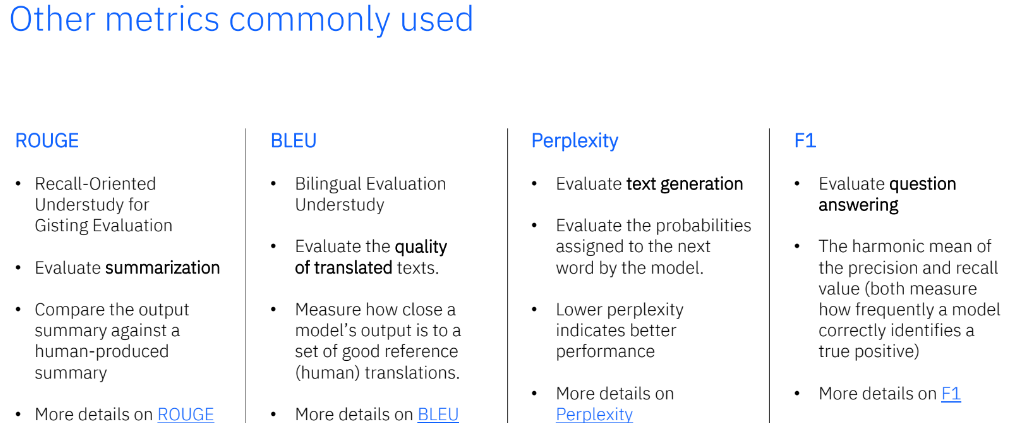
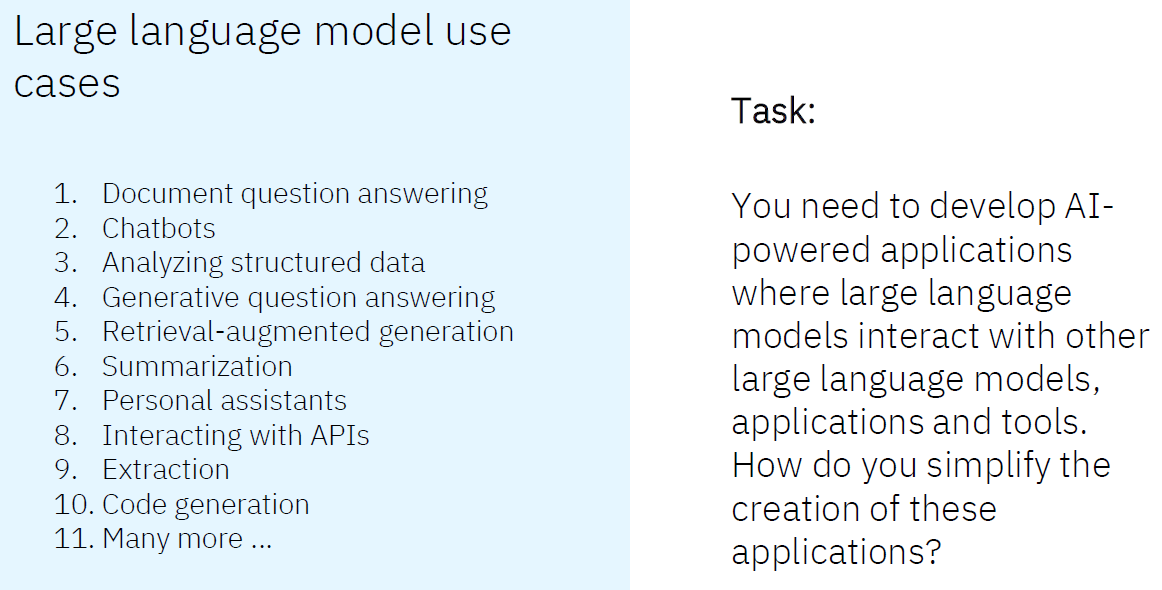
# Foundation models and metrics [watsonx.ai PoX L4]

How to evaluate a foundation models



# Introduction to LangChain [watsonx.ai PoX L4]

LangChain is a popular open source framework that’s becoming a de facto standard for LLM application development. LangChain simplifies implementation of many tasks that are typical in LLM applications, such as:  
• Using prompt templates  
• Parsing output of LLMs  
• Creating a sequence of calls to LLMs  
• Maintaining session state between LLM calls (memory)  
• A systematic approach for implementation of RAG use cases.



