Aware
now part of mimecast

Série sur l'avenir des données de travail

Questions essentielles à poser à votre fournisseur d'IA

Chez Mimecast, nous avons très tôt fait le choix d'intégrer l'IA au cœur de notre stratégie. Ce guide, élaboré avec nos experts en science des données, a été conçu pour aider les acheteurs professionnels à évaluer précisément l'IA pour des cas d'usage en entreprise, sans avoir besoin d'un diplôme en data science.

Peu importe le fournisseur que vous consultez, voici les questions qu'il doit être capable de clarifier et pourquoi il est essentiel de les poser.

TABLE DES MATIÈRES

Les bases

Lors de l'évaluation d'un fournisseur d'IA, il est utile de comprendre les types de modèles d'IA utilisés, car cela peut influencer les coûts, la précision et d'autres facteurs.

Quels types de modèles ML/IA votre technologie de base emploie-t-elle ?

Les réponses possibles comprennent des références aux modalités d'entraînement, aux familles de modèles et aux types de modèles. Vous ne recherchez pas un type d'IA spécifique, mais le fournisseur doit savoir et pouvoir expliquer quels modèles il utilise, comment ils fonctionnent et pourquoi ils sont les plus appropriés pour votre cas d'utilisation.

Dans de nombreux cas, il sera plus rapide et moins coûteux de déployer un modèle relativement simple et très ciblé que d'investir dans un grand réseau neuronal complexe pour répondre à des besoins simples.

Quelle infrastructure est nécessaire pour exécuter les modèles ? Le client (auto-hébergé) ou le fournisseur (SaaS) dispose-t-il du matériel nécessaire ?

Les modèles d'IA peuvent nécessiter des ressources considérables à mesure qu'ils se développent. Quel impact les coûts ou le manque de ressources (par exemple, le manque d'accès aux GPU) pourraient-ils avoir sur la capacité à adapter l'IA aux charges de travail de l'entreprise ?

Les Données

« Garbage in, garbage out ». Sans données d'entrée pertinentes et de haute qualité, les modèles d'IA ne peuvent pas produire des résultats précis ou exploitables. De plus, si votre fournisseur d'IA ne peut pas préciser sur quelles données ses modèles ont été entraînés, il n'y a aucun moyen de savoir comment les modèles fonctionnent ou quels sont les facteurs qui influencent les résultats qu'ils produisent.

Décrivez le processus utilisé pour créer et entraîner votre modèle.

Recherchez les réponses qui mentionnent des divisions de données de type « entraînement-validation-test ». Il s'agit de lots de données utilisés pour entraîner les modèles d'IA et d'apprentissage automatique, évaluer et affiner leurs résultats, et tester les résultats finaux. La division des données en différents ensembles pour chaque fonction permet aux spécialistes en science des données de fixer des points de référence pour évaluer avec précision les performances et l'amélioration des modèles.

Comment surveillez-vous la qualité des données avant que le modèle ne soit développé ?

Le fournisseur doit être en mesure d'expliquer le processus de collecte et de validation des données et, le cas échéant, la manière dont la qualité des données étiquetées est vérifiée.

Quel est le type, la source et le volume de données nécessaires pour entraîner vos modèles ?

Cela vous permet de comprendre la qualité et la quantité des données requises par le modèle d'IA. Les résultats plus complexes nécessitent plus de données pour l'entraînement.

À titre de référence, un petit modèle de classification peut nécessiter environ 20 000 exemples de haute qualité par classe. Un grand modèle de langage nécessite 20 à 30 jetons (mots, segments de code, etc.) par paramètre du modèle. Même un petit LLM, comme Llama-2, possède 7 milliards de paramètres. Cela signifie qu'il a fallu au moins 140 milliards de jetons pour l'entraîner.

Les données (suite)

À quelle fréquence ingérez-vous des données pour entraîner et mettre à jour les modèles ?

Dès qu'un modèle d'IA est publié, il devient obsolète. Il est essentiel de trouver un équilibre entre le coût et la complexité de la mise à jour des modèles et leur obsolescence. Certains modèles peuvent n'être mis à jour qu'une fois par an, voire moins, tandis que d'autres doivent l'être beaucoup plus régulièrement pour rester utiles.

Comment les modèles supervisés sont-ils étiquetés ? D'où viennent les étiquettes ?

L'apprentissage supervisé nécessite d'entraîner l'IA sur des ensembles de données étiquetés. Cette opération peut être réalisée en interne par l'équipe qui entraîne le modèle, externalisée, croudsourcée ou automatisée. Certaines étiquettes peuvent également utiliser des données synthétiques, qui sont générées artificiellement pour correspondre aux critères requis. Un mauvais étiquetage produit de mauvais résultats. Les étiquettes doivent donc être affinées et vérifiées pour assurer clarté et cohérence.

