Aware
now part of mimecast

FUTURE OF WORK DATA SERIES

Questions to ask your AI vendor

At Mimecast, we made the decision early on to be an AI-native company. This guide was created in consultation with our data scientists to enable corporate buyers to accurately assess AI for use cases across the enterprise without needing a data science degree to do it. Whatever vendor you choose, here are the questions they should be able to answer and the reasons why asking them matters.

TABLE OF CONTENTS

The Basics

When vetting any AI vendor, it's helpful to understand the types of AI models in use, as that can inform cost, accuracy and more.

What types of ML/AI models does your core technology employ?

Potential answers include references to training modalities, model families, and model types. You're not looking for a specific type of AI here, but the vendor should know and be able to explain what models they use, how they work, and why they're the most appropriate AI for your use case.

In many cases, it will be faster and cheaper to deploy a relatively simple, highly targeted model than to invest in a large, complex neural network to solve straightforward needs.

What infrastructure is required to run the models? Does the customer (self-hosted) or vendor (SaaS) have the necessary hardware?

AI models can demand significant resources as they grow. How would cost or resource scarcity (e.g., lack of access to GPUs) impact the ability to scale the AI for enterprise workloads?

The Data

Garbage in = garbage out. Without high quality, relevant input data, AI models cannot produce accurate or actionable outputs. Further, if your AI vendor cannot answer what data their models were trained on, there's no way of knowing how the models work or what factors influence the results they produce.

Describe the process used to build and train your model.

Look for responses that mention "train-validation-test" data splits. These are batches of data used to teach AI and machine learning models, evaluate and fine-tune their outputs, and test the final results. Splitting the data into different sets for each function allows data scientists to set benchmarks by which performance and improvement of the models can be accurately assessed.

How do you monitor data quality before the model is developed?

The vendor should be able to explain the process of data gathering, validation and, when relevant, how labeled data is checked for quality.

What is the type, source, and volume of data required to train your models?

This helps you understand the quality and quantity of the data required by the AI model. More complex outputs require more data to train.

For reference, a small classification model might need around 20,000 high quality examples per class. A large language model requires 20-30 tokens (words, code segments, etc.) per parameter in the model. Even a small LLM, like Llama-2, has 7 billion parameters, meaning it required at least 140 billion tokens to train.

The Data (continued)

How often do you ingest data to train and update models?

The moment an AI model is released, it's out of date. Balancing the cost and complexity of refreshing models against obsolescence is critical. Some models may only be refreshed yearly or even less, while others must be refreshed much more regularly to remain useful.

How are supervised models labeled? Where do the labels come from?

Supervised learning requires training AI on labeled datasets. This can be done in-house by the team training the model, outsourced, crowdsourced, or automated. Some labels may also use synthetic data, which is artificially generated to match the required criteria. Bad labeling produces bad results, so labels should be refined and checked for clarity and consistency.

Who is responsible for training and validating models? Do you have an in-house machine learning team, and how big is it?

Equally, if the customer is responsible for fine-tuning the model(s), do they have a machine learning team—with the relevant expertise—available to do it?

The Models

Ensuring the ongoing accuracy of AI models is essential to their continued usefulness. Without a plan to refresh and update them regularly, your business risks making decisions based on bad data.

How do you monitor model accuracy and performance? Who is responsible for this?

Models should be continually assessed for accuracy, resource consumption, API usage, request volume, and more. Dashboards and alerting triggers can support these processes by automating many monitoring processes. However, it's important to review who is doing the monitoring—e.g., if an SRE or infrastructure team is responsible, they may be limiting monitoring to uptime performance and overlooking other factors such as output quality.

Can you provide tests of your models' accuracy?

Reviewing test results will give you the mostcomprehensive understanding of what factors are monitored, how often, and if they routinely surface issues with the AI.

Can your AI be customized to meet individual customers' needs?

Depending on the reason for your AI purchase, customizing it for your specific needs may produce far superior results compared to a generic model.

How do you address data drift?

"Drift" refers to changes over time that impact the properties of the underlying training and input data. A model designed to score the sentiment of speech, for instance, can be quickly made obsolete as the meaning of words changes. A vendor should be able to explain how—and how fast—data drift is detected.

How do you capture and incorporate feedback into your models?

