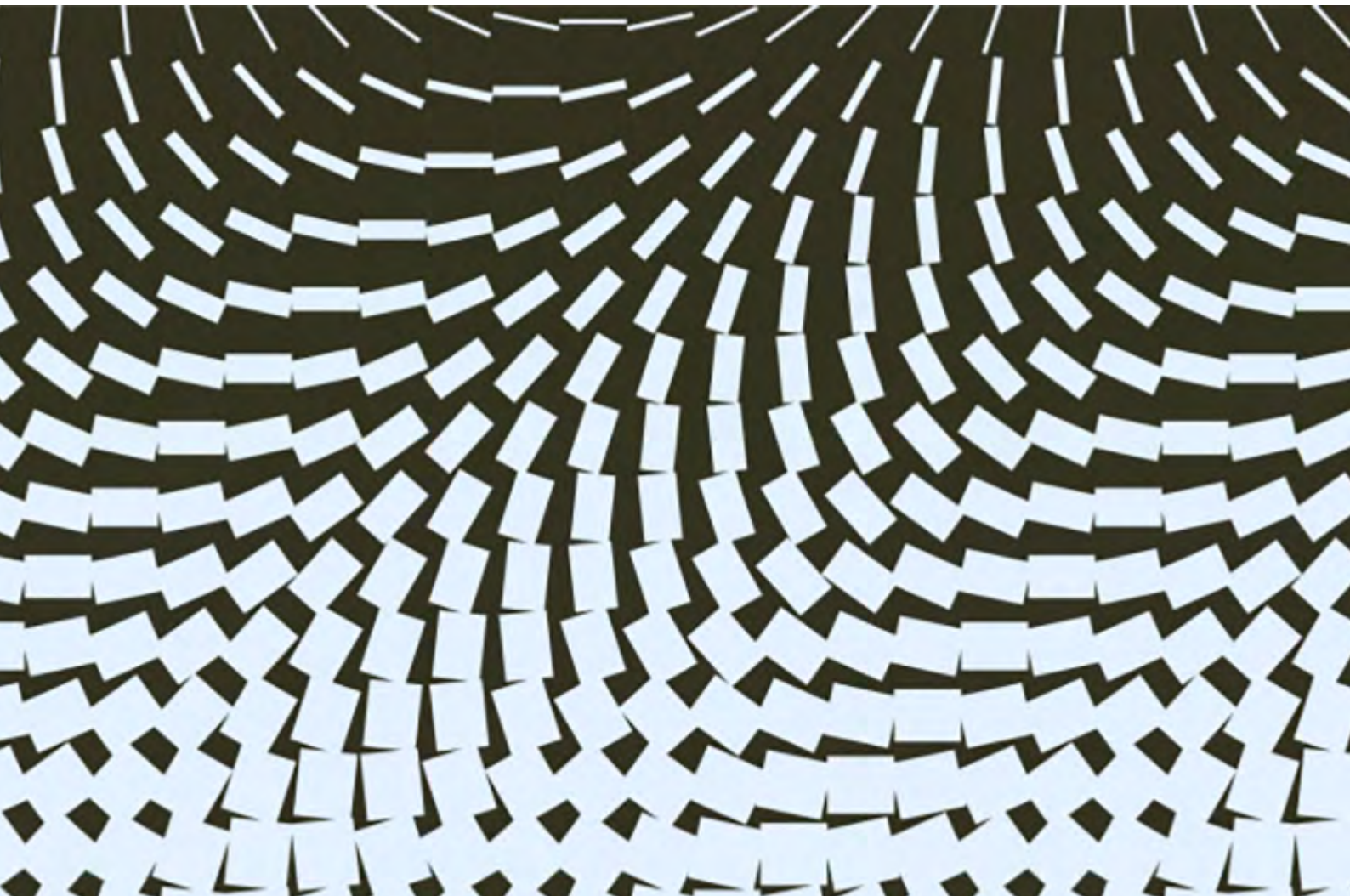
# ChatGPT di OpenAI: tutto quello che c'è da sapere sulla piattaforma che ha rivoluzionato i chatbot

Testi a cura di Pierluigi Sandonnini



# INDICE DEGLI ARGOMENTI

*ChatGPT è un modello di pari livello di InstructGPT, rilasciato a gennaio 2023 da OpenAi e progettato per fornire una risposta dettagliata a una istruzione in un prompt. In questo documento spieghiamo come funziona, cosa chiedergli, quali sono i limiti, quali le applicazioni. Infine, una breve cronistoria dei modelli di AI precedenti.*

**ChatGPT (Chat Generative Pretrained Transformer)** è un chatbot progettato per rispondere alle domande, un nuovo modello linguistico, **sviluppato OpenAI, la società che sta dietro AlphaFold e GPT-3**.

