

**CHATGPT, L'EVOLUZIONE DELL'INTELLIGENZA  
ARTIFICIALE E LA SUA APPLICAZIONE NEL  
CAMPO GIURIDICO  
SPUNTI PER UN DIBATTITO**

FRANCO BENASSI

*l'uomo, con tutti i suoi errori e le inesattezze,  
è irrequieto, in continua tensione e ricerca,  
ma grazie a queste spinte è capace  
di creare arte, di produrre idee, di amare*

SOMMARIO: 1. Premessa; 2. L'"intelligenza artificiale nella giustizia; 3. Applicazioni dell'AI in campo giuridico; 3.1 Il sistema di intelligenza artificiale "IBM Watson"; 3.2 Il servizio di intelligenza artificiale "Google Cloud Natural Language"; 4. La novità di ChatGPT-4, con un cenno al settore legale; 5. I sistemi di machine learning per la redazione di sentenze e la valutazione di decisioni giudiziarie; 6. Il Progetto "Predicting Judicial Decisions of the European Court of Human Rights" sulle decisioni della Corte Europea dei Diritti dell'Uomo; 7. L'algoritmo sviluppato dall'Università di Pennsylvania sulle sentenze della Corte Suprema degli Stati Uniti; 8. Le esperienze della Repubblica Popolare Cinese nell'uso dell'intelligenza artificiale in campo giudiziario; 9. Il problema etico nel rapporto uomo-macchina; 10. Linee guida etiche per un'AI affidabile; 11. Conclusione: i timori ancestrali sono prematuri.

## 1. Premessa

Il tema dell'intelligenza artificiale (AI<sup>1</sup>) è diventato in questi ultimi mesi di forte attualità, attirando l'attenzione del grande pubblico informato dei progressi compiuti da sistemi di machine learning che si prestano a dialogare con gli utenti utilizzando un linguaggio molto simile a quello umano. L'AI sta, inoltre, rivoluzionando il modo in cui le aziende e le organizzazioni gestiscono i loro processi e prendono decisioni, aprendo nuove opportunità nei rapporti con i loro clienti ed il pubblico in generale.

L'attenzione su questi temi si è concentrata sulla piattaforma OpenAI, la quale consente a qualsiasi utente di intrattenere un dialogo con un "chatbot" (termine utilizzato per definire i robot che interagiscono con gli utenti attraverso messaggi di testo), di porre domande su argomenti di attualità, di letteratura ed anche su temi di notevole complessità scientifica, come matematica, fisica e medicina. La piattaforma offre pure assistenza concreta nella risoluzione di problemi e nella scrittura o correzione di testi, del codice di programmazione dei programmi per elaboratori nei diversi linguaggi utilizzati dagli sviluppatori.<sup>2</sup>

OpenAI ha inoltre messo a disposizione strumenti che consentono di interagire con programmi e applicazioni creati dagli utenti per interrogare direttamente il chatbot GPT. Detti strumenti, denominati Application Programming Interface (API)<sup>3</sup>, offrono funzionalità molto differenti tra loro, quali, ad

---

<sup>1</sup> In lingua inglese, l'acronimo più comunemente usato per "intelligenza artificiale" è "AI" (Artificial Intelligence).

<sup>2</sup> OpenAI, fondata nel 2015, dichiara di avere attivato una piattaforma con lo scopo di sviluppare l'intelligenza artificiale in modo etico e sicuro, e con l'obiettivo di creare "intelligenza artificiale generale", ossia una forma di intelligenza in grado di comprendere e risolvere problemi in molteplici campi. Sempre secondo OpenAI, l'attenzione si concentra sulla creazione di algoritmi di machine learning avanzati, capaci di apprendere autonomamente da grandi quantità di dati e di migliorare le prestazioni nel tempo.

<sup>3</sup> Le API (Application Programming Interface) di GPT (Generative Pre-trained Transformer) sono strumenti che consentono agli sviluppatori di integrare la tecnologia di GPT nei propri prodotti e servizi tramite richieste HTTP, che possono essere inviate da un'applicazione client a un server che ospita l'API.

esempio, la funzione di completamento del testo che consente agli sviluppatori di utilizzare la tecnologia di GPT per generare testo in modo naturale e fluido e di classificarlo in base a diverse categorie. Le API di GPT sono dunque adoperate in una vasta gamma di applicazioni, tra cui assistenti vocali, chatbot, motori di ricerca, software di scrittura automatica e applicazioni di analisi del testo.

Va anche segnalata la recentissima attivazione da parte della nota multinazionale Microsoft Corporation di diverse funzioni di AI all'interno di Bing, il suo motore di ricerca<sup>4</sup>, il cui algoritmo di machine learning utilizza le informazioni sul comportamento degli utenti per migliorare i risultati delle ricerche, introducendo anche la funzione di ricerca visiva che permette di usare l'AI per identificare oggetti, monumenti e luoghi in immagini e fornire informazioni aggiuntive. Microsoft ha infatti annunciato di aver integrato la tecnologia di OpenAI in Bing allo scopo di migliorare l'esperienza di ricerca degli utenti.<sup>5</sup>

## 2. L'intelligenza artificiale nella giustizia

Algoritmi e sistemi di machine learning si prestano ad essere utilizzati anche nell'ambito dell'amministrazione della giustizia e del diritto in generale, in quanto possono cambiare il modo in cui vengono gestiti e persino risolti i casi giudiziari. La loro efficienza è una logica conseguenza della creazione di un sistema informatico dotato della capacità di analizzare, in tempi molto rapidi, notevoli quantità di dati e di elaborarli per dare risposte mediante un linguaggio molto simile a quello umano.<sup>6</sup>

---

<sup>4</sup> Raggiungibile all'indirizzo <https://www.bing.com/>

<sup>5</sup> Benché sia importante sottolineare che le fluttuazioni del mercato azionario dipendono da molteplici fattori e non possono essere attribuite esclusivamente a un singolo evento, è molto probabile che l'accordo con OpenAI abbia avuto un impatto positivo sull'andamento delle azioni della Microsoft Corporation le quali hanno registrato nell'ultimo periodo un considerevole aumento di prezzo.

<sup>6</sup> Sull'argomento si vedano anche S. PESUCCI, *Diritto e intelligenza artificiale: opportunità e dilemmi nell'era della automazione*, in

Naturalmente, l'uso di queste tecnologie nel campo giudiziario desta non poche preoccupazioni riguardo alla possibilità di bias<sup>7</sup> e discriminazioni che possono affliggere gli strumenti informatici di cui stiamo parlando, soprattutto se non vengono attentamente impostati, gestiti e valutati.

L'utilizzo di algoritmi e sistemi di machine learning nell'amministrazione della giustizia può, infatti, portare a un aumento delle disuguaglianze e dell'iniquità in un settore della società dove il controllo costante da parte dell'uomo è già di per sé fonte di aspre discussioni e confronti anche a livello politico.

Un profilo assai delicato è, infatti, costituito dalla base dati fornita alla macchina, la cui formazione potrebbe essere influenzata da dati storici inadeguati o da fattori discriminanti.

Inoltre, l'impiego di algoritmi e sistemi di machine learning potrebbe portare a un indebolimento della responsabilità umana e della trasparenza decisionale in considerazione del fatto che questi strumenti, oltre ad essere molto complessi, possono anche presentarsi come assai poco trasparenti in ragione di fattori tecnici non facilmente comprensibili. Ciò impedirebbe ai soggetti interessati al caso giudiziario ed ai cittadini in generale, nel cui interesse la giustizia viene amministrata, di verificare ed avere adeguata consapevolezza delle modalità con le quali i dati forniti alle macchine vengono scelti e gestiti nonché della motivazione posta a fondamento delle decisioni.

---

*ristrutturazioniaziendali.ilcaso.it*, 6 marzo 2022; D. ROQUE VITOLO, *Insolvenza e Intelligenza Artificiale*, in *ristrutturazioniaziendali.ilcaso.it*, 11 feb 2022.

<sup>7</sup> Nel linguaggio utilizzato nell'informatica, il termine "bias" si riferisce a un errore sistematico che si verifica durante la fase di sviluppo o di utilizzo di un algoritmo, che può portare a risultati distorti o sbagliati. Un bias può essere causato da diversi fattori, come la selezione non rappresentativa dei dati di addestramento, la presenza di dati mancanti o la presenza di informazioni non rilevanti per il processo di decisione. Ad esempio, un algoritmo di selezione del personale che si basa solo sulla conoscenza della lingua inglese può essere affetto da un bias culturale che penalizza i candidati che provengono da paesi dove l'inglese non è la lingua madre, anche se possono essere altamente qualificati per il lavoro.

