato a svolgere una **pluralità di compiti e a essere integrato in una pluralità di altri sistemi** di IA, mentre i *foundation models*, nel testo del Parlamento, sono definiti come un sottoinsieme della GPAI¹⁴². Inoltre, il Parlamento scrive nel preambolo del testo che la definizione di intelligenza artificiale utilizzata in UE deve essere “strettamente allineata con il lavoro delle organizzazioni internazionali che lavorano sull’intelligenza artificiale per garantire certezza giuridica, armonizzazione e ampia accettazione”: questo ha assicurato l’allineamento con **la definizione elaborata in ambito OCSE** nel 2019 e con la versione recentemente aggiornata della stessa (2023) proprio per meglio includere le nuove tecnologie.

Nel rispetto dell’articolo 28b, tutte le tecnologie IA che, come i *foundation models*, possono essere usati per una generalità di scopi, ivi comprese applicazioni “ad alto rischio”, debbono essere conformi ai requisiti stabiliti per tale categoria. Pertanto, secondo la bozza proposta dal Parlamento la realizzazione e la fornitura di GPAI e *foundation models* devono avvenire evitando i “rischi ragionevolmente prevedibili”: in questo modo **l’obbligo di identificare i rischi è posto in capo alle imprese produttrici** di tali modelli, essendo impossibile per il legislatore immaginare tutti i loro possibili casi d’uso per classificare i rispettivi livelli di rischio. Gli sviluppatori sono inoltre obbligati a fare scelte progettuali volte a garantire adeguati livelli di prestazione, prevedibilità, interpretabilità, correggibilità, sicurezza e *cybersecurity*. Essi sono altresì tenuti a diversi **obblighi di trasparenza**¹⁴³, tra cui l’obbligo di fornire documentazione tecnica e istruzioni comprensibili alle imprese a valle che fanno uso della tecnologia, nonché ad offrire un rapporto dettagliato sui dati impiegati per l’addestramento. Infine, le imprese sviluppatrici di questi sistemi di IA sono tenute a riferire alla Commissione su eventuali gravi incidenti (occorsi anche in fase di *test*) e sull’efficienza energetica dei modelli, nonché a condurre una valutazione d’impatto sui diritti fondamentali (FRIA). In aggiunta, per i **modelli di fondazione considerati “ad alto impatto”, ovvero con impieghi sistemici**, sono stati introdotti requisiti più stringenti, poiché la loro posizione a monte della catena del valore potrebbe produrre conseguenze imprevedibili a valle.

Come è chiaro, le modifiche introdotte dal Parlamento non hanno sovvertito l’approccio basato sul rischio, ovvero sulla prevenzione del danno, ideato dalla Commissione nel 2021. La bozza per l’AI Act con le modifiche del Parlamento è stata approvata il 14 giugno 2023¹⁴⁴. Si è dunque aperta la fase del ‘trilogo’ tra Commissione, Parlamento e Consiglio. Un momento di tensione in questa fase è stato causato dalla richiesta di Italia, Francia e

¹⁴² Mentre la proposta iniziale della Commissione utilizzava una definizione di intelligenza artificiale ampia e neutrale dal punto di vista tecnologico, i due insiemi di sistemi IA introdotti dal Parlamento sono definiti l’uno in base a una caratteristica (la capacità di servire più scopi), e l’altro in riferimento a uno specifico tipo di tecnologia.

¹⁴³ I cittadini avranno inoltre il diritto di richiedere spiegazioni sulle decisioni che ledono i loro diritti e sono state assunte con l’ausilio di sistemi ad alto rischio.

¹⁴⁴ Il voto sull’AI Act è passato con una maggioranza schiacciante ed è stato annunciato come uno degli sviluppi più importanti a livello mondiale nella regolamentazione dell’IA. La presidente del Parlamento europeo, Roberta Metsola, l’ha definita “una legislazione che senza dubbio definirà lo standard globale per gli anni a venire”.

Germania di ridurre gli obblighi in capo alle imprese sviluppatrici dei modelli di fondazione, inserendo unicamente l'obbligo a adottare **sistemi di autoregolamentazione** (codici di condotta, impegni dei CDA)¹⁴⁵. Tuttavia, a quanto è noto, le disposizioni del Parlamento dedicate a GPAI e modelli di fondazione sono state mantenute nell'accordo politico finale.

