



Le allucinazioni dell'IA

COSA SONO LE ALLUCINAZIONI DELL'INTELLIGENZA ARTIFICIALE?

Immagina di chiedere a un amico un'informazione e lui, invece di cercare la risposta, te la inventa. Ecco, questo è in sostanza ciò che accade con le allucinazioni dell'intelligenza artificiale.

Un modello di linguaggio è addestrato su enormi quantità di dati. Impara a riconoscere pattern e a prevedere la parola successiva in una frase. Ma a volte, soprattutto quando si trova di fronte a domande complesse o poco chiare, può "inventare" informazioni che sembrano plausibili ma non sono corrette.

Questo fenomeno viene chiamato "allucinazione".

PERCHÉ AVVENGONO LE ALLUCINAZIONI?

Ci sono diverse ragioni per cui un'IA può "allucinare":

- **dati di addestramento incompleti o errati:** se i dati su cui viene addestrato il modello contengono errori o sono incompleti, l'IA potrebbe apprendere informazioni false;
- **ambiguità della domanda:** se la domanda formulata nel prompt è troppo vaga o aperta a interpretazioni multiple, l'IA potrebbe generare risposte imprecise o addirittura errate;
- **complessi modelli linguistici:** i modelli di linguaggio più avanzati sono estremamente complessi e possono generare risposte creative e originali, ma a volte questo può portare a risultati inaspettati e imprecisi.

Esempi di allucinazioni

- **Fatti inventati:** l'IA potrebbe affermare che un evento storico si è verificato in un anno diverso da quello reale.
- **Citazioni false:** potrebbe attribuire una citazione a una persona sbagliata o inventare completamente la citazione.
- **Invenzione di prodotti o servizi:** l'IA potrebbe descrivere un prodotto o un servizio che non esiste.

COME SI POSSONO PREVENIRE LE ALLUCINAZIONI?

- **Miglioramento dei dati di addestramento:** è fondamentale utilizzare dati di alta qualità e accurati per addestrare i modelli, come dati accademici, scientifici e coperti da diritti d'autore.
- **Chiarificazione delle domande:** è importante formulare le domande in modo chiaro e preciso, evitando ambiguità, usando tecniche di ingegneria del prompt.
- **Valutazione critica dei risultati:** è sempre consigliabile verificare le informazioni fornite dall'IA con altre fonti e sviluppare quindi il pensiero critico.