While there are AI models that exceed human performance, these models can still make errors. Vendors need to be able to explain how errors caught by the customer can be incorporated into the subsequent retraining of the model in the future. Often this is provided by a simple feedback mechanism. But more complex models may need direct customer-vendor interaction to better understand the errors being made and how the models can be updated to resolve the issue.

The Models (continued)

How quickly are new and refreshed models deployed into production? How are changes and bug fixes deployed?

Learn in advance about any issues that could impact the timely roll-out of updates and fixes, and whether there will be any impact on the end users' access to the AI during updates.

What are the benefits to using our data to train the models?

The vendor should explain how they responsibly use your corporate data to train their AI models, refining results to align with how your organization communicates and operates. This approach enables them to deliver more accurate, customized, and effective solutions tailored to your specific needs, translating into improved ROI, valuable insights, and enhanced protection. As their AI continually adapts and improves, it provides you with robust security and a solution that evolves alongside your organization.

Describe the data processing pipeline that allows you to deploy new and refreshed models to customers.

AI model pipelines can often be complicated. A vendor should be able to explain the process through which these pipelines can be deployed to customers quickly and efficiently and with minimal downtime.

The Cost

AI models often require vast computing power to function. Understanding the scale of the AI in use and the factors impacting cost are essential to properly scoping the right AI for your needs.

What is the estimated cost to build a model?

What is the model run-time cost?

What is the typical size of your models?

Remember that many AI models grow over time as they are refined against new data.

Scalability

Scalability refers to the ability of the model to handle more data, users, and tasks without losing performance. This is especially important in enterprise settings.

How fast is the model inference?

Model inference is the process of providing a prediction for a given set of data (e.g., estimating the likelihood a customer will churn based on a recent call center interaction). In many cases, inferences can be made in near-real time as new data is ingested. However, depending on the model type, data needs, and cost considerations, it may make more sense to process in batches. As a customer, you must consider the business implications of batch versus real-time inference.

How scalable is the AI in terms of ingesting data?

How does your data get into the system for model tuning, and can you exclude data you don't want the model to ingest?

How does the AI integrate with my existing systems?

What kind of IT lift is required to connect the AI? Is there are cost to using APIs? Who writes and maintains this code? And what infrastructure is needed on the customer end to make the integration work?

Responsibility

AI has received some bad press for unethical data harvesting and management, and legal and infosec leaders may be hesitant to authorize the use of AI as a result. The answers to these questions can help alleviate those concerns.

What kind of commitment, promise or pledge does your company offer around the use of AI, data and insights to build trust and transparency?

What is your policy and commitment on following responsible AI/ML practices that prioritize privacy, fairness, transparency, and interpretability, while ensuring safety and accountability through human oversight. We also recognize the importance of sustainability and integrate environmentally conscious practices into our AI initiatives.

How does the AI model handle fairness and bias?

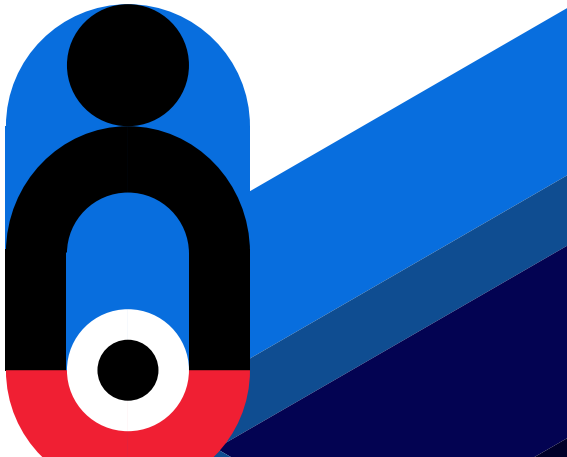
Depending on the use case of the AI, bias and fairness can mean very different things and have variable significance to model outputs. Assess the measures the vendor outlines to counteract bias against the impact on the end results.

How is the data secured and governed?

Will the data be customer- or vendor-hosted? What infrastructure exists to manage the data and how will it be protected?

Final Thoughts

AI is a hot topic in the business world, and it can unlock incredible advantages across the enterprise. However, before making an AI purchase it's important to align the technology you select with the outcomes you want to achieve. Evaluate the answers to the above questions with your end needs in mind to make the right decision when introducing AI to your organization.




Aware
now part of mimecast