Nel corso del trilogio, il Consiglio ha ottenuto che parte dei divieti riguardanti **applicazioni a rischio 'insostenibile'** venissero attenuati. Il compromesso ha riguardato l'utilizzo dell'IA per le politiche di sicurezza, competenza degli Stati: alcuni Paesi¹⁴⁶ hanno infatti chiesto e ottenuto la possibilità di utilizzare l'IA per il **riconoscimento biometrico dei volti negli spazi pubblici** per fini particolari, come la prevenzione di atti terroristici, l'identificazione di persone sospettate di aver compiuto alcuni crimini (ad es. omicidio, rapimento, crimine organizzato), la ricerca di persone vittime di alcuni reati (ad es. rapimento, tratta e sfruttamento sessuale)¹⁴⁷. È stato anche ammesso l'impiego di IA **per il riconoscimento delle espressioni facciali** e delle emozioni nei luoghi di lavoro e studio, esclusivamente per motivi di sicurezza (es. un conducente di mezzi pubblici che rischia di addormentarsi). Anche in questi casi, le forze dell'ordine nazionali dovranno rendere conto alle autorità di competenza europee circa l'impiego dell'IA. Altre esenzioni ai divieti previsti dalla bozza del PE riguardano la possibilità che le forze dell'ordine utilizzino, in caso di emergenza, strumenti di IA ad alto rischio prima che questi abbiano superato la valutazione di conformità.

L'8 dicembre scorso è stato raggiunto un accordo provvisorio tra Commissione, Parlamento e Consiglio. Questo testo è stato approvato all'unanimità dal Consiglio dell'UE il 2 febbraio 2024, mentre dovrà passare a marzo o a aprile in Parlamento per l'approvazione finale, diventando legge a maggio, con la pubblicazione in Gazzetta Ufficiale UE. Mentre alcune disposizioni del regolamento si applicheranno a partire dall'adozione, altre (in particolare riguardanti i sistemi di IA ad alto rischio) saranno applicabili **solo alla fine di un periodo transitorio di due anni** (che intercorre tra entrata in vigore e data di applicabilità della normativa).

Le violazioni alle norme contenute nell'AI Act sono punite con **sanzioni variabili** (le più gravi prevedono sanzioni pari a 35 milioni di euro o al 7% del fatturato globale annuo; le violazioni meno gravi prevedono sanzioni di 7,5 milioni di euro ovvero dell'1,5% del fatturato annuale). Tuttavia, si applicano massimali proporzionali per le piccole e medie imprese e le *start-up*¹⁴⁸.

La governance del settore prevede anche la nascita di **nuovi organi**. Un Ufficio per l'IA è stato istituito presso la Commissione Europea, con il compito di supervisionare i modelli di intelligenza artificiale più avanzati,

¹⁴⁵ Il documento condiviso dai tre Paesi afferma che l'UE necessita di "quadro normativo che promuova l'innovazione e la concorrenza, in modo che gli attori europei possano emergere e portare la nostra voce e i nostri valori nella corsa globale dell'intelligenza artificiale".

¹⁴⁶ Francia, Ungheria, Italia.

¹⁴⁷ Il compromesso è stato criticato da diverse associazioni per i diritti umani, tra cui Amnesty International, che ha sottolineato anche come l'UE non abbia scelto di vietare l'esportazione all'estero delle tecnologie vietate in Europa.

¹⁴⁸ È stato notato che questo potrebbe incentivare le imprese più grandi ad affidare la progettazione dei sistemi di IA più rischiosi alle loro controllate più piccole.

promuovere gli *standard* e il rispetto della regolamentazione. Sarà invece composto da **rappresentanti degli Stati**, il Comitato per l'IA (AI Board) che fungerà da piattaforma di coordinamento e da organo consultivo per la Commissione. Il coinvolgimento della società civile nel processo di *governance* dell'IA sarà assicurato dal Forum consultivo degli *stakeholder* (industria, società civile e mondo accademico).