La parte coinvolta nel singolo caso giudiziario dovrebbe infatti essere in grado, come avviene con le decisioni delle corti, di verificare il percorso motivazionale e, dall'altro lato, qualunque cittadino dovrebbe essere posto in condizioni di comprendere come le macchine vengono "nutrite" dei dati che portano alla decisione.

Per questi motivi, è essenziale che l'impiego di algoritmi e sistemi di machine learning nell'ambito della giustizia sia guidato da principi etici e da standard di trasparenza e responsabilità. Solo a queste condizioni, ammesso che tale obiettivo sia effettivamente raggiungibile, gli strumenti di cui parliamo potranno essere utilizzati in modo efficace ed equo per supportare le decisioni giudiziarie e le scelte amministrative, migliorando così la qualità del servizio offerto ai cittadini.

### **3. Applicazioni dell'AI in campo giuridico**

Un aspetto di rilievo che non può essere trascurato, è quello dell'utilizzo dei sistemi di AI non per amministrare la giustizia, bensì come supporto a disposizione di magistrati, avvocati e di coloro che operano nel settore giuridico per effettuare analisi predittive, identificare rischi e opportunità e generare raccomandazioni sulle possibili soluzioni di casi giudiziari. Questo profilo si presta a dar vita ad un mercato di sicuro rilievo e potrà dunque essere gestito in termini di servizio scientifico e informativo anche da operatori che si occupano di informazione nel campo giuridico.

In questa seconda ipotesi, non dovrebbero porsi particolari problemi di carattere etico, così come già ora la presenza di vari operatori nel campo della documentazione giuridica garantisce un adeguato confronto e un buon livello di affidabilità dei servizi offerti.

A questo proposito, appare evidente come l'utilizzo dell'AI potrebbe svolgere compiti che con i sistemi attuali sarebbero sostanzialmente impossibili, come quello, solo per fare un esempio, di catalogare le decisioni di giurisprudenza in funzione della probabilità di successo di una determinata domanda.

Invece di limitarsi a catalogare le decisioni in base alle tradizionali ripartizioni per materia o per argomenti trattati, l'AI potrebbe, infatti, essere addestrata per analizzare i fatti e le prove nella disponibilità dalle parti al fine di valutare le probabilità di esito favorevole o sfavorevole di una determinata pretesa.

Questa tecnica si rivelerebbe utile non solo per i professionisti del diritto, i quali avrebbero una maggiore comprensione della probabilità di successo delle questioni cui prestano assistenza, ma anche per gli stessi clienti per avere una visione più chiara dei loro casi.

Gli strumenti in parole si prestano ad essere utilizzati anche dagli studiosi del diritto per effettuare ricerche e organizzare il materiale disponibile sugli argomenti oggetto di speculazione ed approfondimento. Ed a ben vedere, sono già fruibili diverse applicazioni di intelligenza artificiale, come i motori di ricerca legali, che si avvalgono di algoritmi di machine learning per fornire risultati sempre più raffinati e pertinenti. Dette applicazioni possono aiutare a creare sommari automatizzati per analizzare testi giuridici in modo da identificare parole chiave e tendenze, semplificando così la ricerca e l'analisi di grandi quantità di documenti.

Per condurre l'analisi di testi normativi e della letteratura scientifica sono già disponibili alcune piattaforme che offrono servizi di analisi testuale basati su tecniche di machine learning, quali IBM Watson<sup>8</sup> e Google Cloud Natural Language<sup>9</sup>. Questi servizi consentono di analizzare grandi quantità di testo e di estrarre informazioni rilevanti, come ad esempio i concetti chiave, le entità menzionate e le relazioni tra di esse.

---

<sup>8</sup> <https://www.ibm.com/it-it/cloud/watson-studio>

<sup>9</sup> <https://cloud.google.com/natural-language?hl=it>

### 3.1 Il sistema di intelligenza artificiale "IBM Watson"

IBM Watson è un sistema di intelligenza artificiale che integra tecniche di elaborazione del linguaggio naturale basato su evidenze<sup>10</sup>.

Il funzionamento di questo sistema dipende da una serie di moduli software, ognuno dei quali ha una *specifica funzione e competenza* quali, come ad esempio:

- il modulo di *analisi del linguaggio naturale*, che permette di comprendere e analizzare il significato del testo in lingua naturale e di estrarre informazioni rilevanti;
- il modulo di *apprendimento automatico*, che consente al sistema di apprendere continuamente da nuovi dati e di migliorare le proprie prestazioni;
- il modulo di *ragionamento* fondato su evidenze, per elaborare ipotesi e valutare la loro plausibilità in base alle evidenze disponibili;
- il modulo di *elaborazione del linguaggio naturale*, che genera risposte in linguaggio naturale alle domande poste dagli utenti.

Si tratta, in sostanza, di un sistema che utilizza algoritmi avanzati di apprendimento automatico e di analisi del linguaggio per comprendere il significato del testo, elaborare le informazioni e generare risposte adeguate alle domande poste dall'utente.

Ma ciò che appare più rilevante e che lo rende sicuramente **interessante per un uso professionale** è il fatto che il sistema è stato progettato per essere *altamente flessibile e*

---

<sup>10</sup> Il ragionamento basato su evidenze è un approccio utilizzato nei sistemi di intelligenza artificiale per trarre conclusioni logiche sulla base delle informazioni disponibili, spesso in combinazione con altre tecnologie come l'elaborazione del linguaggio naturale e l'apprendimento automatico. V. JUDEA PEARL, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, September 1988, <https://dl.acm.org/doi/10.5555/534975>; SHAFER, G. (1976), *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press, <http://www.glennshafer.com/books/amte.html>.

*personalizzabile*, in modo da poter essere adattato a diverse esigenze e applicazioni in diversi settori e ambiti, tra cui:

- *sanità*: aiuta a diagnosticare e individuare le cure più appropriate per i pazienti attraverso l'analisi di grandi quantità di dati medici, scientifici e di ricerca.
- *finanza*: fornisce analisi e raccomandazioni per la gestione del rischio, la valutazione dei portafogli di investimento, la rilevazione delle frodi e l'ottimizzazione delle decisioni di business.
- *retail*: aiuta i negozi a personalizzare l'esperienza d'acquisto dei clienti, a migliorare la supply chain e a identificare le tendenze di mercato.
- *sicurezza*: supporta la sorveglianza delle attività sospette, l'individuazione di minacce e la prevenzione dei crimini.
- *educazione*: aiuta a creare piattaforme di apprendimento personalizzate, a identificare i punti deboli degli studenti e a fornire materiali di studio specifici.
- *trasporti*: supporta la gestione e l'ottimizzazione dei flussi di traffico, la prevenzione degli incidenti stradali e la pianificazione dei viaggi.

Questi sono solo alcuni esempi delle molteplici applicazioni dell'IBM Watson in continua espansione.

### **3.2 Il servizio di intelligenza artificiale "Google Cloud Natural Language"**

E' un servizio di analisi del linguaggio naturale, anch'esso basato sull'intelligenza artificiale, il quale esamina il testo inserito dall'utente, lo elabora e fornisce informazioni come l'identificazione della lingua, l'analisi del sentiment, la categorizzazione del testo in base alle entità e alle categorie grammaticali, l'estrazione delle parole chiave e la sintesi di frasi chiave<sup>11</sup>.

---

<sup>11</sup> Per utilizzare il servizio, è necessario accedere alla piattaforma di Google Cloud e registrarsi per ottenere un API key. Una volta ottenuta la chiave, è

Il funzionamento di Google Cloud Natural Language è fondato su algoritmi che permettono al sistema di analizzare il testo inserito dall'utente e di estrarre informazioni come relative a persone, luoghi, organizzazioni, eventi, etc. e le relazioni tra tali entità intercorrono.

Un esempio concreto di come si può utilizzare l'API di Google Cloud Natural Language<sup>12</sup> per la comprensione di **documenti legali** è **l'analisi di un contratto aziendale** allo scopo di identificare i termini e le clausole principali. In questa modalità, essa, come il sistema IBM Watson, può aiutare ad estrarre

---

possibile utilizzare l'API di Google Cloud Natural Language per analizzare il testo e ottenere le informazioni desiderate.

