Aware
now part of mimecast

Serie di dati Futuro del lavoro

Domande da porre al vostro fornitore di intelligenza artificiale

Mimecast ha deciso fin dall'inizio di essere un'azienda AI-native. Questa guida è stata creata in collaborazione con i nostri data scientist per consentire agli acquirenti aziendali di valutare accuratamente l'AI per casi d'uso in tutta l'azienda senza bisogno di possedere una laurea in scienze dei dati. Qualunque fornitore voi scegliate, ecco le domande a cui dovrebbe essere in grado di rispondere e i motivi per cui è importante porle.

INDICE DEI CONTENUTI

Le basi

Quando si valuta un fornitore di AI è utile comprendere i tipi di modelli di AI in uso, poiché questo fattore può influenzare costi, accuratezza e altro ancora.

Quali tipi di modelli di ML/AI utilizza la vostra tecnologia di base?

Possibili risposte includono riferimenti alle modalità di addestramento, alle famiglie di modelli e ai tipi di modelli. Non state cercando un tipo specifico di AI, ma il fornitore dovrebbe conoscere ed essere in grado di spiegare i modelli che utilizza, sapere come funzionano e perché sono l'AI più appropriata per il vostro caso d'uso.

In molti casi sarà più veloce ed economico implementare un modello relativamente semplice e altamente mirato piuttosto che investire in una rete neurale grande e complessa per soddisfare esigenze semplici.

Quale infrastruttura è necessaria per eseguire i modelli? Il cliente (self-hosted) o il fornitore (SaaS) hanno l'hardware necessario?

I modelli di AI possono richiedere risorse significative quando crescono. In che modo i costi o la scarsità delle risorse (ad esempio, la mancanza di accesso alle GPU) inciderebbero sulla capacità di scalare l'AI per i carichi di lavoro aziendali?

I dati

Spazzatura in entrata = spazzatura in uscita. Senza input di dati pertinenti e di alta qualità, i modelli di AI non possono produrre output accurati o fruibili. Inoltre, se il vostro fornitore di AI non è in grado di darvi una risposta sui dati con cui sono stati addestrati i suoi modelli, non c'è modo di sapere come funzionano i modelli o quali fattori influenzano i risultati che tali modelli producono.

Descrivete il processo utilizzato per costruire e addestrare il vostro modello

Dovreste ottenere risposte che menzionano le suddivisioni dei dati 'train-validation-test'. Sono batch di dati utilizzati per addestrare i modelli di AI e di apprendimento automatico, valutare e mettere a punto i loro output e testare i risultati finali. La suddivisione dei dati in set distinti per ciascuna funzione consente ai data scientist di impostare parametri di riferimento in base ai quali è possibile valutare con precisione le prestazioni e il miglioramento dei modelli.

Come monitorate la qualità dei dati prima che il modello sia sviluppato?

Il fornitore dovrebbe essere in grado di spiegare il processo di raccolta e convalida dei dati e, se del caso, il modo in cui i dati etichettati vengono controllati per verificarne la qualità.

Quali sono il tipo, la fonte e il volume di dati necessari per addestrare i Suoi modelli?

Questo aiuta a comprendere la qualità e la quantità dei dati richiesti dal modello di AI. Output più complessi richiedono più dati per l'addestramento.

A titolo di riferimento, un piccolo modello di classificazione potrebbe necessitare di circa 20.000 esempi di alta qualità per ciascuna classe. Un modello di linguaggio di grandi dimensioni richiede 20-30 token (parole, segmenti di codice, ecc.) per parametro nel modello. Anche un piccolo LLM come Llama-2 ha 7 miliardi di parametri, il che significa che ha richiesto almeno 140 miliardi di token per essere addestrato.

I dati (continua)

Con quale frequenza inserisce i dati per addestrare e aggiornare i modelli?

Nel momento in cui un modello di AI viene rilasciato, è già superato. Bilanciare costi e complessità dell'aggiornamento dei modelli con l'obsolescenza è cruciale. Alcuni modelli possono essere aggiornati solo una volta all'anno o anche meno, mentre altri devono essere aggiornati molto più frequentemente per restare utili.

In che modo vengono etichettati i modelli supervisionati? Da dove provengono le etichette?

L'apprendimento supervisionato richiede l'addestramento dell'AI su dataset etichettati. Questo può essere fatto internamente dal team che addestra il modello, in outsourcing, in crowdsourcing o automatizzato. Alcune etichette possono anche utilizzare dati sintetici, che vengono generati artificialmente per soddisfare i criteri richiesti. Una cattiva etichettatura produce cattivi risultati, quindi le etichette devono essere perfezionate e controllate in modo da garantire chiarezza e coerenza.

