IBM annuncia la disponibilità su watsonx del modello di AI open source Mistral, che amplia la scelta di modelli per aiutare le imprese a scalare l’AI con fiducia e flessibilità

• IBM offre una versione ottimizzata di Mixtral-8x7B con comprovate capacità di ridurre la latenza fino al 75%

• Si aggiunge e amplia il portafoglio di modelli IBM, di terze parti e open-source per offrire ai clienti più scelta e flessibilità

• Ultimo modello open-source disponibile sulla piattaforma di AI e dati watsonx con funzionalità specifiche per le imprese in ambito governance, archivio dati e studio AI

ARMONK, N.Y., 29 febbraio 2024 - IBM (NYSE: IBM) ha annunciato oggi la disponibilità del noto LLM (large language model) open-source Mixtral-8x7B, sviluppato da Mistral AI, sulla sua piattaforma di dati e AI watsonx, nell'ottica di un continuo impegno ad ampliare le funzionalità per aiutare i clienti a innovare con i foundation model propri di IBM e quelli di una gamma di provider open source.

IBM offre una versione ottimizzata di Mixtral-8x7b che, nei test interni, è stato in grado di aumentare la velocità di elaborazione dati - o la quantità di dati che può essere elaborata in un determinato periodo di tempo - del 50% rispetto al modello base. Ciò potrebbe potenzialmente ridurre la latenza del 35-75%, a seconda della dimensione del batch, accelerando il tempo per ottenere gli insight. Tale risultato viene raggiunto grazie a un processo noto come quantizzazione, che riduce la dimensione del modello e i requisiti di memoria per gli LLM e, di conseguenza, può accelerare l'elaborazione per contribuire a ridurre i costi e il consumo di energia.

La disponibilità di Mixtral-8x7B amplia la strategia multi-modello e aperta di IBM per andare incontro alle

esigenze dei clienti e offrire loro scelta e flessibilità per eseguire la scalabilità delle soluzioni AI durante l’intero processo. Grazie alla ricerca e allo sviluppo dell'AI sul quale sta lavorando da decenni, alla collaborazione aperta con Meta e Hugging Face e alle partnership con i leader del settore, IBM sta ampliando la propria offerta di soluzioni di AI e il proprio portafoglio di modelli watsonx.ai, introducendo nuove capacità, linguaggi e modalità.

La scelta di IBM di foundation model pronti per le imprese all’interno della piattaforma watsonx AI, permette ai clienti di utilizzare l'AI generativa per ottenere nuovi insight, efficienze e creare nuovi modelli di business basati su principi di affidabilità. IBM consente ai clienti di selezionare il giusto modello per ciascun progetto tenendo conto delle esigenze in termini di prezzo/prestazioni in aree specifiche, come ad esempio in ambito finanza.

Mixtral-8x7B è stato costruito utilizzando la combinazione di modellazione Sparse - una tecnica innovativa che trova e utilizza solo le parti più essenziali dei dati per creare modelli più efficienti - e la tecnica Mixture-of-Experts, che combina diversi modelli ("esperti") in grado di risolvere specifiche problematiche. Il modello Mixtral-8x7B è ben noto per la sua capacità di elaborare e analizzare rapidamente grandi quantità di dati per fornire insight pertinenti al contesto.

"I clienti chiedono scelta e flessibilità per implementare modelli che si adattino al meglio ai loro specifici progetti e requisiti", ha dichiarato **Kareem Yusuf, Ph.D, Senior Vice President, Product Management & Growth, IBM Software.** *"Mixtral-8x7B e altri modelli su watsonx ci permettono di offrire loro più scelta nelle modalità di implementazione dell'AI, ma anche di favorire la creazione di un solido ecosistema di sviluppatori di AI e leader aziendali che promuovano l'innovazione in diversi settori e domini, attraverso strumenti e tecnologie".*

Questa settimana, IBM ha annunciato anche la disponibilità di ELYZA-japanese-Llama-2-7b, un modello LLM giapponese open-source di ELYZA Corporation, su watsonx. IBM offre, all’interno della piattaforma watsonx, anche i modelli open-source di Meta Llama-2-13B-chat e Llama-2-70B-chat e altri di terze parti. Inoltre, nei prossimi mesi è prevista l’introduzione in watsonx di nuovi modelli.

Le dichiarazioni relative all'orientamento e alle intenzioni future di IBM sono soggette a modifica o a ritiro senza preavviso e rappresentano solo scopi e obiettivi.

IBM

IBM è un’azienda leader a livello mondiale nel settore del cloud ibrido, dell’AI e dei servizi alle imprese e opera con le imprese di oltre 175 Paesi aiutandole a capitalizzare sugli insight dei loro dati, a semplificare i processi aziendali, a ridurre i costi e a ottenere un vantaggio competitivo nei loro settori d’industria. Quasi 4.000 enti governativi e aziende in aree infrastrutturali critiche come quelle dei servizi finanziari, delle telecomunicazioni e sanità si basano sulla piattaforma cloud ibrida di IBM e su Red Hat OpenShift per realizzare la loro trasformazione digitale in modo rapido, efficiente e sicuro. Le innovazioni di IBM nell’ambito dell’AI, del quantum computing, delle soluzioni cloud specifiche per settore d’industria e nei servizi sono offerte con opzioni open e flessibili. Tutto questo è supportato dal ben noto impegno di IBM per la trasparenza, la responsabilità, l'inclusività e il servizio. Per maggiori informazioni, visitate il sito www.ibm.com.

LinkedIn: IBM

Contatti
Morgana Stell - *External Relations Leader, IBM Italia*
email: morgana.stell@it.ibm.com
mobile: 335 7693528

---

https://it.newsroom.ibm.com/mistral