<sup>12</sup> Ricordiamo che, in linguaggio informatico, le API (Application Programming Interface), come in minima parte già accennato in una nota che precede, sono insiemi di regole, protocolli e strumenti che consentono a software diversi di comunicare e interagire tra loro in modo efficiente e sicuro. Le API definiscono come un'applicazione o un servizio può richiedere funzionalità o dati da un'altra applicazione o servizio, e come quest'ultimo risponde alle richieste. Esse sono importanti in quanto semplificano lo sviluppo di software e permettono agli sviluppatori di integrare facilmente funzionalità e servizi di terze parti nelle loro applicazioni. Alcuni vantaggi delle API includono: 1) modularità: le API consentono agli sviluppatori di suddividere il software in componenti modulari e riutilizzabili, il che rende più semplice la manutenzione e l'aggiornamento del software; 2) risparmio di tempo e risorse: utilizzando le API per integrare funzionalità e servizi esistenti, gli sviluppatori possono risparmiare tempo e risorse che altrimenti avrebbero dovuto dedicare allo sviluppo di funzionalità simili da zero; 3) interoperabilità: le API facilitano l'interazione tra diverse applicazioni e piattaforme, consentendo la condivisione di dati e funzionalità in modo più agevole; 4) standardizzazione: le API possono stabilire standard e best practice per l'interazione tra software, migliorando la compatibilità e riducendo il rischio di errori o malfunzionamenti. Per facilitare la comprensione di questo strumento molto diffuso, val la pena di accennare ai seguenti tipi di API: a) API web: che utilizzano protocolli basati su Internet, come HTTP, per permettere la comunicazione tra applicazioni e servizi online, le API RESTful (REpresentational State Transfer) sono un esempio comune di API web; b) API di sistema operativo: queste API forniscono accesso alle funzionalità e ai servizi offerti dai sistemi operativi, come la gestione dei file o la comunicazione tra processi; c) API di librerie e framework: le quali permettono agli sviluppatori di utilizzare funzionalità e componenti offerti da librerie e framework software, semplificando lo sviluppo di applicazioni. Esse, in sostanza, svolgono un ruolo cruciale nello sviluppo di software moderno e nell'integrazione di servizi e applicazioni in un'ampia gamma di settori.

informazioni chiave dal testo, come le parti coinvolte, le condizioni di pagamento, le responsabilità, le restrizioni e le esclusioni. La finalità è quella di fornire agli utenti una comprensione più rapida e dettagliata del contenuto del contratto, prima di dedicarsi ad una più attenta lettura del testo integrale.

Il sistema è inoltre in grado di valutare il **sentiment del testo**, ovvero se esso è positivo, neutrale o negativo, e di identificare le categorie grammaticali dei termini ivi presenti (sostantivi, verbi, aggettivi, etc.). Questa funzionalità si pone l'obiettivo di aiutare a valutare l'atteggiamento emotivo delle parti, come la soddisfazione o la frustrazione, e quindi di prevedere eventuali problemi o conflitti futuri.

L'API di Google Cloud Natural Language può dunque essere adoperata per semplificare l'analisi e la comprensione dei documenti legali, migliorando la precisione e l'efficienza del lavoro degli operatori del diritto.

#### **4. La novità di ChatGPT-4, con un cenno al settore legale**

Come si è molto spesso verificato con riferimento ai fenomeni che hanno avuto una enorme quanto rapida diffusione sulla rete Internet, ChatGPT ha guadagnato una notevole popolarità e riconoscimento anche grazie alla *disponibilità gratuita* di alcune delle sue tecnologie, come GPT-3 e, in seguito, GPT-4. Questa strategia ha, infatti, permesso a un'ampia gamma di utenti, tra i quali sviluppatori e ricercatori, di sperimentare le funzionalità offerte dai suoi modelli di machine learning.

Ma ciò che ha suscitato notevole stupore sono i dati assolutamente **straordinari** di espansione degli utenti che hanno fruito di questa applicazione. Secondo uno studio di UBS, si stima, infatti, che ChatGPT abbia raggiunto i **100 milioni di utenti attivi mensili a soli due mesi dal lancio, diventando**

## **così l'applicazione consumer<sup>13</sup> con la crescita più rapida nella storia.<sup>14 15</sup>**

E' tuttavia importante sottolineare che il successo di questa piattaforma non è da ricondurre esclusivamente all'accessibilità gratuita delle sue tecnologie, ma va attribuita soprattutto alla notevole spinta innovativa applicata ai sistemi di machine learning.

GPT-4, l'ultima versione dell'architettura GPT sviluppata da OpenAI, pur essendo basato su precedenti release di intelligenza artificiale, presenta, invero, migliorie e caratteristiche che lo distinguono dai suoi predecessori e che possono essere in estrema sintesi così riassunte:

- *Capacità di apprendimento*: GPT-4 è addestrato su un insieme di dati più vasto e diversificato rispetto alle versioni precedenti, migliorando la sua comprensione del linguaggio naturale e la sua conoscenza generale.
- *Modello più grande*: GPT-4 dispone di un numero molto più elevato di parametri rispetto alle versioni precedenti, il che

---

<sup>13</sup> Un'applicazione consumer, o applicazione per consumatori, è un software o un servizio creato per essere utilizzato direttamente dai clienti o dagli utenti finali, piuttosto che dagli sviluppatori o dalle aziende. Si tratta di applicazioni progettate per soddisfare le esigenze dei consumatori in termini di *funzionalità, facilità d'uso e accessibilità*. Le applicazioni consumer sono spesso progettate con *un'interfaccia utente intuitiva e attraente*, in modo che gli utenti possano utilizzare l'applicazione senza difficoltà e senza conoscenze tecniche avanzate.

<sup>14</sup> Krystal Hu, *ChatGPT stabilisce il record per la base di utenti in più rapida crescita nota dell'analista*, <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>

<sup>15</sup> Mettere a disposizione le proprie tecnologie gratuitamente o attraverso piani di accesso più accessibili ha diversi vantaggi per un'organizzazione come OpenAI: a) visibilità e riconoscimento: la disponibilità gratuita delle tecnologie di OpenAI aiuta a far conoscere l'organizzazione e le sue innovazioni a un pubblico più ampio; b) adozione e innovazione: consentendo agli sviluppatori di utilizzare i modelli GPT gratuitamente o a costi accessibili, OpenAI incoraggia l'adozione delle sue tecnologie e stimola l'innovazione in una vasta gamma di settori e applicazioni; 3) feedback e miglioramento: con un numero maggiore di utenti che utilizzano i modelli GPT, OpenAI riceve preziosi feedback e informazioni sulle prestazioni e l'utilizzo dei modelli in vari contesti che possono essere utilizzati per apportare miglioramenti e ottimizzazioni future.

lo pone in grado di comprendere contesti più ampi, generare risposte più coerenti e fornire soluzioni più accurate.

○ *Miglioramenti nella generazione del testo*: è progettato per generare risposte più pertinenti, accurate e coerenti grazie ai progressi nella sua capacità di analizzare il contesto delle domande, ridurre le risposte evasive e offrire informazioni utili.

○ *Capacità di ragionamento*: è in grado di eseguire ragionamenti più complessi, risolvendo problemi che richiedono una comprensione più profonda del contesto e delle informazioni fornite.

○ *Fine-tuning*: GPT-4 **può essere ulteriormente addestrato su specifici domini o applicazioni**, migliorando le sue prestazioni e rendendolo più adatto a specifiche esigenze.

○ *Interazione multi-lingua*: supporta un numero maggiore di lingue rispetto alle versioni precedenti.

GPT-4 è dunque sistema di intelligenza artificiale più avanzato e sofisticato, capace di dare soluzioni più precise e utili, nonché di interagire con gli utenti in una gamma più ampia di contesti e lingue<sup>16</sup> e ciò anche grazie alle *tensor processing*

---

<sup>16</sup> OpenAI dichiara di aver fatto progressi significativi nel campo dell'intelligenza artificiale attraverso diversi fattori chiave quali: *Ricerca avanzata*: OpenAI è un'organizzazione di ricerca leader nel campo dell'IA e continua a condurre ricerche all'avanguardia. Attraverso la pubblicazione di numerosi articoli di ricerca e la partecipazione a conferenze accademiche, OpenAI è in grado di rimanere aggiornata sulle ultime scoperte e sviluppi nel campo. *Collaborazione*: OpenAI collabora con altre organizzazioni di ricerca, università e aziende per condividere conoscenze, risorse e competenze. Questo consente di attingere a una vasta gamma di esperienze e idee per migliorare le proprie tecnologie. *Raccolta di dati*: OpenAI ha accesso a grandi quantità di dati testuali provenienti da varie fonti, come il web, libri, articoli scientifici e altro ancora. Questo consente a OpenAI di addestrare i propri modelli su un insieme di dati vasto e diversificato, migliorandone la comprensione del linguaggio naturale e la conoscenza generale. *Potenza di calcolo*: OpenAI utilizza infrastrutture di calcolo di alto livello per addestrare i propri modelli, come unità di elaborazione grafica (GPU) e tensor processing units (TPU). Queste risorse di calcolo consentono a OpenAI di addestrare modelli di apprendimento profondo su larga scala in modo efficiente. *Innovazione nell'apprendimento profondo*: OpenAI si basa su innovazioni nell'apprendimento profondo e nelle reti neurali, che hanno portato a notevoli miglioramenti nella capacità dei modelli di comprensione del linguaggio naturale e di generazione del testo.