A simple LLM chain, where aprompt template component is passed to an LLM component to get a response from the language model.

Features

LangChain is:

* Aframework for developing applications powered by language models
* A feature-rich, fast-growing
* API

Popular “utility” features are:

* Prompt templates
* LLM output parsers
* Wrappers around a series of single components (chains)
* Maintaining session state between LLM calls (memory)
* Support for RAG (specific pattern)

How to access LangChain

From IBM Cloud:

* Create a Watson Studio Project
* Create and associate a Watson Machine (WML) service with the Project
* Install ibm-watson-machine-learning Python library
* Define WML credentials
* Install langchain

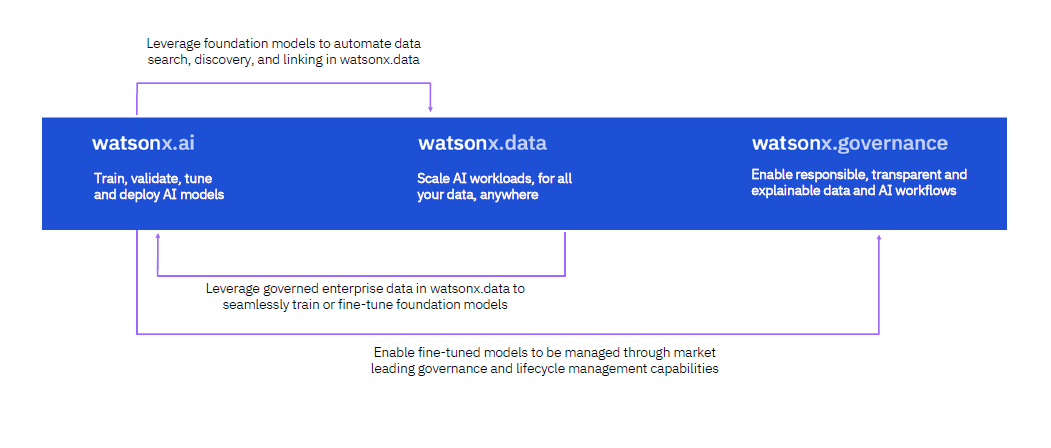
From a Python Integrated Development Environment (IDE) such as PyCharm:

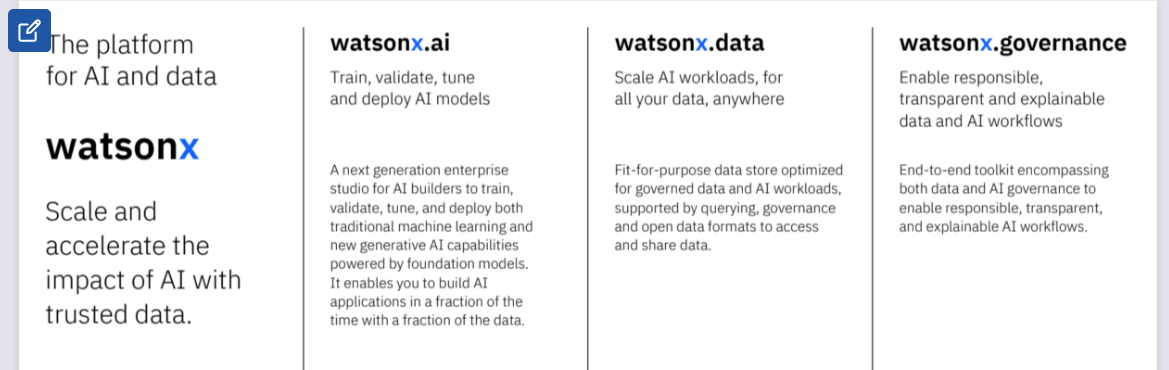
* Create a Watson Studio Project with associated WML service in the IBM Cloud
* pip install ibm-watson-machine-learning Python library in IDE
* Define WML credentials
* Install langchainin IDE