La maggioranza dei commentatori non ha ancora espresso opinioni approfondite sull'accordo politico raggiunto l'8 dicembre, in attesa della pubblicazione del testo **ufficiale** in Gazzetta. Tuttavia, alcune osservazioni espresse precedentemente all'accordo restano tuttora attuali.

Il direttore del CEPS, Andrea Renda, che siede in diversi comitati del G7 e del TTC dedicati alla *governance* dell'IA, pur esprimendo un generale apprezzamento per il contenuto della proposta del Parlamento, ha sottolineato come i lunghi tempi richiesti dal lavoro svolto dal Parlamento per aggiornare la proposta della Commissione abbiano messo a rischio la preminenza dell'UE nella regolamentazione del settore e la sua capacità di influenzare altre legislazioni tramite il *Brussels effect*; ciò, peraltro, senza poter realizzare appieno l'obiettivo di formulare una **legislazione adattiva**, per tenersi a passo con una tecnologia in veloce evoluzione.

Una critica riportata sul sito del *think tank* Bruegel¹⁴⁹ riguarda, invece, la natura dell'approccio europeo, incentrato sulla prevenzione del danno in base alle diverse categorie di rischio, e quindi inadatto a garantire una *governance* sufficiente per i *foundation model*, soprattutto in quanto muove dal presupposto che ciascuna applicazione di intelligenza artificiale possa essere assegnata a una **specificata categoria di rischio in base all'uso a cui è destinata**¹⁵⁰. Tale critica, simile a quella espressa più volte da Kai Zenner, consigliere per le politiche digitali dell'eurodeputato Axel Voss, *rapporteur* del comitato AIDA, appare tuttavia superata proprio grazie all'introduzione dell'articolo 28b nel testo, che equipara la tecnologia dei *foundation models* a qualsiasi applicazione dell'IA considerata ad alto rischio.

Più in generale, potrebbe rivelarsi miope l'intenzione di regolamentare l'IA con il solo obiettivo di prevenire i rischi, senza inquadrare tale regolamentazione in un **progetto di sviluppo economico** che miri ad integrare l'esistenza dell'IA nel sistema economico e sociale europeo, salvaguardando le caratteristiche essenziali di quest'ultimo, in considerazione delle significative implicazioni sociali e geopolitiche dell'intelligenza artificiale.

¹⁴⁹ Marcus, J. Scott. "Adapting the European Union AI Act to Deal with Generative Artificial Intelligence." 19 July 2023.

¹⁵⁰ Questa struttura riflette in gran parte la tradizionale legislazione europea sulla responsabilità per danno da prodotti difettosi, in cui un prodotto ha uno scopo unico e ben definito.

RIFERIMENTI BIBLIOGRAFICI

Publicazioni scientifiche, studi e rapporti (ordine alfabetico)

Acemoglu, Daron, and Pascual Restrepo. "Artificial intelligence, automation, and work." *The economics of artificial intelligence: An agenda*, University of Chicago Press, 2018. 197-236.

Acemoglu, Daron. "The Simple Macroeconomics of AI", National Bureau of Economic Research (2024).

Baldwin, Richard, and Toshihiro Okubo. "Are software automation and teleworker substitutes? Preliminary evidence from Japan." *The World Economy* (2023).

Benhamou, Salima, et al. "Intelligence artificielle et travail." Paris: France Stratégie. Benhamou S. Janin L (2018).

Beckett, Charlie, and Mira Yaseen. "Generating Change. A global survey of what news organisations are doing with artificial intelligence." Polis, London School of Economics and Political Science (2023).

Berkeley, Edmund. "Giant brain: or machine that think." (1949).

Boucher, Philip Nicholas. "What if AI Regulation Promoted Innovation?" European Parliamentary Research Service, Scientific Foresight Unit (STOA), PE 729.515, aprile 2022.

Brown, Tom, et al. "Language models are few-shot learners." *Advances in neural information processing systems* 33: 1877-1901 (2020).

Bubeck, Sébastien, et al. "Sparks of artificial general intelligence: Early experiments with gpt-4." arXiv preprint arXiv:2303.12712 (2023).

Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond. "Generative AI at work." No. w31161. National Bureau of Economic Research, 2023.