*units* (TPU), un tipo di acceleratore hardware per l'apprendimento profondo sviluppato da Google.<sup>17</sup>

Per quanto in questa sede può interessare, particolare attenzione, merita il processo di addestramento di GPT-4 *su specifici domini o applicazioni* denominato *fine-tuning* o "raffinamento". In sostanza, GPT-4, una volta che sia stato addestrato su un vasto insieme di dati generali, **può essere personalizzato per specifici compiti o settori, migliorando così le sue prestazioni in tali contesti.**

L'addestramento avviene su un insieme di dati più piccolo e specifico, che è strettamente correlato al *dominio*<sup>18</sup> o

---

Focus sulla sicurezza e l'etica: OpenAI si impegna a sviluppare tecnologie di IA sicure ed etiche. L'organizzazione dedica risorse e sforzi per affrontare problemi come la polarizzazione, il pregiudizio e l'uso improprio dell'IA, contribuendo a garantire che le sue tecnologie siano utilizzate in modo responsabile e benefico.

<sup>17</sup> Le Tensor Processing Units (TPU) sono un tipo di acceleratore hardware per l'apprendimento profondo sviluppato da Google. Le TPU sono progettate specificamente per eseguire calcoli su tensori, che sono strutture di dati multidimensionali utilizzate nelle reti neurali e in altri algoritmi di apprendimento automatico. Le TPU offrono un'elaborazione ad alte prestazioni e a basso consumo energetico per i calcoli matematici necessari per addestrare e utilizzare modelli di apprendimento profondo. Sono state ottimizzate per eseguire operazioni su tensori in modo efficiente, come la moltiplicazione di matrici, che è un'operazione fondamentale nelle reti neurali. Le TPU si distinguono da altre unità di elaborazione, come le unità di elaborazione grafica (GPU) e le unità di elaborazione centrale (CPU), in quanto sono progettate specificamente per i **calcoli legati all'apprendimento profondo**. Mentre le GPU sono state inizialmente sviluppate per l'elaborazione grafica e le CPU per un'ampia gamma di operazioni, le TPU sono state progettate esclusivamente per l'apprendimento automatico, offrendo vantaggi in termini di velocità e efficienza energetica. Esse sono utilizzate in diversi prodotti e servizi di Google, come Google Cloud, Google Photos e Google Translate, per migliorare le prestazioni e la velocità di elaborazione delle reti neurali; sono inoltre disponibili per gli sviluppatori e i ricercatori tramite la piattaforma Google Cloud, consentendo loro di sfruttare la potenza delle TPU per addestrare e utilizzare modelli di apprendimento profondo su larga scala.

<sup>18</sup> Nel contesto dell'intelligenza artificiale e dell'apprendimento automatico, un "dominio" si riferisce a un'area specifica di conoscenza o a un settore di applicazione. Ad esempio, il settore medico, legale o finanziario sono tutti domini distinti, ciascuno con le proprie conoscenze specializzate, terminologia e problemi tipici. Quando si parla di addestrare un modello su un dominio specifico, ci si

all'applicazione di interesse, il che consente al modello di apprendere informazioni e relazioni più dettagliate e pertinenti a quel determinato contesto.

Ad esempio, se si desidera servirsi di GPT-4 per rispondere a domande nel campo della medicina, lo si dovrà addestrare ulteriormente su un insieme di dati costituito da testi medici, articoli di ricerca, domande e risposte comuni nel settore sanitario. In questo modo, GPT-4 diventa più competente e preciso nel fornire risposte e soluzioni specifiche al settore medico.

Se si desidera utilizzare GPT-4 per rispondere a domande nel **campo legale**, lo si addestrerà su un insieme di dati specifico per questo settore formato da legislazione, giurisprudenza e articoli di ricerca in ambito giuridico. Una volta addestrato sul dominio così arricchito, il sistema potrà rispondere a domande sulla legislazione e la giurisprudenza, generare riassunti di casi giudiziari, interpretazioni di leggi o di saggi di dottrina.

Naturalmente, il sistema in esame, anche se ben addestrato, **ben difficilmente potrà sostituire un avvocato o un consulente esperto**. Esso può dare informazioni e risposte fondate su dati di addestramento, ma non è in grado di fornire consulenza legale specifica adeguata alle singole situazioni.

## 5. I sistemi di machine learning per la redazione di sentenze e la valutazione di decisioni giudiziarie

Ci sono diverse esperienze che hanno utilizzato i sistemi di machine learning per la redazione di sentenze e la valutazione di

---

riferisce al processo di personalizzazione del modello per comprendere e gestire meglio le informazioni e le sfide associate a quel particolare settore. L'addestramento su un insieme di dati più piccolo e specifico implica l'utilizzo di un set di dati di addestramento che contiene informazioni strettamente correlate al dominio o all'applicazione di interesse. Ad esempio, nel caso del settore legale, l'insieme di dati potrebbe includere documenti legali, legislazione, giurisprudenza e altri materiali correlati. Utilizzando un insieme di dati più piccolo e specifico per il *fine-tuning*, *si addestra il modello a concentrarsi sulle informazioni e le relazioni pertinenti a quel dominio*, migliorando così le sue prestazioni in quel contesto.

decisioni giudiziarie. Un esempio interessante è rappresentato dal progetto "Predicting Judicial Decisions of the European Court of Human Rights" condotto dalla London School of Economics, che ha analizzato 5847 sentenze della Corte Europea dei Diritti dell'Uomo mediante algoritmi di machine learning.

Altra esperienza è quella dei ricercatori dell'Università della Pennsylvania, i quali hanno sviluppato un algoritmo per l'analisi e la classificazione delle sentenze della Corte Suprema degli Stati Uniti, addestrato su un set di 2400 decisioni, con l'obiettivo di prevedere l'esito di future cause.

Un algoritmo ideato per l'amministrazione della giustizia è quello denominato *COMPAS* (Correctional Offender Management Profiling for Alternative Sanctions), ideato al fine di valutare il rischio di recidiva di un individuo condannato. Esso raccoglie le risposte ad una serie di domande poste all'imputato riguardo alla sua storia criminale, alla sua personalità e ad altri fattori che possono influenzare la sua probabilità di recidiva. Le risposte vengono poi analizzate dall'algoritmo, che produce un punteggio di rischio. Sviluppato negli Stati Uniti, il sistema si serve di un questionario di circa 137 domande in ordine a vari aspetti della vita del soggetto, come la sua storia penale, l'impiego, l'istruzione, la vita familiare e sociale, le abitudini di consumo di droghe o alcol, ecc.. In ragione delle risposte date, il sistema assegna un punteggio che indica il grado di pericolosità del soggetto e che dovrebbe essere utilizzato dai giudici per decidere se concedere la libertà condizionata o impostare misure alternative alla detenzione.<sup>19</sup>

E' ovvio che tutto ciò ha suscitato critiche per presunte disuguaglianze razziali, dal momento che alcuni studi hanno evidenziato come il sistema abbia una tendenza a classificare in modo sbilanciato gli imputati di colore come più pericolosi

---

<sup>19</sup> JON KLEINBERG, HIMABINDU LAKKARAJU, JURE LESKOVEC E JENS LUDWIG, *Human Decisions and Machine Predictions*, in *Quarterly Journal of Economics*, 2018.

rispetto a quelli bianchi, anche in presenza di un'identica storia penale e di risposte simili al questionario.<sup>20</sup>

Un altro esempio di algoritmo impiegato nel campo giuridico è il modello predittivo *Luminance*, il quale si serve dell'intelligenza artificiale per analizzare documenti legali e identificare le clausole critiche. Il sistema è stato sviluppato per aiutare gli avvocati e gli operatori del diritto a risparmiare tempo nella revisione dei documenti e a individuare più rapidamente eventuali rischi legali. *Luminance* impara man mano che analizza i documenti in numero crescente e diventa sempre più preciso nel rilevare le criticità.

Anche questa esperienza è stata oggetto di critiche sulla scorta di studi i quali hanno prospettato un effetto discriminatorio nei confronti di minoranze etniche e sociali, a causa della presenza di pregiudizi impliciti nei dati utilizzati per addestrare l'algoritmo.<sup>21</sup>

## **6. Il Progetto "Predicting Judicial Decisions of the European Court of Human Rights" sulle decisioni della Corte Europea dei Diritti dell'Uomo**

Il Progetto "Predicting Judicial Decisions of the European Court of Human Rights" è stato condotto dalla London School of Economics e ha analizzato un campione di 5847 sentenze della Corte Europea dei Diritti dell'Uomo. L'obiettivo del progetto era quello di sviluppare un algoritmo di machine learning in grado di prevedere le decisioni della Corte sulla base

---

<sup>20</sup> MEGAN T. STEVENSON E CHRISTOPHER SLOBOGIN, *Assessing Risk Assessment in Action* in *Stanford Law Review*, 2018, <https://www.readcube.com/articles/10.2139%2Fssrn.3016088>. JULIA ANGWIN, JEFF LARSON, SURYA MATTU AND LAUREN KIRCHNER, *Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks*, in *ProPublica*, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

<sup>21</sup> CRYSTAL S. YANG, *Do Judges 'Like' Ties? An Empirical Analysis of Clustering Effects in Federal Sentencing* in *The Journal of Legal Studies*, 2017.

di un set di variabili selezionate, come il paese di origine delle parti, la sezione della Corte che giudica il caso e la materia.

