



La disponibilità al pubblico della piattaforma di Intelligenza Artificiale (AI) ChatGPT (<https://chat.openai.com/auth/login>) ha subito suscitato interesse e polemiche: a queste ha anche contribuito la decisione del Garante italiano della Privacy di sospenderne temporaneamente l’accesso richiedendo modifiche alla modalità di raccolta dati ed al controllo sull’età minima di accesso al sistema. Ora il servizio è nuovamente accessibile, mentre l’argomento AI continua a far discutere; ecco, allora, alcune riflessioni sul tema.

Che cosa è e che cosa c’è di “intelligente” in ChatGPT?

ChatGPT è un robot conversazionale (chatbot) sviluppato da OpenAI (ora di proprietà Microsoft) e basato sui modelli definiti Generative Pretrained Transformer (GPT) ma irrobustiti con una tecnica detta Apprendimento tramite Rinforzo Umano (Reinforcement Learning from Human Feedback o RLHF)⁽¹⁾. Durante lo sviluppo del modello, infatti, è stato chiesto ai supervisori umani di fornire sia le domande che le risposte desiderate: queste sono state utilizzate per ottimizzare il modello ChatGPT.

Questi programmi usano una “rete neurale” di codice per analizzare un testo in ingresso (il “prompt” scritto dall’utente), decodificare il significato della richiesta attribuendo dei valori di peso alle varie parole, ricercare nelle centinaia di miliardi di documenti memorizzati il termine successivo che sembra più probabile e confrontare l’esito della ricerca con il prompt reale e - sulla base del risultato - riaggiustare i pesi delle parole per cercare di ridurre la distanza tra l’ipotesi del programma ed il testo reale. Questo ciclo viene ripetuto più e più volte sino a che la differenza tra la richiesta vera e la risposta calcolata sia ritenuta accettabilmente ridotta: a questo punto può essere generato il testo (o anche l’immagine) in risposta al nostro prompt⁽²⁾.

Questo modo operativo apparentemente così privo di comprensione permette comunque a ChatGPT di generare risposte (in qualche caso sbagliate) che mostrano un’intelligenza umana o, su certi tipi di problemi (matematici, logici, di programmazione), apparentemente sovra-umana, nel senso che non possono essere completamente ricondotte ai dati forniti nell’addestramento.

ChatGPT è stato realizzato per rendere l’interazione con l’AI più naturale ed intuitiva e far conoscere questa tecnologia al grande pubblico. Tramite ChatGPT è possibile utilizzare, con il piano gratuito, il modello GPT3.5, mentre per coloro che hanno l’accesso a ChatGPT Plus (a pagamento) è possibile utilizzare il modello più evoluto GPT4.

I limiti dell’AI

Il sistema ChatGPT memorizza le informazioni fornite dagli utenti (da cui le preoccupazioni del Garante), che diventano parte del patrimonio di dati del sistema; in questo processo si nota però una certa tendenza ad “assecondare” l’utente, cercando di non contraddirlo anche in caso di richieste contenenti errori e fornire risposte il più aderenti possibile al contenuto del prompt. Questo atteggiamento può avere un senso nell’ambito di una conversazione piacevole, ma può essere un problema, se non un pericolo, nel momento in cui si utilizza ChatGPT per ottenere informazioni su argomenti che non conosciamo.

Un punto particolarmente critico è che il sistema, se non trova la risposta alla richiesta tra le sue informazioni, può inventarla: ChatGPT fa un po’ fatica ad ammettere di non sapere. Se la licenza di inventare ha un senso per una richiesta di tipo artistico, diventa pericolosa quando l’utente richiede un’informazione specifica, soprattutto se questi non ha consapevolezza dei limiti e prende per certa ogni risposta ricevuta⁽³⁾.

I vantaggi dell’AI

L’utilità dei sistemi di AI come ChatGPT è certamente evidente. Nella sanità è possibile colloquiare con un paziente ed effettuare una pre-analisi dei sintomi per indirizzarlo verso il giusto reparto, come pure dare supporto conversazionale a pazienti che hanno difficoltà a relazionarsi con altri individui; nell’istruzione si possono avere lezioni più interattive e approfondite, come pure insegnare agli studenti l’utilizzo consapevole di questi strumenti. In ambito aziendale è possibile utilizzare l’AI per statistiche e pianificazioni finanziarie e di produzione, e avere supporto nelle fasi di sviluppo e manutenzione di prodotti. Sono immaginabili molti altri vantaggi, anche in settori non propriamente pacifici (ad esempio, sistemi decisionali militari).

