ARTICOLO DI ELEARNINGNEWS

Anno 9 - numero 351 di Mercoledì 06 ottobre 2025

Allucinazioni dell'intelligenza artificiale: i rischi per l'eLearning

In questo articolo cerchiamo di capire se le allucinazioni dell'IA saranno la prossima frontiera della disinformazione

L'intelligenza artificiale (IA), in particolare i modelli linguistici di grandi dimensioni (LLM), è diventata una componente molto importante delle esperienze digitali moderne. Dai chatbot e dagli assistenti virtuali ai sistemi avanzati di eLearning, le tecnologie di IA promettono efficienza, scalabilità e interazioni personalizzate. Tuttavia, uno dei limiti più preoccupanti di questi sistemi è la loro tendenza a produrre le cosiddette "**allucinazioni**", ovvero **output inventati, fuorvianti o errati, ma presentati con apparente sicurezza**.

Nel contesto dell'istruzione e della formazione, tali errori possono avere conseguenze profonde. Comprendere perché si verificano le allucinazioni, come si manifestano e i loro potenziali rischi è fondamentale, soprattutto perché le piattaforme basate sull'IA continuano a plasmare il modo in cui viene trasmessa la conoscenza.

Cosa sono le allucinazioni dell'IA?

Le **allucinazioni dell'IA** si verificano quando un sistema di IA genera contenuti che sembrano logici o fattuali, ma che in realtà sono inaccurati, inesistenti o inventati. A differenza dei semplici errori, le allucinazioni sono problematiche perché di solito vengono fornite in modo fluido e autorevole, rendendo difficile per gli utenti distinguerle dalle informazioni legittime.

Ad esempio, un'IA potrebbe:

- Fornire citazioni di articoli di ricerca che non esistono.
- Generare soluzioni passo dopo passo che contengono errori logici critici.
- Offrire consigli medici, legali o tecnici con sicurezza, ma privi di fondamento nella realtà.

Perché si verificano le allucinazioni dell'IA?

Le allucinazioni derivano dal modo in cui gli LLM vengono addestrati e dal loro funzionamento. I fattori chiave includono:

Apprendimento basato su modelli senza una reale comprensione

Gli LLM come GPT vengono addestrati su grandi quantità di dati testuali e generano risposte prevedendo la sequenza di parole più probabile. Non "conoscono" i fatti, ma riconoscono i modelli. Quando si trovano di fronte a lacune nelle conoscenze, possono costruire output plausibili ma falsi.

Limiti dei dati

Se il modello non è stato esposto a dati specifici e affidabili su un argomento, può "riempire gli spazi vuoti" con dettagli inventati che si adattano ai modelli linguistici.

Prompt ambigui

L'ambiguità o le domande formulate in modo errato possono indurre l'IA a generare risposte speculative o imprecise, che spesso appaiono come allucinazioni.

Pregiudizio di eccessiva sicurezza nella generazione

I modelli sono progettati per generare testi fluenti e dal suono naturale: questa "falsa sicurezza" stilistica può mascherare le imprecisioni, dando l'illusione della credibilità.

Complessità delle conoscenze specifiche di un determinato settore

In campi specializzati come la medicina, il diritto o l'ingegneria, una piccola imprecisione può trasformarsi in un errore grave. I sistemi di IA che non hanno accesso a database verificati di un determinato settore sono particolarmente inclini alle allucinazioni in tali ambiti.

Troppi contenuti online generati dall'IA

La presenza online in forte aumento di contenuti generati dall'intelligenza artificiale è problematica perché questi contenuti vengono presi dai sistemi LLM come fonti per generare nuovi output. Se questi dati sono errati, anche solo in parte, diventano le basi per la creazione delle allucinazioni dell'IA.

Implicazioni e rischi delle allucinazioni dell'IA

Le allucinazioni introducono sfide significative in termini di fiducia, sicurezza e affidabilità:

Erosione della fiducia

una volta che gli utenti rilevano informazioni false, potrebbero perdere fiducia nel sistema, anche se la maggior parte dei risultati è accurata.

Rischi legali ed etici

La disinformazione in settori come quello legale o medico può esporre le organizzazioni a responsabilità civili se gli utenti agiscono sulla base di allucinazioni generate dall'IA.

Diffusione di disinformazione

Le falsità generate dall'IA possono propagarsi rapidamente online, amplificando il problema più ampio della disinformazione.

Rafforzamento dei pregiudizi

Le allucinazioni possono riflettere e rafforzare i pregiudizi incorporati nei dati di addestramento, perpetuando stereotipi o inesattezze.

L'impatto sull'eLearning e sui sistemi di formazione basati sull'intelligenza artificiale

Vediamo ora l'impatto delle allucinazioni dell'IA sul **settore della formazione**. L'ascesa delle piattaforme basate sull'intelligenza artificiale nel campo dell'istruzione ha trasformato radicalmente il modo in cui le persone apprendono, si esercitano e ricevono feedback. Con significative ripercussioni sul modo in cui le scuole e i reparti di formazione aziendale elargiscono la formazione. Nuove tecnologie come i sistemi di tutoraggio personalizzati, i quiz adattivi, la creazione automatizzata di contenuti e i feedback loop basati sull'intelligenza artificiale promettono di ampliare l'istruzione e renderla più accessibile. Tuttavia, la presenza di allucinazioni dell'intelligenza artificiale introduce diversi rischi.

1. Risultati di apprendimento compromessi

Quando i sistemi di IA presentano **spiegazioni o esempi imprecisi**, gli studenti possono assorbire e ripetere informazioni errate senza rendersene conto. A

2. Integrità della valutazione

Molte piattaforme di eLearning utilizzano l'IA non solo per fornire lezioni, ma anche per **valutare le prestazioni degli studenti**. Se uno strumento di valutazione basato sull'IA interpreta erroneamente il compito di uno studente, rischia di fornire voti o feedback fuorvianti. Ciò compromette l'equità della valutazione e potrebbe scoraggiare gli studenti che ritengono che i loro sforzi siano stati giudicati in modo errato.

3. Rischi nella formazione professionale

Oltre all'istruzione accademica, l'IA è sempre più utilizzata nella **formazione professionale**, ad esempio nei programmi di conformità, nelle simulazioni mediche o nell'aggiornamento professionale della forza lavoro. In questi contesti, un'allucinazione può comportare rischi immediati nel mondo reale. Ad esempio, uno strumento di formazione sulla sicurezza basato sull'intelligenza artificiale per i lavoratori edili potrebbe generare una procedura errata per la manipolazione di materiali pericolosi o indicare in modo errato l'uso corretto dei dispositivi di protezione. Se i dipendenti seguissero queste istruzioni, potrebbero verificarsi incidenti sul lavoro, infortuni o persino violazioni normative. In settori ad alto rischio come l'aviazione, la sanità o la sicurezza informatica, i rischi di disinformazione si moltiplicano.