L'algoritmo si fondava sull'uso di modelli di *regressione logistica*, una tecnica di apprendimento supervisionato che prevede di addestrare il modello su un insieme di dati di input (in questo caso, gli elementi presenti nelle sentenze della Corte) e di output (le decisioni adottate dalla Corte). In particolare, le variabili di *input* (o variabili indipendenti) fornite al sistema si riferiscono alle caratteristiche e alle informazioni *estratte dai testi delle sentenze* della Corte Europea dei Diritti dell'Uomo. Si tratta di elementi come le violazioni contestate, le leggi applicabili, le circostanze del caso e altre informazioni rilevanti. La variabile di *output*, o variabile dipendente, è la stessa *decisione della Corte*, che è binaria, assume cioè solo due possibili valori, che, nel caso di specie, potrebbe essere se una sentenza è stata classificata come "violazione" o "non violazione" dei diritti umani.

Il progetto (pubblicato nel 2016 dalla rivista scientifica "PeerJ Computer Science" con il titolo "Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective"),<sup>22</sup> ha dimostrato che l'algoritmo sviluppato era in grado di prevedere le decisioni della Corte con un'accuratezza del 79%, comprovando così come l'uso di algoritmi di machine learning nell'analisi delle decisioni giudiziarie possa essere effettivamente utile.

Anche in questo caso, è importante sottolineare che il modello di regressione logistica non fornisce una previsione certa sulle decisioni future della Corte, ma piuttosto una stima basata sulle caratteristiche delle sentenze prese in esame. Inoltre, il modello è soggetto ad eventuali errori di classificazione e addestramento, errori, questi, che rappresentano due delle sfide principali nell'uso dei modelli di regressione logistica e di altre tecniche di machine learning. I primi, quelli di classificazione, si verificano quando il modello cataloga erroneamente un'osservazione come

---

<sup>22</sup> <https://peerj.com/articles/cs-93/>

appartenente a una classe diversa da quella a cui appartiene in realtà. Gli errori di addestramento si producono invece quando il modello è istruito su un insieme di dati che non sono rappresentativi della popolazione (il "dominio" di riferimento di cui abbiamo già detto) sul quale si vogliono ottenere previsioni.

Se il set di dati di addestramento è sbilanciato rispetto alle variabili di input o di output, il modello può infatti produrre previsioni non rappresentative della realtà, come nell'ipotesi in cui l'algoritmo di predizione delle decisioni viene addestrato su un insieme di sentenze che non rappresenta la gamma completa di casi e variabili giuridiche.

Inoltre, se i dati di addestramento contengono **pregiudizi o discriminazioni**, il modello può perpetuare questi difetti nelle sue previsioni.

Per mitigare tali problemi, è necessario garantire una corretta raccolta dei dati, che includa una vasta gamma di casi e variabili giuridiche e che sia priva di pregiudizi e discriminazioni ed è, altresì, importante valutare continuamente le prestazioni del modello ed aggiornarlo sulla base di nuovi dati e delle nuove conoscenze acquisite.<sup>23</sup>

## **7. L'algoritmo sviluppato dall'Università di Pennsylvania sulle sentenze della Corte Suprema degli Stati Uniti**

Un team di ricercatori dell'Università della Pennsylvania ha sviluppato un algoritmo per l'analisi e la classificazione delle sentenze della Corte Suprema degli Stati Uniti, il quale è stato

---

<sup>23</sup> Sull'importanza fondamentale dei dati di addestramento nella creazione di modelli di machine learning e sulla necessità di garantire che i dati utilizzati siano accurati e rappresentativi della realtà che si vuole "modellizzare" si vedano: *Introduzione alla creazione del set di dati*, <https://developers.google.com/machine-learning/data-prep/construct/construct-intro?hl=it>; *Dimensioni e qualità di un set di dati*, <https://developers.google.com/machine-learning/data-prep/construct/collect/data-size-quality?hl=it>. CHRISTIAN KÄSTNER, *Model Quality: Measuring Prediction Accuracy*, <https://ckaestne.medium.com/model-quality-measuring-prediction-accuracy-38826216ebcb>.

addestrato su un set di dati costituito da 2400 decisioni con l'obiettivo di prevedere l'esito di future cause.<sup>24</sup>

Esso utilizza una tecnica di machine learning nota come Support vector machine (SVM), la quale ha lo scopo di creare un confine di decisione<sup>25</sup> tra i dati che rappresentano le diverse classi. In questo caso, l'algoritmo è stato addestrato per prevedere l'esito delle cause della Corte Suprema degli Stati Uniti in base alle informazioni contenute nelle sentenze precedentemente emesse.

Come nel caso dell'algoritmo COMPAS, la sua efficacia dipende dalla qualità dei dati utilizzati per addestrarlo.

Inoltre, i ricercatori hanno anche evidenziato il rischio di un *eccessivo affidamento* all'algoritmo, senza che gli esseri umani

---

<sup>24</sup> "Predicting Supreme Court Decisions Mathematically: A Quantitative Analysis of the "Right to Counsel" Cases" (2013) e "Supreme Court Forecasting Project" (2016).

<sup>25</sup> Il termine "confine di decisione" si riferisce a una superficie (o iperpiano) che separa i punti dati appartenenti a diverse classi in uno spazio multidimensionale. Nel contesto degli algoritmi di apprendimento automatico come Support Vector Machine (SVM), il confine di decisione viene utilizzato per classificare nuove osservazioni in base alle loro caratteristiche. L'idea principale dietro al SVM è quella di trovare il confine di decisione ottimale tra le diverse classi che massimizza il margine tra i punti più vicini delle classi. Questi punti vicini sono chiamati "support vectors" (vettori di supporto) e sono i punti dati che influenzano direttamente la posizione e l'orientamento del confine di decisione. Per esemplificare il concetto di confine di decisione, immaginiamo un semplice caso bidimensionale. Supponiamo di avere un grafico a dispersione con due gruppi distinti di punti che rappresentano due classi diverse. Il confine di decisione sarebbe una linea retta che separa i due gruppi di punti in modo che tutti i punti di una classe si trovino da un lato della linea e tutti i punti dell'altra classe si trovino dall'altro lato. In casi più complessi e multidimensionali, il confine di decisione può essere un piano o un'iperpiano che separa i punti dati in uno spazio di caratteristiche di dimensioni superiori. Sull'argomento: VAPNIK, V. N. (1995), *The Nature of Statistical Learning Theory*. Springer, New York, <https://link.springer.com/book/10.1007/978-1-4757-3264-1>; Hastie, T., Tibshirani, R., & Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, <https://www.amazon.it/Elements-Statistical-Learning-Inference-Prediction/dp/0387848576>. CORTES, C., & VAPNIK, V. (1995). "Support-vector networks". *Machine Learning*, 20(3), 273-297, <https://link.springer.com/article/10.1007/bf00994018>.

svolgano una valutazione critica dei risultati ottenuti, con l'ovvia ma opportuna avvertenza che esso non è stato certo stato creato per sostituire i giudici.

La support vector machine (SVM) è una tecnica di machine learning utilizzata per la classificazione e la regressione. L'obiettivo principale dell'algoritmo SVM è quello di creare un confine di decisione tra i dati che rappresentano le diverse classi, confine, anche chiamato iperpiano, definito come la linea che massimizza la distanza tra i dati delle classi diverse più vicine, chiamati support vectors.

L'algoritmo SVM ha dimostrato di essere efficace in numerosi ambiti, come la classificazione di immagini, la rilevazione di frodi, la diagnosi medica e l'analisi dei sentimenti. Possono essere utili alla comprensione delle funzionalità alcuni esempi:

*Classificazione di email:* l'algoritmo è stato utilizzato per creare un sistema di filtraggio di **spam per le email**. Il sistema utilizza un modello addestrato su un insieme di dati di email etichettati come spam o non spam. L'algoritmo è stato in grado di identificare in modo preciso e affidabile le email indesiderate, riducendo la quantità di spam nella casella di posta in arrivo degli utenti.

*Riconoscimento di caratteri:* l'algoritmo SVM è stato utilizzato per sviluppare sistemi di riconoscimento di caratteri per l'elaborazione di **testi scritti a mano**. Il modello addestrato su un insieme di dati di caratteri scritti a mano è stato in grado di riconoscere in modo preciso e affidabile i caratteri scritti a mano, anche quando la scrittura non era del tutto chiara o leggibile.