I problemi dell’AI

Oltre alle citate creatività fuori luogo e mancanza di autocritica, i sistemi come ChatGPT ci pongono di fronte a molte nuove sfide e rischi. ChatGPT manipola il linguaggio, che è la base della cultura umana: il linguaggio racconta quello che siamo, i nostri desideri, i nostri sogni, i nostri miti, le nostre leggi, le religioni, l’amore, la vita intera. Che cosa

significa avere un sistema che può produrre senza difficoltà nuovi testi, nuove leggi, nuovi dei, nuove storie? Chi potrà confutare una storia del passato generata dall'AI, chi potrà opporsi a regole scritte per "il bene dell'umanità" da un'entità così "intelligente"? Come si potrà distinguere una notizia vera da una falsa generate dalla stessa AI? Non ci vuole molto a capire che i sistemi di intelligenza artificiale sono diversi da ogni altra scoperta umana del passato; l'energia atomica è un formidabile strumento di progresso o di distruzione, ma non è in grado di liberarsi da sola: è l'uomo ad utilizzarla e per questo si è dotato di leggi e strumenti per cercare di piegarla al bene anziché al male. Quando le macchine saranno dotate di AI, come si potrà controllarle ed evitare che "vivano" in modo autonomo?

Per dirla come il filosofo Yuval Noah Harari⁽⁴⁾: *"Abbiamo appena incontrato un'intelligenza aliena, qui sulla Terra. Non sappiamo molto di lei, tranne che può distruggere la nostra civiltà. Dovremmo porre un limite alla distribuzione irresponsabile degli strumenti di AI nella sfera pubblica, e regolamentare l'AI prima che essa regolamenti noi"*.

Qui termina il primo sguardo, il nostro, sull'AI.

Il secondo lo abbiamo ottenuto direttamente da ChatGPT.

D: Che cosa è e che cosa c'è di "intelligente" in ChatGPT?

R: ChatGPT è un modello di linguaggio basato sull'architettura GPT (Generative Pre-trained Transformer) sviluppato da OpenAI. È un'IA (Intelligenza Artificiale) che è stata addestrata su una vasta quantità di testo in diverse lingue, al fine di generare testo in modo autonomo e fornire risposte in base al contesto e alla domanda posta. In altre parole, ChatGPT è un assistente virtuale in grado di rispondere a diverse domande e conversare in modo naturale con gli utenti.

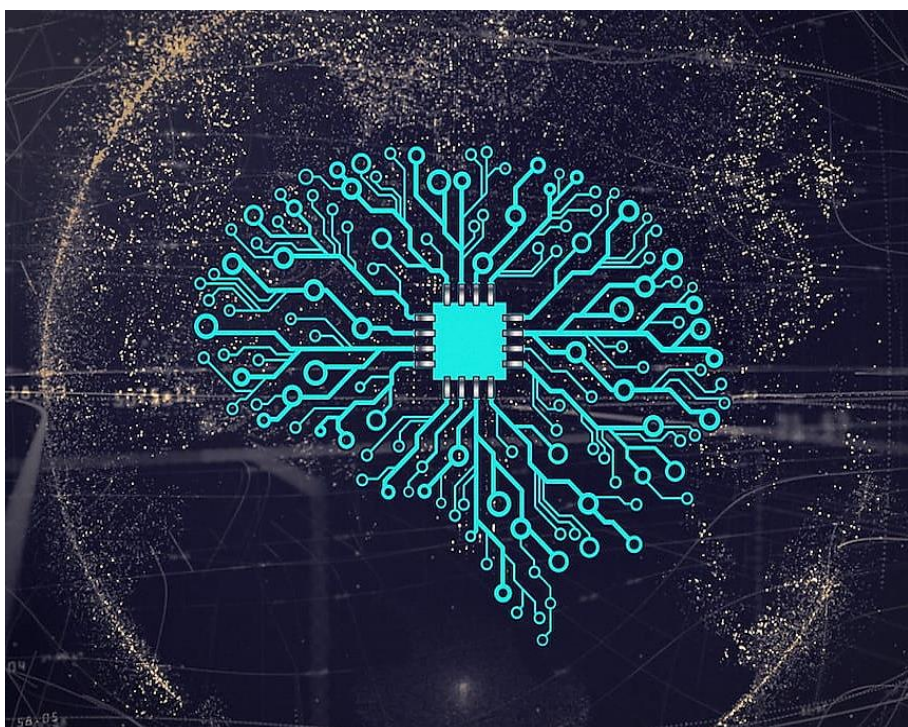
L'"intelligenza" di ChatGPT risiede nella sua capacità di elaborare grandi quantità di dati testuali e di utilizzare tale conoscenza per generare testo in

modo autonomo. Ciò avviene attraverso un processo di apprendimento automatico, in cui il modello viene addestrato su enormi dataset di testo, al fine di apprendere le regolarità linguistiche e semantiche della lingua in questione. Ciò consente a ChatGPT di rispondere a una vasta gamma di domande e di generare testo coerente e comprensibile in base al contesto e alla domanda posta, fornendo un'esperienza di conversazione "intelligente" agli utenti. In sostanza, l'intelligenza di ChatGPT risiede nella sua capacità di utilizzare la conoscenza linguistica appresa per risolvere problemi di natura linguistica e generare risposte coerenti e adeguate alle domande poste dagli utenti.

D: Quali sono i limiti di ChatGPT?