La società, sostenuta da Microsoft, ha sviluppato un **modello linguistico di grandi dimensioni** (LLM) progettato per interagire con gli utenti in" modo colloquiale", che può essere utilizzato anche **per eseguire il debug del codice**.

Il formato di dialogo del modello **consente a ChatGPT di rispondere a domande di follow-up, ammettere i propri errori, sfidare premesse errate e rifiutare richieste inappropriate**.

**ChatGPT** è un modello di pari livello di InstructGPT, rilasciato a gennaio 2023 e progettato per fornire una risposta dettagliata a un'istruzione in un prompt.

# 1 COSA SI PUÒ CHIEDERE A CHATGPT

Sono tantissime le cose che possiamo chiedere, per avere risposte utili.

In sostanza possiamo usarlo sia per avere informazioni in modo diretto (più diretto rispetto a un motore di ricerca) sia per aiutarci a scrivere qualcosa.

## 1.1 Esempi di utilizzo

- richiesta di informazioni di base – ChatGPT può aiutare a ottenere informazioni di base su qualsiasi cosa;

*scrivere:*

- un saggio o un riassunto su qualsiasi argomento
- una canzone o poesia su qualsiasi argomento;
- una e-mail in uno stile definito;
- una sceneggiatura cinematografica;
- un codice;
- eseguire il debug del codice;
- chiedere consigli su qualsiasi cosa: turismo, diete…

2.

# CONSIGLI PER L'USO DI CHATGPT

- Meglio essere specifici. Come spiega Bloomberg, se chiediamo a ChatGPT "Che cos'è il marxismo?", per esempio, avremo risposta passabile, probabilmente non migliore di quella su Wikipedia o Google. Invece, rendete la domanda più specifica: "Quali sono stati gli sviluppi importanti del marxismo francese nella seconda metà del XIX secolo?". ChatGPT farà molto meglio, ed è anche il tipo di domanda a cui è difficile rispondere con Google e Wikipedia.

- ChatGPT farà ancora meglio se gli farete domande una dopo l'altra, per approfondire o spiegare meglio punti specifici. Chiedetegli informazioni sugli specifici marxisti francesi che cita, su cosa facevano e su come si differenziavano dalle loro controparti tedesche.

- ChatGPT è particolarmente bravo nel "confronto e nella contrapposizione". In sostanza, ChatGPT ha bisogno che lo indirizziamo bene. Una domanda ben fatta gli dà più punti di riferimento fissi. Dovete impostare l'atmosfera, il tono e il livello intellettuale della vostra domanda, a seconda del tipo di risposta che desiderate.

- Un altro modo per affinare le capacità di ChatGPT è quello di chiedere risposte con la voce di una terza persona. Se chiediamo: "Quali sono i costi dell'inflazione?", potreste ottenere risposte che non sono esattamente sbagliate, ma nemmeno tanto interessante. Più interessante: "Quali sono i costi dell'inflazione? Rispondete utilizzando le idee di Milton Friedman".

3

# COME FUNZIONA CHATGPT?

ChatGpt è **un tipo di intelligenza artificiale progettata per generare testi simili a quelli umani in base agli input ricevuti**. Utilizza **algoritmi di apprendimento automatico** per elaborare l'input e generare risposte coerenti e appropriate in linguaggio naturale.

Il bot deriva da modelli di apprendimento automatico per l'elaborazione del linguaggio naturale noti come **Large Language Model (LLM)**. Gli LLM analizzano enormi quantità di dati testuali e deducono le relazioni tra le parole all'interno del testo. Colgono insomma le correlazioni statistiche (pattern) tra una parola e l'altra. In questa maniera possono poi "scrivere" una risposta prevedendo una parola dopo l'altra.

Questi modelli sono cresciuti negli ultimi anni grazie ai progressi della potenza di calcolo. I LLM accrescono le loro capacità con l'aumentare delle dimensioni dei dataset di input e dello spazio dei parametri.

L'addestramento più elementare dei modelli linguistici prevede la previsione di una parola in una sequenza di parole. Più comunemente, è quanto viene definito come *next-token-prediction e masked language-modeling*.

I dati di addestramento del bot sono costituiti da una vasta quantità di testi provenienti da diverse fonti, tra cui libri, articoli e forum online. Questo permette al bot di comprendere il contesto e la struttura del linguaggio usato in diverse situazioni e di generare risposte adeguate all'input dato.