*Diagnosi medica:* l'algoritmo SVM è stato utilizzato per creare sistemi di supporto alle decisioni per la **diagnosi medica**. Il modello addestrato su un insieme di dati di sintomi e diagnosi medica è stato in grado di identificare in modo preciso e affidabile le patologie dei pazienti.

*Analisi dei sentimenti:* l'algoritmo SVM è stato utilizzato per sviluppare sistemi di analisi dei sentimenti per l'elaborazione di testi e commenti sui social media. Il modello addestrato su un

insieme di dati di commenti etichettati come positivi o negativi è stato in grado di identificare in modo preciso e affidabile i sentimenti espressi nei commenti.

I problemi riscontrati con riferimento alla affidabilità dei dati riguardano:

*Complessità computazionale*: la complessità computazionale dell'SVM aumenta in modo esponenziale con il numero di campioni di addestramento e di dimensioni del problema, il che lo rende meno adatto a problemi di grandi dimensioni.

*Scelta del kernel*: la scelta del kernel può influenzare significativamente l'accuratezza dell'SVM. Tuttavia, non esiste un kernel universale che funzioni bene per tutti i problemi.<sup>26</sup>

*Interpretabilità*: l'SVM produce un modello di classificazione di difficile interpretazione, che non fornisce una spiegazione esplicita del processo decisionale adottato.

*Overfitting*: se l'SVM viene addestrato troppo a lungo o con un insieme di dati di addestramento troppo piccolo, può verificarsi un fenomeno di overfitting, ovvero l'algoritmo impara a classificare perfettamente i dati di addestramento ma fallisce nel generalizzare correttamente ai nuovi dati.

## **8. Le esperienze della Repubblica Popolare Cinese nell'uso dell'intelligenza artificiale in campo giudiziario**

La Repubblica Popolare Cinese ha mostrato un crescente interesse per l'uso dell'intelligenza artificiale nel settore legale. Nel 2017, la Corte Suprema della Cina ha lanciato una piattaforma di big data e intelligenza artificiale chiamata "Intelligent Assistant for Judges" (智能助手), progettata per aiutare i giudici a *individuare rapidamente* le informazioni rilevanti e a prendere decisioni più informate.

---

<sup>26</sup> Il kernel è una funzione utilizzata nelle Support Vector Machines (SVM) e in altre tecniche di apprendimento automatico per trasformare i dati in un'altra forma o in uno spazio di dimensione superiore, al fine di facilitare la separazione delle classi e la creazione di un confine di decisione; v. nota sul confine di decisione.

Alcune città e province cinesi hanno anche intrapreso iniziative locali per sperimentare l'uso dell'AI nella redazione delle sentenze. Ad esempio, a Shanghai è stato lanciato un programma pilota che utilizza l'intelligenza artificiale per aiutare a preparare bozze di sentenze, in casi civili e amministrativi, mediante algoritmi di apprendimento automatico e tecniche di elaborazione del linguaggio naturale in grado di generare una bozza di sentenza che i giudici possono poi rivedere e modificare.

Nel 2017, il Tribunale di Internet di Hangzhou ha annunciato di aver utilizzato l'intelligenza artificiale per emettere la sua prima sentenza in un caso di violazione del copyright. Nello stesso anno, il tribunale di Shanghai ha iniziato a testare un sistema di intelligenza artificiale per la redazione di sentenze in casi di insolvenza aziendale.<sup>27</sup>

Più di recente, nel 2020, il Tribunale Popolare Intermedio di Guangzhou ha annunciato l'adozione di un sistema per la redazione di sentenze **in cause di lavoro**.

Anche queste esperienze hanno sollevato preoccupazioni riguardo alla trasparenza e all'imparzialità delle soluzioni consigliate dagli algoritmi generando il timore che possano servire per limitare la libertà di espressione e reprimere le opinioni dissidenti.

L'implementazione di un sistema di intelligenza artificiale per la gestione della giustizia nella Repubblica popolare cinese ha subito un **forte accelerazione** a seguito delle linee guida emanate dalla Corte Suprema cinese per la creazione di un **sistema di supporto nel settore giudiziario entro il 2025**.<sup>28</sup> il documento che contiene le *linee guida* è intitolato "Opinioni

---

<sup>27</sup> Sull'argomento dell'organizzazione della giustizia cinese con l'AI si vedano i contributi di D. TODARO, *Efficienza e controllo: l'uso della tecnologia nei tribunali cinesi*, <https://iris.unive.it/handle/10278/3736010>; IVAN CARDILLO, *Disciplina dell'intelligenza artificiale e intelligentizzazione della giustizia in Cina*, in *BioLaw Journal – Rivista di BioDiritto*, n. 3/2022, <https://teseo.unitn.it/biolaw/article/view/2389/2310>

<sup>28</sup> <http://lawinfochina.com/display.aspx?lib=law&id=40187>

sulla regolamentazione e il rafforzamento dell'applicazione dell'intelligenza artificiale in ambito giudiziario" dove l'obiettivo perseguito è quello di migliorare i servizi legali, rafforzare i tribunali e fornire un livello più elevato di giustizia giuridica digitale garantendo la sicurezza e la legalità dell'applicazione di intelligenza artificiale. Le linee guida chiariscono che il sistema non sostituirà i giudici nella decisione dei casi e afferma che i risultati derivanti dall'utilizzo dell'intelligenza artificiale potranno costituire riferimenti supplementari nella supervisione e nella gestione giudiziaria.<sup>29</sup>

Il documento chiede inoltre ai tribunali cinesi di istituire **entro il 2030** un'applicazione per l'utilizzo dell'AI per migliorare l'apparato regolamentare e legislativo e garantire che la tecnologia possa supportare l'intero processo giudiziario.

Il tribunale Supremo ha affermato che il rafforzamento dell'applicazione dell'AI in campo giudiziario non solo aiuterà i giudici a ridurre l'onere di **trattare questioni banali** per migliorare la loro efficienza lavorativa, ma darà anche al pubblico un accesso più facile alle risorse legali per risolvere le loro controversie in modo più efficace. Pare inoltre che più di 90.000 istituti di mediazione e oltre 350.000 operatori si siano collegati ad una piattaforma progettata dalla Corte Suprema per aiutare i litiganti a risolvere le loro controversie direttamente attraverso sistemi di collegamento sulla rete Internet.<sup>30</sup>

## 9. Il problema etico nel rapporto uomo-macchina

Sulla base delle notizie che abbiamo raccolto e delle quali abbiamo parlato nei paragrafi che precedono, è possibile

---

<sup>29</sup><https://www.notizie.ai/entro-il-2025-in-cina-un-sistema-ai-assistera-le-decisioni-dei-giudici/>

<sup>30</sup><https://www.chinadaily.com.cn/a/202212/09/WS6392fa3ba31057c47eba3a3f.html>. V. anche: *La Cina sperimenta la "Robot – justice" per la lotta al crimine*, <https://www.irpa.eu/la-cina-sperimenta-la-robot-justice-per-la-lotta-al-crimine/>; STEPHEN CHEN, *Chinese scientists develop AI 'prosecutor' that can press its own charges*, <https://www.scmp.com/news/china/science/article/3160997/chinese-scientists-develop-ai-prosecutor-can-press-its-own>

constatare come il ruolo dell'uomo non sia solo quello di raccogliere dati per "nutrire" il sistema dell'algoritmo di intelligenza artificiale, ma anche quello di valutare criticamente sia le modalità di raccolta e di consegna dei dati sia i risultati e le decisioni proposte dall'algoritmo, con l'avvertenza che il termine "decisioni" è assolutamente inadeguato spettando le stesse sempre e comunque all'uomo.

L'interpretazione dei dati e la contestualizzazione delle informazioni sono compiti che richiedono il giudizio umano, il quale può essere determinato e influenzato da **fattori che l'intelligenza artificiale non è in grado di considerare.**

Gary Kasparov, ex campione del mondo di scacchi, ha sottolineato l'importanza della collaborazione tra l'uomo e l'AI, affermando che "l'AI e l'uomo possono lavorare insieme per creare *cose incredibili*".<sup>31</sup> Il noto giocatore ha svolto un ruolo attivo nel promuovere la comprensione dell'AI e la diffusione del concetto che la collaborazione tra uomo e intelligenza artificiale può portare a risultati migliori rispetto alla competizione o alla contrapposizione ideologica.

Kasparov ha sperimentato l'AI in prima persona nel 1997, quando ha perso una partita di scacchi contro il computer Deep Blue di IBM. In seguito a questa esperienza, ha affermato che l'AI può essere utilizzata in modo efficace solo se l'uomo è coinvolto nel processo decisionale e decisione finale è presa dall'essere umano.

