REPLY TARGET

Generative AI offline

Why you should know about it

TARGET INNOVATION PILLS

Di cosa parliamo

Cosa sono i LLM offline?

I Large Language Model offline sono in grado di funzionare in locale, cioè direttamente sul dispositivo dell’utente, senza richiedere una connessione ad un server remoto per l’elaborazione delle richieste

Il dettaglio

REPLY TARGET

> I LLM offline sfruttano al massimo le potenti capacità dei moderni modelli di LLM, consentendo agli utenti di ottenere risposte accurate e contestuali alle domande poste, anche quando non sono connessi ad internet. Inoltre, grazie alla loro architettura efficiente e alle capacità di quantizzazione, possono essere eseguiti su una vasta gamma di dispositivi.

Vantaggi (vs online)

Privacy

Non è necessario inviare i dati sensibili a un server remoto per l'elaborazione. Le informazioni personali dell'utente restano sicure e protette sul proprio dispositivo, riducendo il rischio di violazioni della privacy.

Affidabilità

I LLM offline consentono di ottimizzare in modo efficace le prestazioni del modello, risolvere eventuali bug o problemi senza dover attendere l'intervento del provider di un'API esterna, personalizzando il modello per soddisfare al meglio le proprie esigenze.

Indipendenza da servizi di terze parti

L'utilizzo di modelli offline permette di non dipendere da servizi che possono variare il proprio funzionamento senza avvisare o soffrire interruzioni operative.

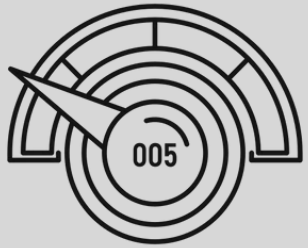
Costi ridotti

Eliminando la necessità di utilizzare e mantenere un'API di terze parti, si possono ridurre i costi operativi e di infrastruttura associati all'elaborazione delle richieste.

Potenziali limiti

Qualità dell’output

Avendo un numero inferiore di parametri rispetto agli online, l’output potrebbe avere limitazoni sulla qualità.



Velocità di risposta

La velocità di risposta dipende dall’hardware utilizzato. Se si utilizza un hardware limitato, la velocità potrebbe essere rallentata.

Da notare che questi modelli sono sempre in continuo sviluppo e queste limitazioni potrebbero essere ridotte anche nel breve tempo.

Use cases in Target Reply

La Gen AI offline trova applicazione nei casi più disparati. Eccone alcuni affrontati:

AutoReply

AutoReply soddisfa le esigenze dei propri utenti implementando un sistema di risposta automatica alle e-mail, in grado di operare offline. Questo sistema gestisce in modo efficiente le richieste comuni del servizio clienti, riducendo al minimo l'impatto sul lavoro degli operatori.
Grazie a questa soluzione, abbiamo alleggerito il loro carico di lavoro, consentendo loro di concentrarsi su attività più complesse e critiche.

Automatic Copywriting

Nel settore dell'e-commerce, si ha la necessità di creare descrizioni accattivanti per diverse pagine, mantenendole costantemente aggiornate con i cambi di stagione e l'arrivo di nuovi prodotti. La nostra soluzione permette di generare descrizioni su misura che si adattano alle esigenze del momento e ai diversi mercati. Il suo funzionamento offline, inoltre, mira a ridurre i costi e i tempi per la generazione di descrizioni, coinvolgendo un numero minore di risorse.

Ci sono domande?

Contattaci!

SCRIVI A:

info.target@reply.it