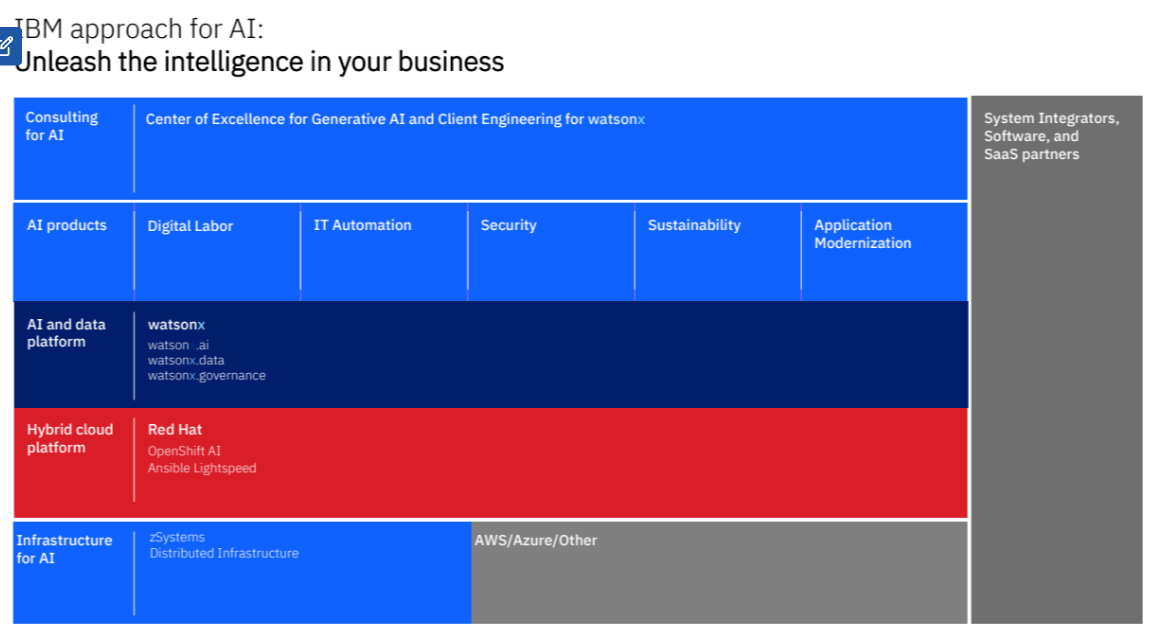
# Environment Setup for Watson

* IBM watsonx.ai (using the IBM Technology Sales TechZone pattern)
* IBM watsonx.ai + Watson Discovery + watsonx Assistant (using the IBM Technology Sales TechZone pattern)
* REST API access for when a client wants to use their own applications (this path does not require a TechZone pattern)

Watsonx.ai is IBM AI builder helps to validate, train and deploy the genAI models







## IBM watsonx.ai

An enterprise-ready next-generation AI studio for AI builders, bringing together traditional machine learning and new generative AI capabilities powered by foundation models. IBM watsonx.ai can train, validate, tune, and deploy AI models.

* Build AI applications in a fraction of the time with a fraction of the data.
* Guide models to meet your needs, with easy-to-use tools for building and refining performant prompts to achieve the desired result.
* Tune models with your enterprise data; your data remains private and secure.
* An enterprise-ready next-generation AI studio for AI builders, bringing together traditional machine learning and new generative AI capabilities powered by foundation models. IBM watsonx.ai can train, validate, tune, and deploy AI models.