Kasparov ha anche sottolineato l'importanza della formazione delle persone per il corretto utilizzo dell'AI e della comprensione delle sue limitazioni e potenzialità, sostenendo che l'AI non deve essere considerata una minaccia all'occupazione umana, ma piuttosto un'opportunità per migliorare la produttività e la qualità del lavoro. A suo parere, i sistemi di cui stiamo parlando possono aiutare gli esseri umani a prendere decisioni migliori, ma devono essere applicati in modo responsabile, adottando un approccio di "centaur chess",

---

<sup>31</sup> Fonte: Forbes, 2017.

in cui un umano gioca insieme ad un sistema artificiale per ottenere la massima efficacia nella partita. Questo approccio combinato potrebbe essere applicato anche in altri ambiti, come la medicina, il diritto e la finanza, per ottenere risultati migliori e più affidabili.

Max Tegmark, fisico e docente al MIT, ha affermato che "l'AI deve essere progettata e controllata dall'uomo e che deve prendere decisioni etiche".<sup>32</sup> Max Tegmark è un fisico svedese che insegna al Massachusetts Institute of Technology (MIT) e che è noto anche per le sue attività di divulgazione scientifica. Egli sostiene che la *rapida evoluzione dell'AI rappresenta una sfida* per l'umanità, poiché può portare a importanti cambiamenti nella società e nella vita delle persone.

In particolare, Tegmark afferma che è importante garantire che l'AI non diventi troppo potente e che non prenda decisioni che vadano contro gli interessi umani. Egli ritiene che essa debba essere guidata da una serie di **principi etici che siano riconosciuti a livello globale**, al fine di garantire che sia applicata in modo responsabile e sostenibile.

Inoltre, Tegmark sostiene che l'AI deve essere sviluppata in modo tale da essere **trasparente e comprensibile** e che gli algoritmi di machine learning e le tecniche di intelligenza artificiale devono essere aperti alla verifica e alla revisione da parte degli esperti e della società nel suo complesso.

Secondo Max Tegmark, ci sono diversi **principi etici** che dovrebbero guidare la progettazione e il controllo dell'AI: la *sicurezza*, ovvero l'idea che l'AI dovrebbe essere progettata in modo da minimizzare i rischi per la società e per l'ambiente; la *trasparenza*, l'idea che le decisioni prese dall'AI dovrebbero essere chiare e comprensibili per gli esseri umani; *l'equità e la non discriminazione* delle persone in base alla loro razza, etnia,

---

<sup>32</sup> Fonte: BigThink, 2017.

genere o altre caratteristiche personali; *il rispetto della privacy e la dignità umana*.<sup>33</sup>

Sul rapporto tra il fattore umano e i sistemi di intelligenza computazionale Stuart Russell ha proposto il concetto di "friendly AI" come approccio etico per un loro sviluppo responsabile. Secondo Russell, l'AI dovrebbe essere progettata in modo tale che le sue azioni siano coerenti con gli obiettivi e i valori dell'umanità, in altre parole, essa dovrebbe essere in grado di capire i *desideri* umani e di agire di conseguenza, *senza causare danni o interferire con i nostri valori fondamentali*.

Per raggiungere questo obiettivo, Russell propone l'utilizzo di sistemi di apprendimento "per rinforzo", in la macchina apprende attraverso un processo di trial-and-error, noto anche come metodo di tentativi ripetuti. Si tratta di una tecnica di risoluzione dei problemi in cui si effettuano tentativi e si valutano i loro risultati fino a quando non si raggiunge la soluzione desiderata. Questo metodo viene utilizzato quando non si conosce la soluzione corretta, ma si hanno a disposizione diverse opzioni tra cui scegliere. Il processo inizia con la scelta di una possibile soluzione, che viene testata per verificare se funziona o meno, se la soluzione non funziona, si passa alla successiva e si ripete il processo finché non si trova una soluzione funzionante.

Anche Russell sottolinea l'importanza della trasparenza e della responsabilità nell'uso di questi nuovi strumenti, in modo che gli **utenti possano comprendere come funzionano e siano coinvolti nelle decisioni che li riguardano**.<sup>34</sup>

---

<sup>33</sup> TEGMARK, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. Knopf, *Ethics and governance of artificial intelligence*. Nature, 563(7731), 439-445, <https://www.amazon.it/Life-3-0-Being-Artificial-Intelligence/dp/1101946598>

<sup>34</sup> STUART RUSSELL, *Human Compatible: Artificial Intelligence and the Problem of Control*, 2019, [https://en.wikipedia.org/wiki/Human\\_Compatible](https://en.wikipedia.org/wiki/Human_Compatible); dello stesso Autore: *Artificial Intelligence: The Future Is Superintelligent*, TED, 2015, <https://www.nature.com/articles/548520a>; *AI vs. AI: Can we ensure that advanced AI will do what we want?*, <https://www.tidio.com/blog/ai-test/>.

## 10. Linee guida etiche per un'AI affidabile

I governi possono svolgere un ruolo importante nell'affrontare le questioni etiche legate alle problematiche di cui stiamo parlando, attraverso la regolamentazione e la definizione di linee guida etiche. Inoltre, i governi possono incentivare la ricerca etica e la formazione degli esperti in questo campo, promuovendo così una maggiore consapevolezza e attenzione a questi aspetti.

Evidentemente, non si tratta solo di un problema di protezione dei dati personali, ma anche di giustizia in senso lato, perché l'AI può influenzare il modo in cui le decisioni sono prese e le conseguenze che hanno per le persone coinvolte. Se i modelli sono progettati per massimizzare determinati obiettivi, come la massimizzazione del profitto o della sicurezza, può invero portare a decisioni che non tengono conto delle esigenze e dei diritti delle persone interessate.

I governi possono fare molto per mitigare questi rischi. Ad esempio, possono creare regolamentazioni e standard etici per l'uso dei sistemi, assicurandosi che siano progettati e utilizzati in modo responsabile e *conforme ai valori umani fondamentali*.

Inoltre, i governi possono lavorare per educare il pubblico sui rischi e le opportunità di queste scoperte, *incoraggiando una discussione pubblica aperta e informata sull'argomento*.

L'Unione Europea ha adottato la *European Ethical Guidelines for Trustworthy AI* (Linee guida etiche per un'AI affidabile), un insieme di principi etici che comprendono la trasparenza, l'equità, la responsabilità e la protezione della privacy e dei dati personali. L'8 aprile 2019 un gruppo di esperti ad alto livello scelto dalla commissione ha presentato le linee guida etiche per un'intelligenza artificiale affidabile, raccolte in una prima bozza nel dicembre 2018, sulla quale sono stati ricevuti più di 500 commenti attraverso una consultazione aperta. Secondo le linee guida, un sistema di intelligenza artificiale affidabile dovrebbe essere: (1) lecito - rispetto di tutte le leggi e regolamenti applicabili; (2) etico - rispetto dei principi e dei valori etici; (3)

robusto - sia dal punto di vista tecnico che tenendo conto del suo ambiente sociale.<sup>35</sup>

Le linee guida presentano una serie di sette requisiti chiave che i *sistemi* in questione dovrebbero soddisfare per essere considerati affidabili e uno specifico elenco di valutazione ha lo scopo di aiutare a verificare l'applicazione di ciascuno dei seguenti requisiti chiave:

*Intervento e sorveglianza umani:* dovrebbero responsabilizzare gli esseri umani, consentendo loro di prendere decisioni informate e promuovendo i loro diritti fondamentali. Allo stesso tempo, devono essere garantiti adeguati meccanismi di supervisione, che possono essere raggiunti attraverso approcci umani nel giro, umani nel giro e umani al comando

*Robustezza tecnica e sicurezza:* devono essere resilienti e sicuri. Devono essere sicuri, garantire un piano di riserva nel caso qualcosa vada storto, oltre ad essere accurati, affidabili e riproducibili. Questo è l'unico modo per garantire che anche i danni non intenzionali possano essere ridotti al minimo e prevenuti.

*Riservatezza e governance dei dati:* oltre a garantire il pieno rispetto della privacy e della protezione dei dati, devono essere assicurati adeguati meccanismi di governance dei dati, tenendo conto della qualità e dell'integrità dei dati e garantendo un accesso legittimato ai dati.

*Trasparenza:* i modelli di business di dati, sistemi e AI dovrebbero essere trasparenti. I meccanismi di tracciabilità possono aiutare a raggiungere questo obiettivo. Inoltre, i sistemi e le loro decisioni dovrebbero essere spiegati in modo adeguato alle parti interessate. Gli esseri umani devono essere consapevoli che stanno interagendo con un sistema di intelligenza artificiale e devono essere informati delle capacità e dei limiti del sistema.