Qui est responsable de l'entraînement et de la validation des modèles ? Avez-vous une équipe interne chargée de l'apprentissage automatique, et quelle est sa taille ?

De même, si le client est responsable de l'ajustement du(des) modèle(s), dispose-t-il d'une équipe spécialisée en apprentissage automatique, avec l'expertise nécessaire pour le faire efficacement ?

Les Modèles

Il est essentiel de garantir l'exactitude permanente des modèles d'IA pour qu'ils conservent leur utilité. Sans un programme régulier d'ajustement et de mise à jour, votre entreprise risque de prendre des décisions sur la base de données erronées.

Comment surveillez-vous la précision et les performances du modèle ? Qui en est responsable ?

Les modèles doivent être évalués en permanence en termes de précision, de consommation de ressources, d'utilisation de l'API, de volume de requêtes, etc. Les tableaux de bord et les déclencheurs d'alerte peuvent contribuer à ces processus en automatisant de nombreuses tâches de surveillance. Cependant, il est important de savoir qui est chargé de la surveillance : par exemple, si une équipe de SRE ou d'infrastructure est responsable, il se peut qu'elle limite la surveillance aux performances de disponibilité et néglige d'autres facteurs tels que la qualité des résultats.

Pouvez-vous fournir des tests de précision de vos modèles ?

L'examen des résultats des tests vous permettra de mieux comprendre quels sont les facteurs surveillés, à quelle fréquence et s'ils révèlent régulièrement des problèmes liés à l'IA.

Votre IA peut-elle être personnalisée pour répondre aux besoins de chaque client ?

Selon la raison de votre achat d'une solution optimisée par l'IA, il peut être utile de la personnaliser en fonction de vos besoins spécifiques pour obtenir des résultats bien supérieurs à ceux d'un modèle générique.

Comment abordez-vous le Data Drift ou dérive des données ?

La « dérive » désigne les changements au fil du temps qui ont un impact sur les propriétés des données d'entraînement et d'entrée sous-jacentes. Un modèle conçu pour évaluer le sentiment d'un discours, par exemple, peut rapidement devenir obsolète à mesure que le sens des mots évolue. Le fournisseur doit être en mesure d'expliquer comment et à quelle vitesse la dérive des données est détectée.

Comment capturez-vous et intégrez-vous les retours dans vos modèles ?

Bien que certains modèles d'IA surpassent les capacités humaines, ils peuvent tout de même commettre des erreurs. Le fournisseur doit pouvoir expliquer comment les erreurs détectées par le client peuvent être intégrées dans le réentraînement futur du modèle. Souvent, cela se fait par un simple mécanisme de retour d'information. Cependant, les modèles plus complexes peuvent nécessiter une interaction directe entre le client et le fournisseur pour mieux comprendre les erreurs commises et comment les modèles peuvent être mis à jour pour résoudre le problème.

Les Modèles (suite)

À quelle vitesse les nouveaux modèles et modèles actualisés sont-ils déployés en production ? Comment les modifications et les corrections de bugs sont-elles déployées ?

Renseignez-vous à l'avance sur les problèmes qui pourraient avoir une incidence sur le déploiement en temps voulu des mises à jour et des correctifs, et sur l'impact éventuel sur l'accès des utilisateurs finaux à l'IA pendant les mises à jour.

Quels sont les avantages de l'utilisation de nos données pour entraîner les modèles ?

Le fournisseur doit expliquer comment il utilise de manière responsable les données de votre entreprise pour entraîner ses modèles d'IA, en affinant les résultats pour qu'ils correspondent à la manière dont votre organisation communique et fonctionne. Cette approche permet de fournir des solutions plus précises, personnalisées, efficaces et adaptées à vos besoins spécifiques. Cela se traduit par un meilleur retour sur investissement, des informations exploitables et une protection renforcée. Au fur et à mesure que l'IA s'adapte et s'améliore, elle vous offre une sécurité robuste et une solution qui évolue parallèlement à votre organisation.

Décrivez le pipeline de traitement des données qui permet de déployer de nouveaux modèles ou des modèles actualisés auprès des clients.

Les pipelines de modèles d'IA peuvent souvent être complexes. Le fournisseur doitre être en mesure d'expliquer le processus permettant de déployer ces pipelines rapidement et efficacement auprès des clients, avec un minimum de temps d'arrêt.

Le coût

Les modèles d'IA nécessitent souvent une grande puissance de calcul pour fonctionner. Il est essentiel de comprendre l'échelle de l'IA utilisée et les facteurs qui influent sur le coût pour choisir l'IA qui répondra à vos besoins.