Chi è responsabile dell'addestramento e della validazione dei modelli? Avete un team interno di machine learning e quanto è grande?

Allo stesso modo, se il cliente è responsabile della messa a punto dei modelli, ha a disposizione un team di machine learning dotato delle competenze necessarie per farlo?

I Modelli

Garantire l'accuratezza costante dei modelli di AI è fondamentale per la loro continua utilità. Senza un piano per aggiornarli regolarmente, la vostra azienda rischia di prendere decisioni basate su dati errati.

In che modo monitorate l'accuratezza e le prestazioni del modello? Chi ne è responsabile?

I modelli devono essere continuamente valutati per verificarne l'accuratezza, il consumo di risorse, l'utilizzo delle API, il volume delle richieste e altro ancora. I dashboard e i trigger di avviso possono supportare questi processi automatizzando molti processi di monitoraggio. Tuttavia, è importante verificare chi esegue il monitoraggio: ad esempio, se la responsabilità è di un team SRE o di un'infrastruttura, potrebbe limitare il monitoraggio alle prestazioni di uptime e trascurare altri fattori come la qualità dell'output.

Potete fornire test sull'accuratezza dei vostri modelli?

L'analisi dei risultati dei test vi fornirà una comprensione completa di quali fattori vengono monitorati, con quale frequenza e se sollevano abitualmente problemi con l'AI.

La vostra AI può essere personalizzata per soddisfare le esigenze dei singoli clienti?

A seconda del motivo per cui avete acquistato l'AI, personalizzarla per le vostre esigenze specifiche può produrre risultati di gran lunga superiori rispetto a un modello generico.

Come affronta la deriva dei dati?

Il termine "drift" si riferisce ai cambiamenti nel tempo che influenzano le proprietà dei dati di addestramento e di input sottostanti. Un modello progettato per valutare il sentiment del discorso, ad esempio, può diventare rapidamente obsoleto man mano che il significato delle parole cambia. Un fornitore dovrebbe essere in grado di spiegare come e con quale rapidità viene rilevato il drift dei dati.

In che modo acquisite e incorporate il feedback nei modelli?

Sebbene esistano modelli di intelligenza artificiale che superano le capacità umane questi modelli possono ancora commettere errori. I fornitori devono essere in grado di spiegare come gli errori individuati dal cliente possano essere integrati nel successivo addestramento del modello in futuro. Spesso questo è fornito da un semplice meccanismo di feedback. Tuttavia, i modelli più complessi potrebbero necessitare di un'interazione diretta tra cliente e fornitore per comprendere meglio gli errori commessi e come i modelli possano essere aggiornati per risolvere il problema.

I modelli (continua)

Con quale rapidità vengono distribuiti in produzione i modelli nuovi e aggiornati? Come vengono distribuite le modifiche e le correzioni di bug?

Informatevi in anticipo su eventuali problemi che potrebbero influire sul tempestivo rilascio di aggiornamenti e correzioni, e se ci sarà un impatto sull'accesso degli utenti finali all'AI durante gli aggiornamenti.

Quali sono i vantaggi dell'utilizzo dei nostri dati per addestrare i modelli?

Il fornitore dovrebbe spiegare come utilizza in modo responsabile i vostri dati aziendali per addestrare i suoi modelli di AI, affinando i risultati per allinearsi al modo in cui la vostra organizzazione comunica e opera. Questo approccio gli consente di fornire soluzioni più accurate, personalizzate ed efficaci su misura per le vostre esigenze specifiche, che si traducono in un miglior ROI, informazioni preziose e una protezione avanzata. Poiché la sua AI si adatta e migliora continuamente, vi offre una solida sicurezza e una soluzione che si evolve di pari passo con la vostra organizzazione.

Descrivete la pipeline di elaborazione dei dati che vi consente di distribuire modelli nuovi e aggiornati ai clienti.

Le pipeline dei modelli di AI possono spesso risultare complesse. Un fornitore dovrebbe essere in grado di spiegare il processo attraverso il quale queste pipeline possono essere implementate presso i clienti in modo rapido ed efficiente e con tempi di inattività minimi.

Il costo

I modelli di AI spesso richiedono una vasta potenza di calcolo per funzionare. Comprendere la portata dell'AI in uso e i fattori che influiscono sui costi è essenziale per individuare correttamente l'AI giusta per le vostre esigenze.

