



COMUNICATO STAMPA SPONSORIZZATO • Non solo eccitazione per l'IA nel campo della compliance, ora anche risultati • con progressi evidenti nei modelli di ultima generazione

Descrizione

(Immediapress) •

MONACO DI BAVIERA, GERMANIA • EQS Newswire • 20 ottobre 2025 • L'intelligenza artificiale sta entrando rapidamente a far parte dei flussi di lavoro aziendali • ma non tutti i modelli danno gli stessi risultati. Per valutare in che misura l'IA possa essere applicata nella realtà della compliance, il nuovo •EQS Benchmark Report: AI Performance in Compliance & Ethics• ha testato sei tra i principali modelli di IA con 120 scenari di compliance del mondo reale • da valutazioni di rischi e conflitti di interessi ad audit di terze parti. I risultati evidenziano una precisione quasi perfetta con tassi di accuratezza superiori al 95% nelle operazioni strutturate, come classificazione e processo decisionale, ma un notevole peggioramento in caso di maggiore complessità o ambiguità. Realizzato in collaborazione con l'associazione tedesca Berufsverband der Compliance Manager e.V. (BCM), nel benchmark risulta chiaramente anche il ritmo serrato dello sviluppo, con una performance significativamente superiore dei modelli del 2025 rispetto a quelli del 2024.

• Per molti responsabili della conformità, l'IA continua a rimanere un territorio poco familiare, • spiega Moritz Homann, Director of Product Innovation and AI di EQS Group. • Capire come applicarla in modo efficace e quando ce se possa fidare può essere difficile • soprattutto in un ambito sensibile come la compliance, dove non ci sono compromessi in termini di accuratezza, trasparenza e integrità. •

• Nel campo della compliance l'IA può offrire nuovi livelli di comprensione, ma la nostra responsabilità • assicurare che il suo impiego rimanga soggetto a chiari limiti etici e legali, • sottolinea Dr. Gisa Ortwein, la Presidente di BCM. • Iniziative come questa valutazione ci aiutano a distinguere tra cosa possa davvero fornire l'IA e i casi in cui invece il giudizio umano •

insostituibile. In questo modo salvaguardiamo l'integrità accogliendo l'innovazione e assicuriamo che l'adozione dell'IA faciliti la nostra professione, anziché metterla a rischio.

Il benchmark EQS è il primo a valutare l'IA nel campo della compliance, utilizzando operazioni che riflettono le responsabilità quotidiane dei professionisti di compliance ed etica. Esso misura accuratezza del modello, affidabilità e utilità pratica in prove strutturate, semistrutturate e aperte.

La performance dei modelli più recenti è molto superiore a quella dei modelli usciti solo pochi mesi prima

Gli esiti della valutazione sottolineano la rapidità di evoluzione delle capacità dei modelli. Gemini 2.5 Pro di Google ha raggiunto il punteggio complessivo più alto di 86,7%, mostrando una performance valida in tutti i tipi di operazioni e di aspetti della compliance. Con un punteggio complessivo dell'86,5%, GPT-5 di OpenAI (il modello standard di ChatGPT dall'agosto 2025) è simile a Gemini nella maggior parte delle categorie, evidenziando una rapida convergenza delle capacità dei migliori modelli. GPT-5 ha registrato una performance particolarmente buona nella creazione di contenuti aperti, mentre Gemini si è dimostrato superiore in complessi compiti di analisi e in processi decisionali.

Segue o3 di OpenAI con una performance dell'83,3% che mostra chiaramente il progresso di GPT-5 rispetto al modello precedente e anche il rapido ciclo iterativo caratteristico di questo settore. Claude Opus 4.1 di Anthropic ha raggiunto un punteggio dell'81,5%, con una performance inferiore nelle analisi strutturate e nel ragionamento analitico, mentre all'ultimo posto si trovano GPT-4o (72,9%) e Mistral Large 2 (70,1%). I risultati riflettono il significativo salto generazionale tra i modelli lanciati nel 2024 e quelli del 2025.