Cagnin, C., Muench, S., Scapolo, F., Störmer, E., Vesnic-Alujevic, L. Shaping and securing the EU's open strategic autonomy by 2040 and beyond. Publications Office of the European Union, Luxembourg, 2021.

Carugati, Christophe. "Competition in generative artificial intelligence foundation models.", Bruegel Working Paper, 18 giugno 2023.

Carugati, Christophe. "The Competitive Relationship Between Cloud Computing and Generative AI.", Bruegel Working Paper, 11 dicembre 2023.

Cazzaniga and others. "Gen-AI: Artificial Intelligence and the Future of Work." IMF Staff Discussion Note SDN2024/001, International Monetary Fund, Washington, DC, 2024

Communication from the Commission on boosting startups and innovation in trustworthy artificial intelligence, 24.1.2024, COM(2024) 28 final.

Dell'Acqua, Fabrizio, et al. "Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality." Harvard Business School Technology & Operations Mgt. Unit Working Paper 24-013 (2023).

Eloundou, Tyna, et al. "Gpts are gpts: An early look at the labor market impact potential of large language models." arXiv preprint arXiv:2303.10130 (2023).

Formosa, Paul, et al. "Medical AI and human dignity: Contrasting perceptions of human and artificially intelligent (AI) decision making in diagnostic and medical resource allocation contexts." *Computers in Human Behavior* 133 (2022): 107296.

Franke, Ulrike. "Artificial intelligence diplomacy - Artificial Intelligence Governance as a New European Union External Policy Tool." AIDA, European Parliament, June (2021).

Gans, Joshua S. "Artificial intelligence adoption in a competitive market." *Economica* 90.358 (2023)

Girotra, Karan, et al. "Ideas are dimes a dozen: Large language models for idea generation in innovation." Available at SSRN 4526071 (2023).

Gmyrek, P., Berg, J., Bescond, D. "Generative AI and jobs: A global analysis of potential effects on job quantity and quality." ILO Working Paper 96. Geneva: International Labour Office, 2023.

Hendrycks, Dan, Mantas Mazeika, and Thomas Woodside. "An Overview of Catastrophic AI Risks." arXiv preprint arXiv:2306.12001 (2023).

Lewis, Patrick, et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." *Advances in Neural Information Processing Systems* 33 (2020): 9459-9474.

Matthews, David, et al. "Efficient automatic design of robots." arXiv preprint arXiv:2306.03263 (2023).

Merchant, Amil, et al. "Scaling deep learning for materials discovery." *Nature* 624.7990 (2023): 80-85.

OECD, *OECD Employment Outlook 2023*, Paris: OECD publishing, 2023.

OECD Digital Economy Papers, no. 352, "AI Language Models: Technological, Socio-Economic, and Policy Considerations.", OECD publishing 2023.

Samoili, Sofia, et al. "AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence.", AI Watch, JRC Technical Reports, (2020).

Schaeffer, Rylan, Brando Miranda, and Sanmi Koyejo. "Are emergent abilities of Large Language Models a mirage?", 2023.

Schmidt, Albrecht. "The End of Serendipity: Will Artificial Intelligence Remove Chance and Choice in Everyday Life?." CHItaly 2021: 14th Biannual Conference of the Italian SIGCHI Chapter. 2021.

Valle-Cruz, David, Vanessa Fernandez-Cortez, and J. Ramon Gil-Garcia. "From E-budgeting to smart budgeting: Exploring the potential of artificial intelligence in government decision-making for resource allocation." *Government Information Quarterly* 39.2 (2022).

Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).

Wang, Hanchen, et al. "Scientific discovery in the age of artificial intelligence." *Nature* 620.7972 (2023): 47-60.

Webb, Taylor, Keith J. Holyoak, and Hongjing Lu. "Emergent analogical reasoning in large language models." *Nature Human Behaviour* (2023): 1-16.

Wei, Jason, et al. "Emergent abilities of large language models." arXiv preprint arXiv:2206.07682 (2022).

Whittaker, Joe, et al. "Recommender systems and the amplification of extremist content." (2021).