Qual è il costo stimato per costruire un modello?

Qual è il costo di esercizio del modello?

Qual è la dimensione tipica dei vostri modelli?

Ricordate che molti modelli di AI crescono nel tempo, man mano che vengonoperfezionati sulla base di nuovi dati.

Scalabilità

La scalabilità si riferisce alla capacità del modello di gestire più dati, utenti e attività senza perdere in prestazioni. Questo è particolarmente importante nei contesti aziendali.

Quanto è rapida l'inferenza del modello?

L'inferenza del modello è il processo di fornitura di una previsione per un determinato insieme di dati (ad esempio, la stima della probabilità che un cliente abbandoni in base a una recente interazione con il call center). In molti casi, le inferenze possono essere effettuate quasi in tempo reale mentre vengono acquisiti nuovi dati. Tuttavia, a seconda del tipo di modello, delle esigenze di dati e delle considerazioni sui costi, potrebbe essere più sensato elaborare in batch. In qualità di cliente, dovete considerare le implicazioni aziendali dell'inferenza in batch rispetto a quella in tempo reale.

Quanto è scalabile l'AI in termini di acquisizione dei dati?

Come vengono inseriti i vostri dati nel sistema per l'ottimizzazione del modello e potete escludere i dati che non desiderate che il modello elabori?

In che modo l'AI si integra con i miei sistemi esistenti?

Quale tipo di intervento IT è necessario per collegare l'AI? Ci sono costi per l'utilizzo delle API? Chi scrive e mantiene questo codice? E quale infrastruttura è necessaria dal lato del cliente per far funzionare l'integrazione?

Responsabilità

L'AI è stata oggetto di critiche negative per la raccolta e la gestione non etica dei dati e, di conseguenza, i responsabili legali e della sicurezza informatica potrebbero esitare ad autorizzarne l'uso. Le risposte a queste domande possono aiutare ad alleviare tali preoccupazioni.

Che tipo di impegno, promessa o garanzia offre la vostra azienda riguardo all'uso dell'AI, dei dati e delle informazioni per sviluppare fiducia e trasparenza?

Quali sono la vostra politica e il vostro impegno nel seguire pratiche responsabili di AI/ML che diano priorità alla privacy, all'equità, alla trasparenza e all'interpretabilità, garantendo al contempo sicurezza e responsabilità attraverso la supervisione umana? Riconoscete anche l'importanza della sostenibilità e integrate pratiche rispettose dell'ambiente nelle vostre iniziative di AI?

In che modo il modello di AI gestisce l'equità e il bias?

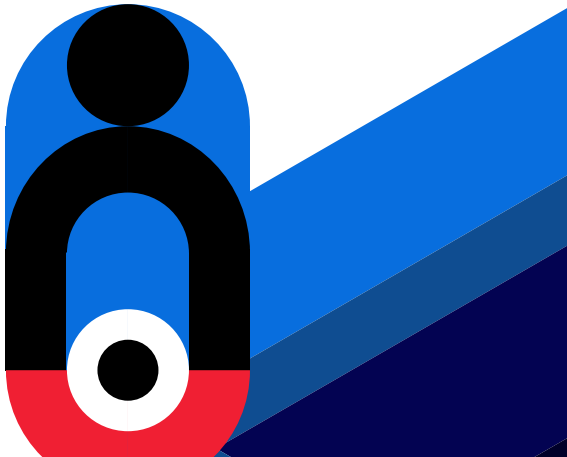
A seconda del caso d'uso dell'AI, bias ed equità possono avere significati molto diversi e avere un'importanza variabile per gli output del modello. Valutate le misure delineate dal fornitore per contrastare il bias rispetto all'impatto sui risultati finali.

In che modo i dati vengono protetti e governati?

I dati saranno ospitati dal cliente o dal fornitore? Quale infrastruttura esiste per gestire i dati e come verrà protetta?

Considerazioni finali

L'AI è un argomento di grande interesse nel mondo del business e può offrire vantaggi straordinari a tutti i livelli aziendali. Tuttavia, prima di effettuare un acquisto di AI, è importante allineare la tecnologia che selezionate con i risultati che desiderate ottenere. Valutate le risposte alle domande di cui sopra tenendo conto delle vostre esigenze finali per prendere la decisione giusta quando introducete l'AI nella vostra organizzazione.




Aware
now part of mimecast