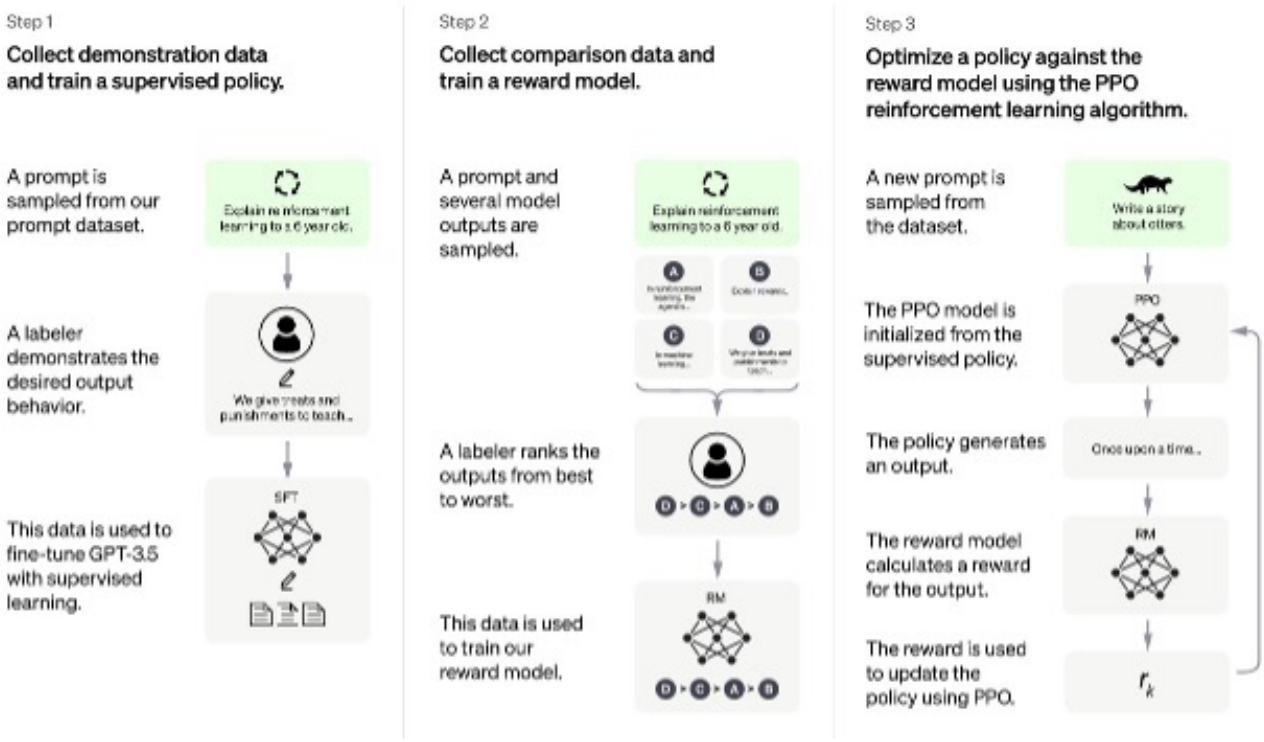
Utilizza una combinazione di tecniche, tra cui una **rete neurale ricorrente (RNN)** e un modello di **trasformazione (transformer)**. La RNN elabora l'input che l'assistente riceve e lo utilizza per generare una sequenza di parole o token. Il modello di trasformazione prende il risultato della RNN e lo usa per generare il testo finale. Questa combinazione di tecniche permette al chatbot di generare risposte coerenti e appropriate in tempo reale.

In particolare, ChatGPT è stato costruito utilizzando l'apprendimento per rinforzo dal feedback umano (RLHF). **Il metodo è progettato per eliminare le risposte errate premiando le risposte corrette mentre corregge quelle che non lo sono**.

Un modello iniziale è stato addestrato utilizzando la messa a punto supervisionata: i formatori di AI umani hanno fornito conversazioni in cui hanno giocato sia l'utente che un assistente AI. I formatori hanno avuto accesso a suggerimenti scritti su modelli per aiutarli a comporre le loro risposte.

## 3.1
## GPT

Il modello stesso è stato costruito e messo a punto con GPT-3.5, una versione rivista del modello di punta di OpenAI che condivide il suo nome ma è costruito per essere migliore nella generazione di testo dettagliato. Sia ChatGPT che GPT 3.5 sono stati addestrati su un'infrastruttura di supercomputing di intelligenza artificiale di Azure, che Microsoft ha creato in collaborazione con OpenAI a maggio 2020.