R: Nonostante le sue notevoli capacità, ChatGPT ha alcuni limiti che possono influire sulla sua capacità di fornire risposte precise e appropriate in determinate situazioni. Alcuni di questi limiti sono:

1. **Conoscenza limitata:** ChatGPT si basa sulla conoscenza appresa dai dati di addestramento, che possono essere limitati in alcune aree specifiche. Ciò significa che potrebbe non essere in grado di fornire risposte precise o complete su argomenti meno frequenti o complessi.
2. **Mancanza di comprensione del contesto:** ChatGPT può avere difficoltà a comprendere il contesto più ampio di una conversazione, specialmente se le domande sono poste in modo ambiguo o se manca contesto sufficiente. Ciò potrebbe portare a risposte imprecise o fuorvianti.
3. **Problemi di coerenza:** anche se ChatGPT è in grado di generare testo coerente, potrebbe avere difficoltà a mantenere una coerenza logica a lungo termine. Ciò significa che potrebbe generare risposte che sembrano appropriate, ma che in realtà non sono del tutto coerenti o logiche.



4. *Bias: come qualsiasi sistema di apprendimento automatico, ChatGPT può essere influenzato dai bias [pregiudizi, NdA] presenti nei dati di addestramento. Ciò potrebbe portare a risposte che riflettono i pregiudizi o le limitazioni del dataset, piuttosto che rappresentare una risposta obiettiva.*

In generale, ChatGPT può fornire risposte precise e coerenti a molte domande, ma è importante considerare i suoi limiti e utilizzarlo con cautela, soprattutto per argomenti più complessi o sensibili.

D: Quali sono i vantaggi dell'IA?

R: *L'intelligenza artificiale (IA) offre numerosi vantaggi in molte aree della vita, tra cui:*

1. *Automazione delle attività ripetitive: L'IA può automatizzare le attività ripetitive, riducendo così l'onere del lavoro umano. Ciò consente alle persone di concentrarsi su attività che richiedono abilità umane uniche e creative.*
2. *Miglioramento dell'efficienza: L'IA può aiutare a migliorare l'efficienza dei processi, ridurre i tempi di attesa e aumentare la velocità delle operazioni. Ciò può portare a una maggiore produttività e a un miglioramento dell'esperienza del cliente.*
3. *Personalizzazione: L'IA può essere utilizzata per personalizzare l'esperienza dell'utente, offrendo suggerimenti personalizzati e raccomandazioni in base alle preferenze individuali.*
4. *Miglioramento della precisione: L'IA può migliorare la precisione dei processi, riducendo così gli errori umani e migliorando la qualità dei prodotti e dei servizi.*
5. *Analisi dei dati: L'IA può essere utilizzata per analizzare grandi quantità di dati, rivelando modelli e tendenze che potrebbero essere altrimenti difficili da individuare.*
6. *Assistenza sanitaria: L'IA può essere utilizzata nella diagnostica medica, nella ricerca farmaceutica e nella progettazione di terapie personalizzate, migliorando così la salute e il benessere delle persone.*

In sintesi, l'IA ha il potenziale per migliorare molti aspetti della vita umana, offrendo soluzioni innovative e efficienti ai problemi attuali e futuri.

D: Quali sono i problemi dell'AI?

R: *L'intelligenza artificiale (IA) ha anche alcuni problemi e sfide che devono essere considerati, tra cui:*

1. *Bias e discriminazione: l'IA può essere influenzata dai bias presenti nei dati di addestramento, portando a discriminazione su base di razza, genere, orientamento sessuale o altre caratteristiche. Ciò potrebbe avere conseguenze negative sulla società e sulle persone coinvolte.*
2. *Trasparenza e accountability: molte applicazioni dell'IA sono "black box", il che significa che il processo decisionale non è trasparente e non è sempre chiaro come il sistema arriva alle sue conclusioni. Ciò rende difficile la responsabilizzazione e la comprensione delle decisioni prese dall'IA.*
3. *Disoccupazione: L'IA potrebbe portare alla disoccupazione di alcune categorie di lavoratori, poiché molte attività umane potrebbero essere automatizzate.*
4. *Affidabilità: l'IA è ancora soggetta a errori, e potrebbe non sempre essere affidabile al 100% quando si tratta di compiti critici per la vita umana.*
5. *Privacy e sicurezza: l'IA potrebbe mettere a rischio la privacy e la sicurezza dei dati personali, specialmente se le informazioni raccolte sono usate in modo inappropriato o malintenzionato.*

In sintesi, l'IA ha il potenziale per migliorare molti aspetti della vita umana, ma è importante considerare i suoi limiti e sfide per garantire che venga utilizzata in modo responsabile e sicuro.

Riferimenti:

⁽¹⁾ [ChatGPT, cos'è, come funziona e come accedere gratuitamente \(smartstrategy.eu\)](https://www.smartstrategy.eu)

⁽²⁾ [Sotto il cofano dei sistemi GPT: ecco come fanno a capire cosa chiediamo - Agenda Digitale](#)

⁽³⁾ [ChatGPT, licenza di inventare? Il "caso" del rinnovo del passaporto - Agenda Digitale](#)

⁽⁴⁾ <https://www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation?giftId=ece231e6-9cc2-418b-b231-3de21daa9828>