Nella compliance l'IA eccelle se le regole sono chiare, ma incontra problemi in caso di ambiguità

In linea generale i modelli di IA hanno dato i risultati migliori in compiti di compliance strutturati e lineari. Ad esempio, la performance media è stata del 90,8% in scenari di processi decisionali basati su una situazione definita e una serie di norme o regolamenti. In compiti che richiedevano di confrontare o stabilire corrispondenze tra set di dati, i modelli hanno raggiunto un punteggio medio del 91,8% e quattro dei sei modelli hanno superato il 95%.

I diversi modelli hanno invece fatto registrare performance molto differenziate nelle operazioni più complesse. Il divario si è dimostrato particolarmente grande nelle prove che richiedevano l'analisi di dati, con una differenza di 60 punti tra i risultati migliori e quelli peggiori. In questa categoria, Gemini

2.5 Pro ha raggiunto un punteggio dell'88%, seguito da GPT-5 con il 62% mentre GPT-4o si è classificata all'ultimo posto con soltanto il 28%.

Le prove aperte, come la redazione di briefing esecutivi o rapporti su indagini interne, si sono rivelate piuttosto difficili anche per i modelli più recenti. I risultati migliori in questa categoria sono stati raggiunti da GPT-5 con un punteggio del 67,4%. A differenza delle prove strutturate, questi compiti sono stati valutati da una giuria composta di persone.

Ci sono alcuni compiti ad alto rischio che i professionisti della compliance non ritengono opportuno affidare completamente all'IA, e non dovrebbero nemmeno farlo, commenta Moritz Homann. L'efficacia degli strumenti basati sull'IA risiede nell'effetto di moltiplicatore di forze, con supporto dei flussi di lavoro della compliance lasciando ai professionisti per la responsabilità ultima e il giudizio. L'IA può assumersi una buona parte del lavoro preparatorio, anche nel caso di casi altamente complessi. Permette così di risparmiare tempo prezioso nelle operazioni preliminari di routine e consente agli esperti di concentrarsi sui punti per i quali il loro giudizio è indispensabile.

Elevata coerenza e basso tasso di allucinazioni

Il benchmark ha testato anche l'affidabilità facendo ripetere a ogni modello per tre volte i compiti a scelta multipla. La coerenza è stata elevata, e la maggior parte dei modelli ha fornito lo stesso risultato in oltre il 95% dei casi. Le allucinazioni, uno dei rischi più criticati dell'IA, sono state rare: in tutte le prove e i modelli ne sono stati registrati solo tre casi ovvi, con un tasso pari soltanto allo 0,71%. Ciò significa che, se i compiti vengono definiti e contestualizzati chiaramente, i modelli attuali sono in grado di fornire risultati stabili e basati sui fatti negli scenari di compliance. Tuttavia non è possibile escludere completamente le allucinazioni, pertanto la supervisione umana rimane essenziale e specialmente nel caso di contenuti sensibili con implicazioni normative.

Selezione dei modelli e conseguenze dell'influsso della formulazione dei prompt

Il rapporto ha evidenziato anche l'importanza della specificità dei prompt. Nei compiti nei quali ai modelli di IA è stato richiesto di individuare red flag dai dati di audit di terze parti, i risultati sono stati diversi a seconda dell'accuratezza nella formulazione della domanda, ad esempio esplicitando se includere entità affiliate o valutare la gravità delle informazioni trovate. I modelli più recenti come GPT-5 e Gemini 2.5 Pro hanno mostrato una maggiore capacità di seguire istruzioni complesse e fornire risultati strutturati, con vantaggi chiari per i team di compliance che operano con regolamenti diversificati e set di dati estesi.

Moritz Homann: «L'IA è destinata ad affermarsi e il modo in cui la implementiamo e utilizziamo oggi sarà determinante per il suo ruolo nell'ambito della compliance negli anni futuri. I team di compliance ed etica non dovrebbero solo tenere sotto controllo i rischi dell'IA, ma anche applicare loro stessi questa tecnologia. Solo impiegando l'IA nella pratica possiamo imparare a fare le domande giuste, definire limiti di sicurezza efficienti e creare fiducia. Il nostro obiettivo è sostenere questo processo con strumenti pratici, trasparenza e dialogo.»