# 4 LIMITI DI CHATGPT

ChatGPT è effettivamente un chatbot e come altri è soggetto a generare **output offensivo**, come nel caso di BlenderBot 3 di Meta. OpenAI afferma di aver preso lezioni dai modelli precedenti, come GPT-3 e Codex, per cercare di migliorare l'affidabilità di ChatGPT, incluso l'utilizzo della sua tecnica di apprendimento per rinforzo per ottenere "sostanziali riduzioni di output dannosi e non veritieri". Ha applicato anche filtri etici per limitare hate speech e disinformazione.

Tuttavia, l'azienda di intelligenza artificiale riconosce che il suo modello ha dei **limiti**. Ad esempio, a volte scrive risposte plausibili ma errate o senza senso. Ha problemi anche nel fare calcoli. A volta ha quelle che gli esperti chiamano “allucinazioni”, confondendo realtà con fantasia. Gpt coglie infatti, come detto, i rapporti statistici tra le parole, ma non ha contezza dei rapporti tra parole e il mondo. Non conosce né capisce insomma, davvero, ciò che scrive.

Può alimentare così la disinformazione, ad esempio in ambito medico o politico, come rilevato da molti esperti, nonostante i filtri presenti.

Il modello è anche prolisso, tende ad avere uno stile piatto e abusa di alcune frasi, come riaffermare che si tratta di un modello linguistico addestrato da OpenAI, che secondo la società deriva da pregiudizi nei dati di addestramento.

OpenAI afferma anche che l'**AI potrebbe avere una conoscenza limitata degli eventi o delle conoscenze dopo il 2021**, in base a quando il modello è stato addestrato.

Per affrontare i limiti del modello, OpenAI ha dichiarato di volersi impegnare a fare aggiornamenti regolari per migliorarlo e che lavorerà anche per fornire un'interfaccia accessibile.

5

# COSA NE PENSANO DI CHAT GPT GLI ESPERTI DI AI

Come per ogni nuova versione di modello, la comunità AI ha reagito in modi molto diversi. Alcuni hanno chiesto a ChatGPT di generare prompt per generatori di testo-immagine, come MidJourney o Stable Diffusion, e li hanno utilizzati su questi modelli.

La prof.ssa Emily M. Bender, direttrice del **Computational Linguistics Laboratory** dell'Università di Washington, ha ricordato che ChatGPT:" *è ancora solo un modello linguistico: solo una macchina di sintesi testuale / generatore casuale di BS".*

Pochi giorni dopo il rilascio del modello, StackOverflow, il sito di domande e risposte dei programmatori, ha temporaneamente vietato agli utenti di pubblicare risposte generate da ChatGPT, dicendo che il volume di risposte errate ma plausibili era troppo grande.

6

# A COSA SERVE CHATGPT IN PRATICA

Strumenti come ChatGPT possono creare grandi opportunità per le aziende che sfruttano la tecnologia in modo strategico. L'intelligenza artificiale basata su chat può aumentare il modo in cui gli esseri umani lavorano automatizzando le attività ripetitive e fornendo interazioni più coinvolgenti con gli utenti.

Ecco alcuni dei modi in cui le aziende e i professionisti possono utilizzare strumenti come ChatGPT:

- Compilazione di ricerche
- Idee di brainstorming
- Scrittura di codice informatico
- Automatizzare parti del processo di vendita
- Fornire servizi post-assistenza quando i clienti acquistano prodotti
- Fornire istruzioni personalizzate
- Razionalizzazione e miglioramento dei processi utilizzando l'automazione
- Traduzione di testo da una lingua all'altra
- Facilitare il processo di onboarding dei clienti
- Aumentare il coinvolgimento dei clienti, con conseguente miglioramento della fidelizzazione e della fidelizzazione

Il servizio clienti è una grande area di opportunità per molte aziende. Le aziende possono utilizzare la tecnologia ChatGPT per generare risposte per i propri chatbot del customer care in modo da poter automatizzare molte attività tipicamente svolte dagli esseri umani e migliorare radicalmente i tempi di risposta.

Secondo un rapporto di Opus Research, il 35% dei consumatori vorrebbe che più aziende utilizzassero i chatbot e al 48% dei consumatori non importa se un essere umano o un chatbot automatizzato li aiuta con una query del servizio clienti. D'altra parte, le opportunità dei chatbot basati sull'intelligenza artificiale possono trasformarsi in minacce se i tuoi concorrenti sfruttano con successo questa tecnologia e la tua azienda no.