Quel est le coût estimé pour construire un modèle ?

Quel est le coût d'exécution du modèle ?

Quelle est la taille typique de vos modèles ?

N'oubliez pas que de nombreux modèles d'IA se développent au fil du temps, au fur et à mesure qu'ils sont perfectionnés en fonction de nouvelles données.

Extensibilité

L'extensibilité désigne la capacité du modèle à traiter plus de données, d'utilisateurs et de tâches sans compromettre les performances. Cet aspect est particulièrement important dans les entreprises.

Quelle est la vitesse d'inférence du modèle ?

L'inférence du modèle est le processus consistant à fournir une prédiction pour un certain ensemble de données (par exemple, estimer la probabilité qu'un client se désabonne après une interaction récente avec un centre d'appels). Dans de nombreux cas, des inférences peuvent être créées en temps quasi réel à mesure que de nouvelles données sont intégrées. Cependant, selon le type de modèle, les besoins en données et les facteurs liés au coût, il peut être plus judicieux de procéder par lots. En tant que client, vous devez prendre en compte les implications commerciales de l'inférence par lots par rapport à l'inférence en temps réel.

Quelle est l'extensibilité de l'IA en termes d'ingestion de données ?

Comment vos données sont-elles intégrées dans le système pour l'ajustement du modèle, et pouvez-vous exclure les données que vous ne souhaitez pas que le modèle absorbe ?

Comment l'IA s'intègre-t-elle à vos systèmes existants ?

Quelle infrastructure informatique est nécessaire pour connecter l'IA ? L'utilisation des API a-t-elle un coût ? Qui rédige et entretient ce code ? Et quelle est l'infrastructure nécessaire côté client pour que l'intégration fonctionne ?

Responsabilité

L'IA a eu mauvaise presse en raison de sa collecte et de sa gestion de données contraires à l'éthique. Par conséquent, les responsables juridiques et de la sécurité informatique peuvent hésiter à autoriser son utilisation. Les réponses à ces questions peuvent contribuer à apaiser ces inquiétudes.

Quel type d'engagement, de promesse ou de garantie votre entreprise propose-t-elle concernant l'utilisation de l'IA, des données et des analyses pour instaurer la confiance et assurer la transparence ?

Quelle est votre politique et votre engagement à suivre des pratiques responsables en matière d'IA/de ML, favorisant la confidentialité, l'équité, la transparence et l'interprétabilité, tout en garantissant la sécurité et la responsabilité grâce à la surveillance humaine ? Nous reconnaissons également l'importance du développement durable et intégrons des pratiques respectueuses de l'environnement dans nos initiatives en lien avec l'IA.

Comment le modèle d'IA prend-il en compte l'équité et les biais ?

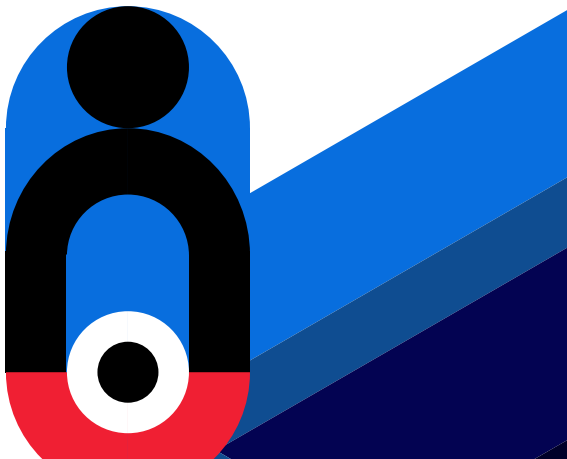
Selon le cas d'utilisation de l'IA, les notions de biais et d'équité peuvent avoir des significations très différentes et un impact variable sur les résultats du modèle. Évaluez les mesures mises en place par le fournisseur pour contrer les biais qui pourraient affecter les résultats finaux.

Comment les données sont-elles sécurisées et gérées ?

Les données seront-elles hébergées par le client ou par le fournisseur ? Quelle est l'infrastructure de gestion des données existante et comment seront-elles protégées ?

Réflexions finales

L'IA est un sujet brûlant dans le monde des affaires, et elle peut apporter des avantages considérables dans une entreprise. Cependant, avant d'acheter une solution optimisée par l'IA, il est important de sélectionner une technologie adaptée aux résultats que vous souhaitez obtenir. Évaluez les réponses aux questions ci-dessus en gardant à l'esprit vos besoins finaux pour prendre la bonne décision lors de l'introduction de l'IA dans votre entreprise.




Aware
now part of mimecast