Gemini AI: l'avanguardia della nuova conoscenza. Ma con quali interrogativi?

di Raffaella Arcangeli e Nicola Diana - pubblicato su "www.irpa.eu" - Osservatorio sullo Stato digitale, 2 ottobre 2024

Gemini, il nuovo modello di intelligenza artificiale sviluppato a Montainview (Google), promette una rivoluzione innovativa in svariati settori, dalla medicina all'ingegneria, dalla finanza al marketing, fino all'uso quotidiano per imprese e individui. Integra le funzionalità del motore di ricerca con l'assistenza, offrendo una prospettiva creativa. La sua capacità multimodale consente di comprendere e distinguere contesti, garantendo affidabilità nei risultati e nei contenuti. L'articolo esplora le potenzialità, le applicazioni e le caratteristiche tecniche di Gemini, oltre a considerare le criticità tecniche e le principali problematiche giuridiche, soprattutto in relazione alla normativa europea sull'intelligenza artificiale, il trattamento dei dati e la privacy.

I *Large Language Models* (LLM) sono modelli di intelligenza artificiale che utilizzano l'apprendimento automatico per capire e generare linguaggio umano, impiegando reti neurali e tecniche di elaborazione del linguaggio naturale**.** Tra questi strumenti deve sicuramente essere annoverato Gemini, sviluppato da Google AI, con capacità avanzate in diverse aree.

Dopo il successo di Mistral e di ChatGPT, Google annuncia il lancio di Bard, una *chatbot* di intelligenza artificiale generativa, che ha cambiato il nome in Gemini, presentando diverse novità. Attualmente è disponibile in oltre 40 lingue e in 230 paesi con: **Gemini 1.5 Pro**, la versione più recente e avanzata, capace di elaborare grandi quantità di dati con multimodalità (in fase di test). **Gemini 1.0** ha invece tre dimensioni di modello (**Ultra**, ideale per attività altamente complesse; **Pro**, con le migliori prestazioni per una vasta gamma di attività; **Nano**, ottimizzato per le attività su dispositivi con risorse limitate). Infine, **Gemini Advance** che integra l'AI con i servizi offerti da Google alle attività professionali.

Grazie alla sua avanzata capacità di relazionarsi con il motore di ricerca, Gemini ha la possibilità di valutare un ampio panorama informativo con **dati precisi e pertinenti**. Attraverso ricerche mirate esplora una varietà di fonti *online*, dai blog settoriali fino a raggiungere le pubblicazioni accademiche, con un flusso costante di **informazioni attendibili ed approfondite**. La **multimodalità** permette di integrare e **analizzare** dati provenienti da **diverse fonti** (grafici, testi e analisi dei sentimenti). Questa competenza gli permette di comprendere in maniera approfondita, facilitando un'interpretazione olistica dei contenuti analizzati. La conseguenza è l'attitudine di Gemini a generare testo di qualità e risposte ad un'ampia gamma di richieste, basandosi su fatti e prove concrete, adattando il proprio **linguaggio alla situazione**. Può scrivere diversi **formati creativi** (poesie, codici, brani, mail, ecc.), può **tradurre** accuratamente, fornite **riassunti cogliendo il contesto**. È in grado di **elaborare file**, **produrre documenti finanziari**, riconoscendo gli elementi, ragionare in modo logico e analitico (*problem solving*) e sviluppare **prodotti**, **generando idee e testando concetti.**

Nell'ambito della distinzione tra modelli aperti e chiusi Gemini, si presenta come **modello tendenzialmente chiuso**. Questo significa che il suo codice sorgente non è pubblicamente

disponibile. Il suo utilizzo è soggetto a restrizioni, anche se sono presenti alcuni elementi di apertura, come l'API (interfaccia di programmazione dell'applicazione) e la documentazione che potrebbe essere pubblicata in merito al suo utilizzo. Inoltre, **Google ha presentato Gemma**, una raccolta di modelli AI leggeri *open source* **rivolta solamente agli sviluppatori**. Gemini utilizza **un'architettura *Transformer* con *encoder-decoder*** che gli consente di elaborare sequenze di parole, comprendere relazioni contestuali e generare *output* coerenti. L'*encoder* analizza l'*input*, catturando le relazioni semantiche, mentre il *decoder* utilizza queste informazioni per **generare *output* coerenti** con il contesto. Un meccanismo di attenzione avanzato permette di focalizzarsi su parti specifiche dell'*input* durante la decodifica. Integra anche **reti neurali convoluzionali per migliorare l'identificazione di modelli** e le relazioni nel linguaggio. **Gemini 1.5 Pro**, può elaborare fino a **1 milione di *token*** e utilizza un'architettura ***Mixture-of-Experts***), che consente al modello di attivare selettivamente percorsi esperti specifici nella rete neurale in base al tipo di *input*.