7

# CHATGPT, I SETTORI AZIENDALI IN CUI POTREBBE ESSERE IMPIEGATO

Secondo McKinsey & co., le funzioni in cui modelli come ChatGPT potrebbero essere usati sono:

- *Marketing e vendite*: creazione di contenuti di marketing, social media e di vendita tecnici personalizzati (inclusi testo, immagini e video); Creazione di assistenti allineati ad attività specifiche, come la vendita al dettaglio
- *Operazioni*: generazione di elenchi di attività per l'esecuzione efficiente di una determinata attività
- *IT/ingegneria*: scrittura, documentazione e revisione del codice
- *Rischio e legale*: rispondere a domande complesse, attingere da grandi quantità di documentazione legale e redigere e rivedere relazioni annuali
- *R&S*: accelerare la scoperta di farmaci attraverso una migliore comprensione delle malattie e la scoperta di strutture chimiche.

**Example use cases[1] (not exhaustive)**

| Marketing and sales | Operations | IT/engineering | Risk and legal | HR | Utility/employee optimization |
|---|---|---|---|---|---|
| **Write marketing and sales copy including text, images, and videos** (eg, to create social media content or technical sales content) | **Create or improve customer support chatbots** to resolve questions about products, including generating relevant cross-sell leads | **Write code and documentation** to accelerate and scale developments (eg, convert simple JavaScript expressions into Python) | **Draft and review legal documents,** including contracts and patent applications | **Assist in creating interview questions for candidate assessment** (eg, targeted to function, company philosophy, and industry) | **Optimize communication of employees** (eg, automate email responses and text translation or change tone or wording of text) |
| **Create product user guides** of industry-dependent offerings (eg, medicines or consumer products) | **Identify production errors, anomalies, and defects** from images to provide rationale for issues | **Automatically generate or auto-complete data tables** while providing contextual information | **Summarize and highlight changes** in large bodies of regulatory documents | **Provide self-serve HR functions** (eg, automate first-line interactions such as employee onboarding or automate Q&A or strategic advice on employment conditions, law, regulations, etc) | **Create business presentations** based on text prompts, including visualizations from text |
| **Analyze customer feedback** by summarizing and extracting important themes from online text and images | **Streamline customer service** by automating processes and increasing agent productivity | **Generate synthetic data** to improve training accuracy of machine learning models with limited unstructured input | **Answer questions from large amounts of legal documents,** including public and private company information | | **Synthesize a summary** (eg, from text, slide decks, or online video meetings) |
| **Improve sales force** by, for example, flagging risks, recommending next interactions such as additional product offerings, or identifying optimal customer interaction that leads to growth and retention | **Identify clauses of interest,** such as penalties or value owed through leveraging comparative document analysis | | | | **Enable search and question answering** on companies' private knowledge data (eg, intranet and learning content) |
| **Create or improve sales support chatbots** to help potential clients understand, including technical product understanding, and choose products | | | | | **Automated accounting by sorting and extracting documents** using automated email openers, high-speed scanners, machine learning, and intelligent document recognition |

*Fonte: McKinsey*

## 7.1 MCKINSEY: CREARE UN TEAM PER RISPONDERE ALLE PRINCIPALI DOMANDE

Per McKinsey, nelle aziende che considerano l'AI generativa, i dirigenti vorranno identificare rapidamente le parti del loro business in cui la tecnologia potrebbe avere l'impatto più immediato e implementare un meccanismo per monitorarla, dato che si prevede che si evolverà rapidamente. Una mossa è quella di riunire un **team interfunzionale**, inclusi professionisti della Data science, esperti legali e leader aziendali funzionali, per riflettere su domande di base, come queste:

- In che modo la tecnologia potrebbe aiutare o interrompere il nostro settore e/o la catena del valore della nostra attività?
- Quali sono le nostre politiche e la nostra posizione? Ad esempio, stiamo aspettando attentamente di vedere come si evolve la tecnologia, investendo in progetti pilota o cercando di costruire un nuovo business? La postura dovrebbe variare tra le aree dell'azienda?
- Dati i limiti dei modelli, quali sono i nostri criteri per selezionare i casi d'uso da targetizzare?
- Come perseguiamo la costruzione di un ecosistema efficace di partner, comunità e piattaforme?
- A quali standard legali e comunitari dovrebbero aderire questi modelli in modo da poter mantenere la fiducia con i nostri stakeholder?

Nel frattempo, è essenziale incoraggiare un'innovazione ponderata in tutta l'organizzazione, posizionando guardrail insieme ad ambienti *sandbox* per la sperimentazione, molti dei quali sono prontamente disponibili tramite il cloud, con più probabilità all'orizzonte.