*Diversità, non discriminazione ed equità:* i pregiudizi ingiusti devono essere evitati, in quanto potrebbero avere molteplici

---

<sup>35</sup><https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

implicazioni negative, dall'emarginazione dei gruppi vulnerabili, all'esacerbazione del pregiudizio e della discriminazione. Promuovendo la diversità, tali sistemi dovrebbero essere accessibili a tutti, indipendentemente da qualsiasi disabilità, e coinvolgere le parti interessate pertinenti durante l'intero ciclo di vita.

*Benessere sociale e ambientale:* i sistemi di intelligenza artificiale dovrebbero beneficiare tutti gli esseri umani, comprese le generazioni future. Occorre quindi garantire che siano sostenibili e rispettosi dell'ambiente. Inoltre, dovrebbero tenere conto dell'ambiente, compresi gli altri esseri viventi, e il loro impatto sociale e societario dovrebbe essere attentamente considerato.

*Accountability:* dovrebbero essere messi in atto meccanismi per garantire responsabilità e i risultati. L'auditability, che consente la valutazione di algoritmi, dati e processi di progettazione, svolge un ruolo chiave in tale contesto, soprattutto nelle applicazioni critiche.

## **11. Conclusione: i timori ancestrali sono prematuri**

Vorrei concludere questa rassegna di informazioni sull'intelligenza artificiale e sulle novità introdotte in questa materia da ChatGPT-4 proponendo alcuni argomenti di riflessione e dando conto del punto di vista esposto da Noam Chomsky in un articolo comparso sul New York Times l'8 marzo 2023.<sup>36</sup>

Lo studioso sostiene che l'origine delle paure e delle preoccupazioni che hanno suscitato nel nostro animo le notizie dei "prodigiosi" progressi dell'AI, derivano da un misto di entusiasmo e preoccupazione.

---

<sup>36</sup> NOAM CHOMSKY , IAN ROBERTSEJEFFREY WATUMUL, *Noam Chomsky: Noam Chomsky: The False Promise of ChatGPT*, *The New York Times*, <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html?searchResultPosition=1>

Il primo, l'entusiasmo, sarebbe dovuto al fatto che ci rendiamo perfettamente conto che l'intelligenza è l'ingrediente chiave per risolvere problemi e, aggiungerei, la qualità in assoluto più importante dell'umanità, che distingue l'uomo da tutte le altre forme di vita.

La preoccupazione, invece, secondo l'Autore, nasce dalla paura che l'approccio predominante e che dell'AI ci viene in questi giorni trasmesso dai media (la funzione di apprendimento automatico della quale abbiamo parlato nei paragrafi che precedono) possa snaturare le nostre capacità, compromettendo le conquiste dell'uomo nel campo del sapere, sminuendole, per giungere in qualche modo a sostituire l'uomo nelle determinazioni di carattere etico ed esistenziale.<sup>37</sup>

Chomsky ci fa però notare che i recenti sistemi di machine learning non fanno altro che dedicarsi all'apprendimento "automatico" elaborando una enorme quantità di dati<sup>38</sup>, individuare modelli all'interno di essi e migliorare progressivamente la loro capacità di produrre risultati statisticamente plausibili, con un linguaggio e un pensiero che "sembrano" umani perché si presentano nelle forme esteriori tipiche del linguaggio umano.

In proposito credo sia prima di tutto opportuno chiedersi se un sistema che si basa su dati possa veramente essere un precursore dell'intelligenza artificiale c.d. generale (AGI)<sup>39</sup>, di quell'intelligenza cioè che ha la capacità di comprendere, apprendere, adattarsi e applicare la conoscenza in un'ampia varietà di compiti a livelli paragonabili o superiori a quelli di un essere umano. L'intelligenza artificiale "generale" è (o vorrebbe essere) un genere di intelligenza che va oltre la semplice esecuzione di compiti specifici, come avviene nell'AI "stretta" o

---

<sup>37</sup> A questo proposito, rimandiamo a quanto già detto nei paragrafi che precedono di coloro che consigliano di gestire con la massima prudenza i sistemi di intelligenza artificiale in fase di applicazione nell'universo giudiziario.

<sup>38</sup> Abbiamo detto come sia sufficiente compromettere i dati o fornirli in modo parziale per rendere inattendibile il modello su cui si basa l'AI.

<sup>39</sup> AGI, dall'inglese Artificial General Intelligence.

"specializzata", e implica invece la capacità di trasferire l'apprendimento da un dominio all'altro, di elaborare il linguaggio naturale, di risolvere problemi complessi e di esercitare il ragionamento astratto. L'AGI rappresenta *il vertice delle ambizioni* nel campo dell'intelligenza artificiale, poiché si tratta di creare un'entità artificiale che possa effettivamente pensare, comprendere e *comportarsi come un essere umano* - o addirittura superare le capacità umane.

Chomsky afferma che forse, un giorno, questo ambizioso obiettivo verrà raggiunto, ma ci avverte che siamo ancora molto lontani, tanto lontani che non se ne vede ancora l'inizio, nonostante le speculazioni dei media e i nostri timori.

Quel giorno, prosegue, potrebbe arrivare, ma nonostante quanto si può dedurre da titoli sensazionalistici e investimenti impulsivi, all'orizzonte non si scorge proprio nulla. Nonostante i programmi di apprendimento automatico come ChatGPT possano risultare utili in determinati ambiti circoscritti, la scienza della linguistica e la filosofia della conoscenza ci mostrano che differiscono in modo significativo dal modo in cui gli esseri umani ragionano e usano il linguaggio. Queste differenze impongono limiti considerevoli a ciò che questi programmi possono realizzare, inscrivendo in essi difetti irrimediabili: essi generano risultati statisticamente probabili sulla base di una grandissima mole di dati, ma - dice Chomsky - non saranno mai come la mente umana che lui definisce "*un sistema sorprendentemente efficiente e persino elegante che opera con piccole quantità di informazioni*", come accade nella mente di un bambino che "*inconsciamente, automaticamente*" e partendo da una minima quantità di dati, "*sviluppa un sistema stupendamente sofisticato di principi e parametri logici*" attraverso il suo sistema operativo "*innato che dota gli esseri umani della capacità di generare frasi complesse e lunghe serie di pensieri*".

L'essere umano, pur essendo soggetto ad errori, è capace di *critiche creative*, di quelle critiche che, attraverso il confronto

dialettico con i suoi simili, danno luogo al continuo progresso del pensiero ed alla evoluzione dell'umanità in generale.

L'essere umano agisce in simbiosi, soprattutto sul piano del confronto delle idee e delle opinioni, con i suoi simili, dei quali, in quasi tutti gli aspetti della sua esistenza, non può sostanzialmente fare a meno. Sotto questo profilo, potremmo dire che i modelli di AI non sentono il "bisogno" di confrontarsi con i propri simili, come fanno invece gli esseri umani, non sono spinti da sentimenti, non si cercano. Essi, invero, non fanno altro che rimanere inerti in attesa di essere riforniti di dati, mentre l'uomo, con tutti i suoi errori e le inesattezze, è irrequieto, in continua tensione e ricerca e grazie a queste spinte è capace di creare arte, di produrre idee, di amare.

L'intelligenza artificiale, anche nei suoi modelli più avanzati, è solamente un insieme di algoritmi e dati, non ha coscienza, emozioni, desideri o bisogni. Non ha un'esperienza soggettiva e non può apprezzare l'arte, innamorarsi o avere un'idea originale, non ha una motivazione intrinseca o un senso di curiosità, si limita a eseguire i compiti per cui è stata programmata, utilizzando i dati che le sono stati forniti.

D'altra parte, noi esseri umani siamo entità complesse e multidimensionali, abbiamo la capacità di pensare in modo critico, di sentire emozioni, di sognare, di creare arte e di amare. Queste caratteristiche uniche ci permettono di vivere esperienze ricche e significative, di imparare e crescere, e di contribuire alla società in modi che vanno oltre la semplice esecuzione di compiti.<sup>40</sup>

L'intelligenza artificiale può essere uno strumento potente e utile, ma non può sostituire l'umanità. Può aiutarci a risolvere problemi complessi, a ottimizzare i processi e a fare scoperte,

---

<sup>40</sup> Su questo argomento, segnalo l'interessante articolo di G. ZAGNI, *Leggeremo solo opere di ChatGPT?* Apr 2023, <https://giovannizagni.substack.com/p/i-libri-artificiali>

ma la creatività, l'empatia, l'intuizione e altre caratteristiche propriamente umane sono fuori dalla sua portata.<sup>41</sup>

---

<sup>41</sup> La parte di testo compresa da "*L'intelligenza artificiale, anche nei suoi modelli più avanzati...*" sino a "*... sono fuori dalla sua portata...*" sono stati elaborati da ChatGPT-4 in risposta alle mie affermazioni contenute nei tre periodi che precedono, tali e quali inseriti nella chat come pura asserzione, non accompagnata da alcuna domanda.