Gemini viene addestrato principalmente utilizzando **l'apprendimento supervisionato**, dove viene fornito un vasto *dataset*. Il modello impara ad associare correttamente l'*input* all'*output*, migliorando la sua capacità di generare un testo accurato e pertinente. Gemini sfrutta anche l'apprendimento **semi- supervisionato**, dove solo una parte dei dati di addestramento è etichettata. Il modello utilizza le informazioni etichettate per apprendere modelli generali e poi le applica ai dati non etichettati per completarne la classificazione. L**'apprendimento rinforzato** viene utilizzato per ottimizzare le prestazioni di Gemini attraverso i *feedback* positivi o negativi, imparando gradualmente.

Le *Large Language Models* mostrano comunque diverse problematiche. Tra queste, le **allucinazioni** rappresentano una sfida significativa. Se il modello viene addestrato su *dataset* con informazioni errate o incomplete o non è sufficientemente ampio, produrrà *output* che con **distorsioni, errori** o **lacune.**

Un'altra preoccupazione riguarda la **violazione della proprietà intellettuale** sia nell'utilizzo delle opere durante il processo di apprendimento, sia nei risultati prodotti dall'intelligenza artificiale stessa, suscitando dubbi su chi abbia effettivamente diritto a tali risultati. Considerando che OpenAI avrebbe impiegato milioni di ore di video di YouTube per addestrare GPT-4, con noti problemi di violazione del *copyright*, è lecito immaginare la vastità dei dati che Google potrebbe potenzialmente utilizzare.

Il Regolamento europeo sull'intelligenza artificiale (**AI Act**), stabilisce un quadro giuridico per la regolamentazione il settore (ne abbiamo parlato QUI), con l'obiettivo di **promuovere lo sviluppo** e l'utilizzo dell'intelligenza artificiale, **garantendo il rispetto dei diritti e la sicurezza dei cittadini**. I sistemi di intelligenza artificiale sono classificati in **quattro categorie di rischio**:

- **rischio minimo,** (calcolatrici o videogiochi)
- **rischio basso (**Chatbot per il servizio clienti, antispam)
- **rischio elevato** (contesti medici o finanziari)
- **rischio inaccettabile** (sorveglianza di massa o la manipolazione del comportamento umano).

L'intelligenza artificiale generativa, come Gemini, **non dovrebbero essere classificate in modo aprioristico** e automatico come ad alto rischio, ma dovranno rispettare i requisiti di trasparenza e la legge sul *copyright* dell'UE, dovendo quindi divulgare che il contenuto è stato

generato dall'intelligenza artificiale e progettare il modello per evitare che generi contenuti illegali e fornire informazioni dei dati protetti da copyright utilizzati per la formazione. I modelli di AI ad alto impatto che potrebbero comportare un rischio sistemico, come il modello di intelligenza artificiale più avanzato GPT-4 o Gemini 1.5, dovrebbero essere sottoposti a valutazioni approfondite e qualsiasi incidente grave dovrebbe essere segnalato alla Commissione europea.

Infine, Gemini suscita preoccupazioni sulla **privacy** (ne abbiamo parlato QUI**)** poiché raccoglie, elabora e utilizza i dati degli utenti. Google registra le conversazioni e acquisisce informazioni sulla posizione, sui servizi utilizzati e sui *feedback* forniti dagli utenti. È sconsigliata la condivisione di informazioni personali per impedire l'accesso ai dati da parte di revisori umani. Anche se Google assicura che le conversazioni sono separate dall'account dell'utente e conservate per massimo tre anni, il mantenimento di enormi quantità di dati solleva preoccupazioni, anche per quanto riguarda la navigazione in incognito.

Prima che i sistemi di AI possano essere affidabili su vasta scala, ci sono ostacoli significativi da superare. Le **sfide come le allucinazioni** ostacolano l'adozione diffusa dei grandi modelli linguistici in applicazioni ad alto rischio. È essenziale che i ricercatori si concentrino su queste aree critiche, **introducendo benchmark più realistici** (come *HaluEval* e *SWE-bench*) per garantire miglioramenti tangibili nell'IA. Le sfide non sono solo legate alle leggi esistenti che si adattano alle nuove tecnologie, ma si evolvono costantemente. Oltre al rischio teorico considerato dagli esperti, sorge anche il rischio pratico dell'abuso tecnologico, come la violazione dei principi e delle norme già stabilite. Questo evidenzia la necessità non solo di regolamenti preventivi, ma anche di **sistemi di controllo e monitoraggio** (come quelli per la tutela del diritto d'autore e della privacy). Infatti, solo affrontando direttamente queste sfide, l'AI può aspirare a diventare affidabile e adottata su vasta scala.