8

# CHATGPT, I POTENZIALI PERICOLI

Ci sono enormi opportunità per le aziende di utilizzare strumenti come ChatGPT per migliorare i loro profitti e creare esperienze migliori per i clienti, ma ci sono anche alcuni potenziali pericoli con questa tecnologia. Nella versione beta del software ChatGPT, OpenAI riconosce gli attuali limiti dell'AI, incluso il potenziale di generare occasionalmente informazioni errate o contenuti distorti.

Ci sono anche **potenziali problemi di privacy, per tutte le informazioni che ottiene dagli utenti**.

ChatGPT può essere vulnerabile agli attacchi di sicurezza informatica, in quanto è connesso a Internet e potrebbe potenzialmente essere utilizzato per diffondere contenuti dannosi o virus. I criminali informatici malintenzionati potrebbero anche manipolare le persone affinché divulghino informazioni personali utilizzando il chatbot, quindi utilizzare tali informazioni per scopi fraudolenti o per attacchi di phishing mirati.

## 8.1
## LE LIMITAZIONI IMPOSTE

Ai primi di gennaio 2023, l'**International Conference on Machine Learning (ICML)** ha impedito agli autori di presentare articoli contenenti testo generato da strumenti come ChatGPT. Inoltre, le scuole di New York City hanno vietato agli studenti di utilizzare ChatGPT per paura che possa essere usato per imbrogliare. Ora è vietato su tutti i dispositivi e le reti nelle scuole pubbliche.

Un portavoce di OpenAI ha dichiarato al Washington Post che sta sviluppando "mitigazioni per aiutare chiunque a identificare il testo generato da quel sistema".

C'è poi il problema della legalità delle opere generate dall'AI; attualmente, solo un'opera ha ottenuto la protezione del copyright: un fumetto negli Stati Uniti, anche se l'Ufficio copyright Usa sta rivedendo la sua decisione.

# 9 COME USARE GRATUITAMENTE CHATGPT

Si può provare ChatGPT gratuitamente durante la fase di anteprima della ricerca. Per accedere al modello, si deve creare un account OpenAI e dichiarare di avere più di 18 anni.

## 9.1 CHATGPT A PAGAMENTO

È uscita già una versione a pagamento, per ora solo negli Usa, di ChatGpt, a 20 dollari al mese, per un uso senza attesa e senza limiti di caratteri.

10

# CHATGPT IN BING ED EDGE DI MICROSOFT

ChatGPT può essere utilizzato per rispondere in modo sintetico alle query di ricerca nel motore di ricerca **Bing** di Microsoft, invece di mostrare semplicemente i collegamenti. **Questa nuova versione potrebbe essere lanciata entro la fine di marzo. Da febbraio è disponibile con accesso limitato (c'è una lista di attesa).**

In questo modo, **Microsoft sta cercando di guadagnare terreno nei confronti di Google Search**. Secondo Statista, Google Search detiene una quota di mercato dell'84% rispetto al 9% per il secondo posto di Bing (a settembre 2022).

Microsoft ha incorporato la tecnologia di OpenAI anche in **Edge**, il suo browser web, come una sorta di assistente di scrittura superpotenziato. Gli utenti possono ora aprire un pannello in Edge, digitare un argomento generico e ottenere un paragrafo, un post sul blog, un'e-mail o un elenco di idee generati dall'IA e scritti in uno dei cinque toni disponibili (professionale, informale, informativo, entusiasta o divertente). Possono incollare il testo direttamente in un browser web, in un'applicazione di social media o in un client di posta elettronica.

Gli utenti possono anche chattare con l'AI integrata in Edge su qualsiasi sito web che stanno visualizzando, chiedendo riassunti o informazioni aggiuntive. In una dimostrazione, un dirigente Microsoft ha navigato sul sito web di *Gap*, ha aperto un file PDF con i risultati finanziari trimestrali più recenti dell'azienda e ha chiesto a Edge di riassumere i dati principali e di creare una tabella di confronto con i risultati finanziari più recenti di un'altra azienda di abbigliamento, *Lululemon*. L'AI ha fatto entrambe le cose, quasi istantaneamente.

L'integrazione di ChatGPT in futuro – come dichiarato dall'azienda – si estenderà a tutti i prodotti Microsoft Office, tra cui Word, PowerPoint e Outlook per **consentire agli utenti di generare contenuti da prompt di testo**.

Una prima integrazione è in Outlook, il suo servizio di posta elettronica, con uno strumento che aiuta i venditori a scrivere e-mail personalizzate.