IBM watsonx.ai components

• Foundation models library

• Prompt lab

• Tuning studio

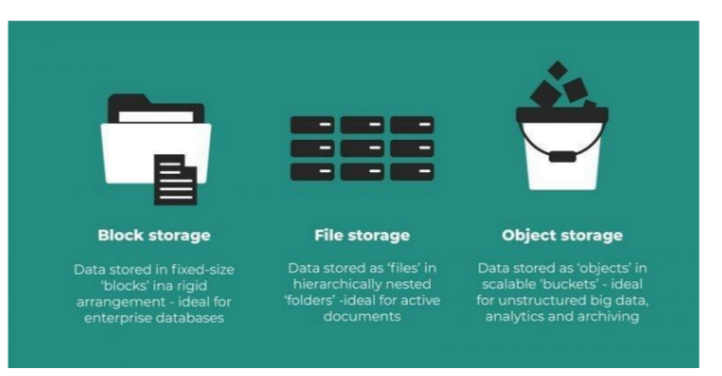
• Synthetic data Generator

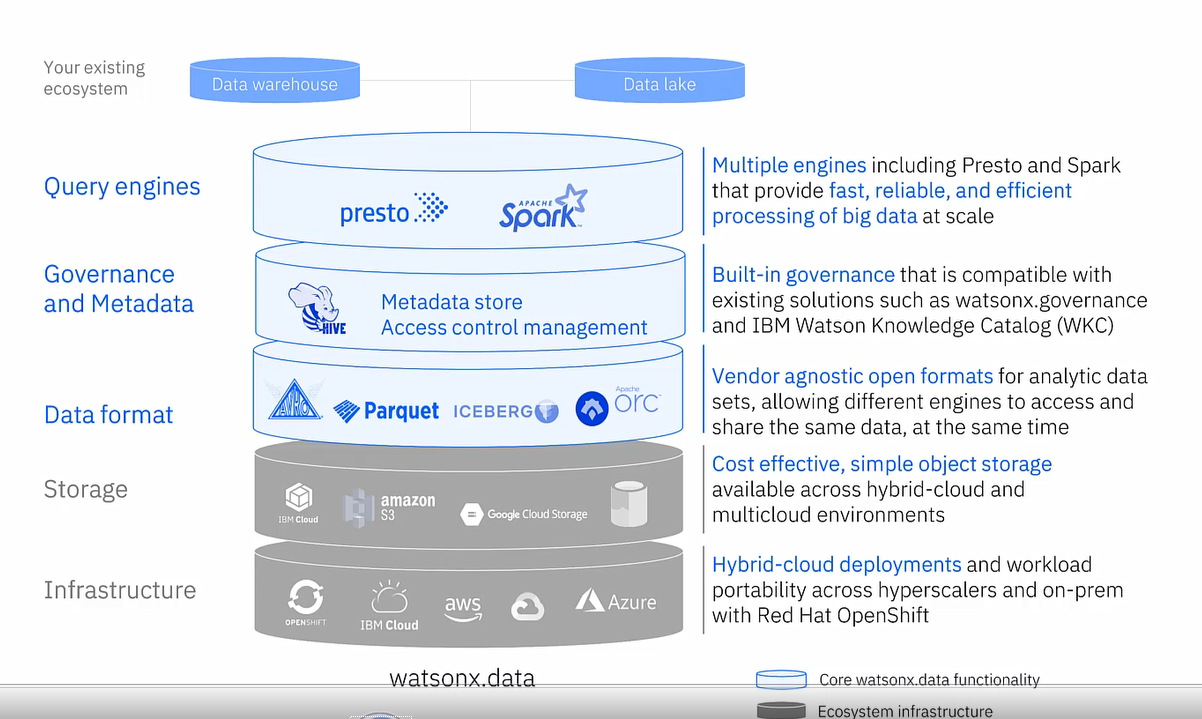
## IBM watsonx.data

## Data Lake && Data Warehouse

Scale AI workloads, for all data, anywhere. IBM watsonx.data can store, manage, enrich, and access all your data for AI.

* Reduce the cost of your data warehouse by up to 50% through workload optimization across query engines and storage tiers
* Get started inminutes with built-in governance, security, and automation
* Access all your data through a single point of entry across all clouds and on-prem environments



s

## IBM watsonx.governance

An enterprise-ready solution that enables responsible, transparent, and explainable workflows across the AI lifecycle.

* Trace and document the origin of datasets, models, and pipelines —so you can explain your AI’s decisions, every time
* Monitor AI models for fairness, bias and drift —and take action real-time if they go awry
* Manage your AI Lifecycle while upholding internal policies and external regulations

## IBM watsonx Code Assistant

Empowering hybrid cloud developers with AI-generated code recommendations.

