Mini dizionario sull'AI: prompt, allucinazioni e effetto Barnum

RICERCA AVANZATA Submit

I recenti progressi nell'iter di approvazione dell'AI Act rendono di grande attualità termini o espressioni poco conosciute come "prompt", "allucinazioni" o "effetto Barnum". Il presente contributo prova a fornire una breve definizione di tali concetti.

di
La Redazione

SOMMARIO

Un **prompt** è un testo in **linguaggio naturale** che ha lo scopo di descrivere il compito, o i compiti, che un sistema di intelligenza artificiale deve eseguire. L'importanza del prompt deriva dal fatto che un sistema di AI può fornire risposte (c.d. output) diverse, a seconda di come esso viene formulato. Il prompt, infatti, può essere una semplice parola, una frase, o un paragrafo strutturato che fornisce al sistema informazioni sufficienti per poter generare un output pertinente.

Quali sono le caratteristiche principali che un buon prompt deve avere?

Per ottenere risultati migliori un prompt deve avere le seguenti caratteristiche:

1. **Contenere istruzioni chiare, specifiche, complete e dettagliate**. A tal fine, è consigliabile chiedere al modello di immedesimarsi in un professionista o adottare lo stile di un personaggio, specificare la lunghezza dell'output che si vuole ottenere e i passaggi necessari per completare l'attività richiesta.
2. **Fornire un testo di riferimento**. Uno dei problemi principali dell'AI, è la sua capacità di inventare e fornire risposte errate o fuorvianti presentandole come vere (c.d. allucinazioni). Per minimizzare questo rischio, è consigliabile fornire un testo di riferimento da utilizzare come spunto per circoscrivere il "campo" di lavoro dell'AI. Un testo di riferimento può essere un URL o una porzione di testo che si vuole analizzare. In questo caso, andranno utilizzati dei delimitatori (ad esempio. """ oppure ###).
3. **Suddividere le attività complesse**. Le attività complesse generano più errori rispetto a compiti semplici. È consigliabile, pertanto, segmentare richieste particolarmente complicate o articolate in più attività ricorsive, in modo che l'AI possa eseguire più operazioni distinte ed essere meno incline agli errori.

Un esempio di prompt potrebbe essere il seguente:
Sono un avvocato esperto di privacy.

Analizza la proposta di regolamento dell'AI Act.
Sintetizza per punti principali e individua le novità significative in tema di privacy.
Una volta individuate, indicami i temi che un avvocato esperto dovrà affrontare.

Che cosa sono le allucinazioni e cos'è l'effetto Barnum?

Le allucinazioni sono un fenomeno che si verifica quando un sistema di AI genera un risultato non basato su dati reali, fornendo un output privo di senso o del tutto impreciso.
Le allucinazioni possono essere causate da diversi fattori, tra cui:

- dati di addestramento di scarsa qualità; se un sistema di AI utilizza tali dati è più probabile che generi output poco accurati o non pertinenti,
- utente che fornisce istruzioni imprecise, errate o incomplete.

L'effetto Barnum, invece, è un **fenomeno psicologico** determinato dal fatto che le persone hanno la tendenza a considerare accurata e rilevante una descrizione generica o vaga, ma che sembra specifica e personalizzata ossia dalla propensione a ritenere credibili informazioni che sono congruenti con le loro aspettative. Le allucinazioni delle AI possono essere viste come una forma di effetto Barnum.
La **caratteristica dei sistemi di AI generativa**, infatti, è quella di dare sempre risposte, anche a costo di inventare. Di conseguenza, quando un sistema genera un output che non è basato su dati reali, le persone possono essere indotte a credere che esso sia accurato e pertinente, anche quando non lo è. Ad esempio, se un oroscopo dice che una persona di un dato segno zodiacale è decisa e meticolosa, è più probabile che quest'ultima creda a questa descrizione, anche se non è basata su alcuna informazione reale.
L'effetto Barnum, in sintesi, è particolarmente insidioso per l'utente, perché può facilmente indurlo a prendere decisioni sbagliate o non informate. È quindi importante che egli ricordi che le descrizioni e gli output generici non sono necessariamente accurati o pertinenti.

Quali sono le principali raccomandazioni che un avvocato deve tenere a mente quando utilizza un sistema di AI?

La **commissione Nuove Tecnologie di FBE (Fédération des Barreaux d'Europe)** ha elaborato un documento contenente dei principi etici e professionali che gli avvocati devono rispettare quando interagiscono con un sistema di AI.
Seguire tali raccomandazioni aiuta l'avvocato a evitare o riconoscere allucinazioni e a scongiurare l'**effetto Barnum**.
Essi sono sintetizzabili nei seguenti punti:

1. comprendere l'IA generativa e il suo funzionamento;
2. riconoscere le limitazioni e il contesto in cui i sistemi di AI operano;
3. rispettare le regole esistenti sull'uso della IA;
4. integrare gli output forniti con la competenza giuridica;
5. rispettare il segreto professionale;
6. garantire la protezione dei dati personali e della privacy;
7. informare il cliente in caso di utilizzo di sistemi di AI e assumersi la responsabilità.