In precedenza, Microsoft ha integrato un sistema OpenAI in uno dei suoi prodotti: **DALL-E**, lo strumento text-to-image utilizzato come base per Microsoft Designer, una piattaforma di progettazione grafica text-to-image AI destinata a rivaleggiare con Canva.

Anche **CoPilot**, basato su GPT-3, è uno strumento Microsoft e serve per assistere nella programmazione.

# 11 COME SI È ARRIVATI A CHATGPT: LA STORIA TECNOLOGICA

Il successo di OpenAI non è arrivato dal nulla. Il chatbot è l'iterazione più raffinata di una serie di modelli linguistici di grandi dimensioni che risalgono ad anni fa. Ecco come siamo arrivati a questo punto.

## 11.1 ANNI '80-'90: RETI NEURALI RICORRENTI

ChatGPT è una versione di GPT-3, un modello linguistico di grandi dimensioni sviluppato da OpenAI. I modelli linguistici sono un tipo di rete neurale che è stata addestrata su moltissimi testi. (Le reti neurali sono software ispirati al modo in cui i neuroni nel cervello degli animali si segnalano a vicenda). Poiché il testo è composto da sequenze di lettere e parole di lunghezza variabile, i modelli linguistici richiedono un tipo di rete neurale in grado di dare un senso a questo tipo di dati. Le reti neurali ricorrenti, inventate negli anni '80, possono gestire sequenze di parole, ma sono lente da addestrare e possono dimenticare le parole precedenti in una sequenza.

Nel 1997, gli informatici Sepp Hochreiter e Jürgen Schmidhuber hanno risolto questo problema inventando le reti **LTSM (Long Short-Term Memory)**, reti neurali ricorrenti con componenti speciali che consentono di conservare più a lungo i dati passati di una sequenza di input. Le LTSM potevano gestire stringhe di testo lunghe diverse centinaia di parole, ma le loro capacità linguistiche erano limitate.

## 11.2
## 2017: I TRASFORMATORI (TRANSFORMER)

La svolta che ha portato all'attuale generazione di modelli linguistici di grandi dimensioni è arrivata quando un team di ricercatori di Google ha inventato i trasformatori, un tipo di rete neurale in grado di tracciare la posizione di ogni parola o frase in una sequenza. Il significato di una parola dipende spesso dal significato di altre parole che la precedono o la seguono. Tenendo traccia di queste informazioni contestuali, i trasformatori possono gestire stringhe di testo più lunghe e catturare il significato delle parole con maggiore precisione.

## 11.3
## 2018-2019: GPT E GPT-2

I primi due grandi modelli linguistici di OpenAI sono arrivati a pochi mesi di distanza l'uno dall'altro. L'azienda vuole sviluppare un'intelligenza artificiale polivalente e generica e ritiene che i modelli linguistici di grandi dimensioni siano un passo fondamentale verso questo obiettivo. **GPT** (acronimo di **Generative Pre-trained Transformer**) ha fatto scuola, battendo i benchmark più avanzati dell'epoca per l'elaborazione del linguaggio naturale.

GPT combinava i trasformatori con l'apprendimento non supervisionato, un modo per addestrare i modelli di apprendimento automatico su dati (in questo caso, moltissimi testi) che non sono stati annotati in precedenza. In questo modo il software è in grado di individuare da solo gli schemi nei dati, senza che gli venga detto cosa sta guardando. Molti successi precedenti nel campo dell'apprendimento automatico si sono basati sull'apprendimento supervisionato e su dati annotati, ma l'etichettatura manuale dei dati è un lavoro lento che limita le dimensioni dei set di dati disponibili per l'addestramento.

Ma è stato il GPT-2 a creare il maggior scalpore. OpenAI ha dichiarato di essere così preoccupata che le persone possano usare GPT-2 "per generare un linguaggio ingannevole, distorto o abusivo" che non avrebbe rilasciato il modello completo. Come cambiano i tempi.

## 11.4
## 2020:GPT-3

Il GPT-2 era notevole, ma il seguito di OpenAI, il GPT-3, ha colpito davvero la comunità scientifica.

La sua capacità di generare testi simili a quelli umani ha rappresentato un grande balzo in avanti. GPT-3 può rispondere a domande, riassumere documenti, generare storie in diversi stili, tradurre tra inglese, francese, spagnolo e giapponese e molto altro ancora. Il suo mimetismo è sorprendente.