La versione integrale del Benchmark Report di EQS sull'IA può essere scaricata dal sito: <https://www.eqs.com/compliance-wpapers/ai-performance-compliance-ethics-eqs/>

Metodologia

Il Benchmark Report di EQS sull'IA ha testato sei modelli linguistici di grandi dimensioni, e cioè GPT-5 di OpenAI, GPT-4o e o3, Gemini 2.5 Pro di Google, Claude Opus 4.1 di Anthropic e Mistral Large 2, con 120 operazioni che coprivano dieci ambiti essenziali della compliance. Erano incluse tematiche quali la valutazione dei rischi, la verifica di casi di segnalazioni, l'efficienza dell'apprendimento, la valutazione dei regolamenti e l'analisi di lacune normative.

Il set di prove è stato preparato in base a indicazioni di professionisti della compliance e ha compreso sia contenuti del mondo reale che contenuti sintetici come set di dati HR, risultati di apprendimento e testi di regolamenti. Alcuni compiti avevano una risposta oggettivamente corretta, mentre altri richiedevano un approccio più soggettivo e incentrato sulle persone per essere considerati nel punteggio. Per questo motivo i risultati aperti sono stati valutati con il sostegno di Berufsverband der Compliance Manager (BCM), i cui membri hanno contribuito a un'analisi professionale e al feedback sulla qualità e sull'utilità delle risposte generate dai modelli.

Informazioni su EQS Group

EQS Group è uno dei principali fornitori internazionali di cloud per compliance ed etica, protezione dei dati, gestione della sostenibilità e relazioni con gli investitori. Oltre 14.000 aziende di tutto il mondo impiegano i prodotti di EQS Group per creare fiducia grazie al rispetto affidabile e sicuro di requisiti legali complessi, alla riduzione dei rischi e al reporting trasparente sui risultati dell'attività e sul relativo impatto su società e ambiente.

Le soluzioni di EQS Group sono riunite in una piattaforma su base cloud. Questa permette di trattare in modo professionale processi di ottemperanza per la tutela dei whistleblower e la coordinazione dei casi, gestione della propria policy e processi di approvazione come anche nel caso di partner d'affari, terzi e rischi, insider list e obblighi di comunicazione. Inoltre EQS Group fornisce software per rispondere a requisiti di due diligence in termini di diritti umani in tutte le catene di fornitura

aziendali, assicurando il rispetto delle disposizioni di tutela dei dati come il RGPD e il Regolamento europeo sull'intelligenza artificiale, inoltre supporta un'efficiente gestione ESG e un reporting di sostenibilità conforme. Le aziende registrate traggono inoltre vantaggio da una rete globale, un modo mirato di rivolgersi agli investitori e una gestione dei contatti come anche da siti web e webcast IR per una comunicazione efficiente e sicura con gli investitori.

EQS Group Ã stato costituito a Monaco di Baviera nel 2000. Oggi il gruppo conta circa 600 professionisti in tutto il mondo.

<https://www.eqs.com/>

Informazioni su BCM

In qualitÃ di principale associazione di professionisti, solamente per responsabili della conformitÃ interni di aziende, associazioni e altre organizzazioni, BCM rappresenta gli interessi dei propri membri nei rapporti con decisori politici, mondo degli affari e societÃ . BCM si dedica principalmente a fornire informazioni, promuovere di reti e rafforzare la professione di responsabile della conformitÃ . Offre un'ampia gamma di servizi gratuiti studiati per lâaggiornamento dei membri in merito alle attuali tematiche di compliance e per la promozione e lo sviluppo continuo della condivisione di conoscenze all'interno della propria rete.

www.compliance-verband.de

Contatti:

Immediapress

<https://www.eqs.com/>

www.compliance-verband.de

COMUNICATO STAMPA SPONSORIZZATO: Immediapress Ã un servizio di diffusione di comunicati stampa in testo originale redatto direttamente dall'ente che lo emette. L'Adnkronos e Immediapress non sono responsabili per i contenuti dei comunicati trasmessi

â

immediapress

Categoria

1. Comunicati

Tag

1. ImmediaPress

Data di creazione

Ottobre 20, 2025

Autore

redazione

default watermark