Wu, Haitao, et al. "A survey on teaching workplace skills to construction robots." *Expert Systems with Applications* 205 (2022): 117658.

Articoli di giornale e blog post in ordine di data di pubblicazione

"Le frontiere dell'umano: i dilemmi dell'algoritmica", Paolo Benanti, intervistato da Giacomo Bottos, Eleonora Desiata, Pandora Rivista, 21 luglio 2021.

"Reflections on Foundation Models", Rishi Bommasani, Percy Liang, Stanford Institute for Human-Centered AI, 18 ottobre 2021.

"L'IA mette a nudo quasi tutte le proteine note, oltre 200 milioni", Ansa, 29 luglio 2022.

"Artificial intelligence: how to get the most from the labour-productivity boost", Georgios Petropoulos, Bruegel, 9 marzo 2023.

"The potentially large effects of artificial intelligence on economic growth (Briggs/Kodhani)", Jan Hatzius, Goldman Sachs, 26 marzo 2023.

"Workforce Ecosystems and AI", David Kiron, Elizabeth J. Altman, and Christoph Riedl, Brookings, 13 April 2023.

"Large, creative AI models will transform lives and labour markets", *The Economist*, 22 aprile 2023.

"Machines of mind: the case of an AI-powered productivity boom", Martin Neil Bailley et al., Brookings, 10 maggio 2023.

"Why neural net pioneer Geoffrey Hinton is sounding the alarm on AI", Sara Brown, Ideas made to matter, blog della MIT Management Sloan School, 23 maggio 2023.

"AI, Geopolitics, Regulation, and Digital Innovation – What is actually going on right now?", Keegan McBride, Oxford Internet Institute, University of Oxford, 13 giugno 2023.

"Even the scientists who build AI can't tell you how it works: We built it, we trained it, but we don't know what it's doing", Noam Hassenfeld, vox.com, 15 luglio 2023.

"L'intelligenza artificiale al lavoro", Luca De Biase e Stefano Scarpetta, Nòva, *il Sole 24 Ore*, 16 luglio 2023.

"Innovative computational approach helps design proteins for cancer treatment", Max Planck Institute, 28 agosto 2023.

note tematiche

"What Large Models Cost You – There Is No Free AI Lunch", Craig Smith, Forbes, 8 settembre 2023.

"Exclusive: ChatGPT-owner OpenAI is exploring making its own AI chips", Anna Tong, Max A. Cherney, Christopher Bing, and Stephen Nellis, Reuters, 6 ottobre 2023.

"Gartner identifies the top 10 strategic technology trends for 2024", Meghan DeLisi, Catherine Howley, gartner.com, 16 ottobre 2023.

"Baidu says its AI is in the same league as GPT-4", Michelle Toh, CNN, 17 ottobre 2023.

"ChatGPT parent OpenAI seeks \$86bn valuation", Financial Times, 19 ottobre 2023.

"Why the Godfather of A.I. Fears What He's Built", Joshua Rothman, The New Yorker, 13 novembre 2023.

"Updates to the OECD's Definition of an AI System Explained", Stuart Russell, Karine Perset, Marko Grobelnik, 29 novembre 2023.

"Intelligenza artificiale, anche l'Europa accende un faro sull'investimento di Microsoft in OpenAI", Filippo Santelli, 9 gennaio 2024.

note tematiche

Ministero dell'Economia e delle Finanze
Dipartimento del Tesoro
Direzione I: Analisi e Ricerca Economico Finanziaria

Indirizzo:
Via XX Settembre, 97
00187 - Roma

Siti Web:
www.mef.gov.it
www.dt.mef.gov.it/it/

e-mail:
dt.segreteria.direzione1@tesoro.it

Telefono:
+39 06 47614202
+39 06 47614197

Fax:
+39 06 47821886

© Copyright:

2024, Anna Maria Catenacci, Gabriele Velpi.

Il documento può essere scaricato dal sito web www.dt.mef.gov.it/it/ e utilizzato liberamente citando la fonte e l'autore.

Comitato di redazione: Riccardo Barbieri Hermitte, Mauro Marè, Libero Monteforte, Francesco Nucci.

Coordinamento organizzativo: Cristina Gambini