Uno dei risultati più notevoli è che i guadagni di GPT-3 sono stati ottenuti grazie al sovradimensionamento di tecniche esistenti piuttosto che all'invenzione di nuove. **GPT-3 ha 175 miliardi di parametri** (i valori di una rete che vengono regolati durante l'addestramento), rispetto agli 1,5 miliardi di GPT-2. Inoltre, è stato addestrato su un numero molto maggiore di reti. Inoltre, è stato addestrato su un numero molto maggiore di dati.

Ma l'addestramento su testi presi da Internet comporta nuovi problemi. GPT-3 ha assorbito gran parte della disinformazione e dei pregiudizi che ha trovato online e li ha riprodotti su richiesta. Come ha riconosciuto OpenAI: "I modelli addestrati su Internet hanno pregiudizi su scala Internet".

## 11.5
## DICEMBRE 2020: TESTO TOSSICO E ALTRI PROBLEMI

Mentre OpenAI era alle prese con i pregiudizi di GPT-3, il resto del mondo tecnologico stava affrontando una resa dei conti di alto profilo sull'incapacità di contenere le **tendenze tossiche dell'AI**. Non è un segreto che i modelli linguistici di grandi dimensioni possano emettere testi falsi o addirittura odiosi, ma i ricercatori hanno scoperto che la risoluzione del problema non è sulla lista delle cose da fare per la maggior parte delle aziende Big Tech. Quando Timnit Gebru, co-direttore del team di etica dell'AI di Google, è stato coautore di un documento che evidenziava i potenziali danni associati ai modelli linguistici di grandi dimensioni (tra cui gli elevati costi di calcolo), non è stato accolto con favore dai dirigenti dell'azienda. Nel dicembre 2020, Gebru è stata allontanata dal suo posto di lavoro.

## 11.6
## GENNAIO 2022: INSTRUCTGPT

OpenAI ha cercato di ridurre la quantità di disinformazione e di testo offensivo prodotto da GPT-3 utilizzando l'apprendimento per rinforzo per addestrare una versione del modello sulle preferenze dei tester umani. Il risultato, **InstructGPT**, è stato migliore nel seguire le istruzioni delle persone che lo utilizzano – noto come "allineamento" nel gergo dell'intelligenza artificiale – e ha prodotto meno linguaggio offensivo, meno disinformazione e meno errori in generale.

## 11.7
## MAGGIO-LUGLIO 2022: OPT, BLOOM

Una critica comune ai modelli linguistici di grandi dimensioni è che il costo del loro addestramento rende difficile per tutti i laboratori, tranne quelli più ricchi, costruirne uno. Ciò solleva il timore che un'intelligenza artificiale così potente venga costruita da piccoli team aziendali a porte chiuse, senza un adeguato controllo e senza il contributo di una comunità di ricerca più ampia. In risposta, una manciata di progetti collaborativi ha sviluppato modelli linguistici di grandi dimensioni e li ha rilasciati gratuitamente a tutti i ricercatori che vogliono studiare e migliorare la tecnologia. Meta ha costruito e regalato OPT, una ricostruzione di GPT-3. Hugging Face ha guidato un consorzio di circa 1.000 ricercatori volontari per costruire e rilasciare BLOOM.

## 11.8
## DICEMBRE 2022: CHATGPT

Persino OpenAI è sbalordita dal modo in cui ChatGPT è stato accolto. Nella prima dimostrazione che l'azienda mi ha fornito il giorno prima del lancio online, ChatGPT è stato presentato come un aggiornamento incrementale di InstructGPT. Come quel modello, ChatGPT è stato addestrato usando l'apprendimento per rinforzo sul feedback di tester umani che hanno valutato le sue prestazioni come interlocutore fluido, accurato e inoffensivo. In effetti, OpenAI ha addestrato GPT-3 a padroneggiare il gioco della conversazione e ha invitato tutti a "giocare".

NETWORK DIGITAL 360

Network Digital360 è il più grande network in Italia di testate e portali B2b dedicati ai temi della Trasformazione Digitale e dell'Innovazione Imprenditoriale, con oltre 50 fra portali, canali e newsletter.
Ha la missione di diffondere la cultura digitale e imprenditoriale nelle imprese e pubbliche amministrazioni italiane e di fornire a tutti i decisori che devono valutare investimenti tecnologici informazioni aggiornate e approfondite.
Il Network è parte integrante di Digital360HUB, il polo di Demand Generation di Digital360, che mette a disposizione delle tech company un'ampia gamma di servizi di comunicazione, storytelling, pr, content marketing, marketing automation, inbound marketing, lead generation, eventi e webinar.

VIA COPERNICO, 38
20125 - MILANO
TEL. 02 92852785
MAIL: MARKETING@DIGITAL4.BIZ