* Reduce time-to-productivity and narrow the IT Automation skills gap with AI-generated code
* Maintain high levels of accuracy and transparency through attribution of generated content recommendations
* Tune the Foundation Model with your own data and customize it with your own standards and best practices.

## IBM watsonx Orchestrate

Leverage AI and automation, empower individuals to do work without expert knowledge of business processes and applications.

* Democratize the availability of automation through natural language
* Create a highly accessible experience for non-technical users to be able to leverage
* Allow IT to build and importcustom capabilities for end users without disrupting existing workflow.

## IBM watsonx Assistant

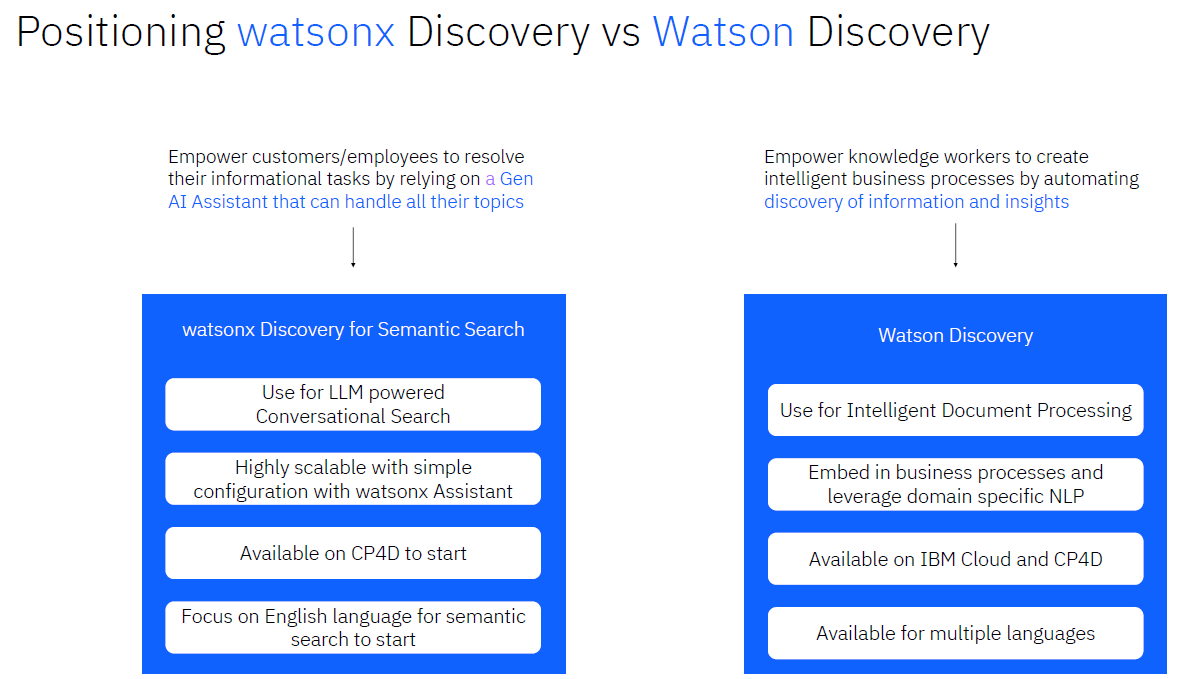
Automate interactions with customers using chatbots or agents across digital and voice channels using a virtual assistant platform that utilizes artificial intelligence techniques, like NLP and ML

* Generative AI-first. Assistant uses LLM-powered AI to achieve the highest accuracy with 80% less training effort, as well as trusted conversational search
* Simplified build experience. No-code and generative authoring makes it 3x faster to build with watsonx Assistant than any other platform
* Designed to connect to all the customer support channels and back-end systems, IBM watsonx Assistantcreates an experience that is focused on self-service.

## IBM watsonx Discovery

Boost the productivity of knowledge workers by automating the discovery of information and insights with advanced NLP and NLU.

* Quickly make business decisions & automate workflows with highly accurate insights from large, unstructured, and complex data
* Teachdomain-specific understanding through custom NLP models to extract the insights you need
* to automate workflows and make business decisions
* Discover deeper insights by using extracting content to segment documents and ignore irrelevant parts.

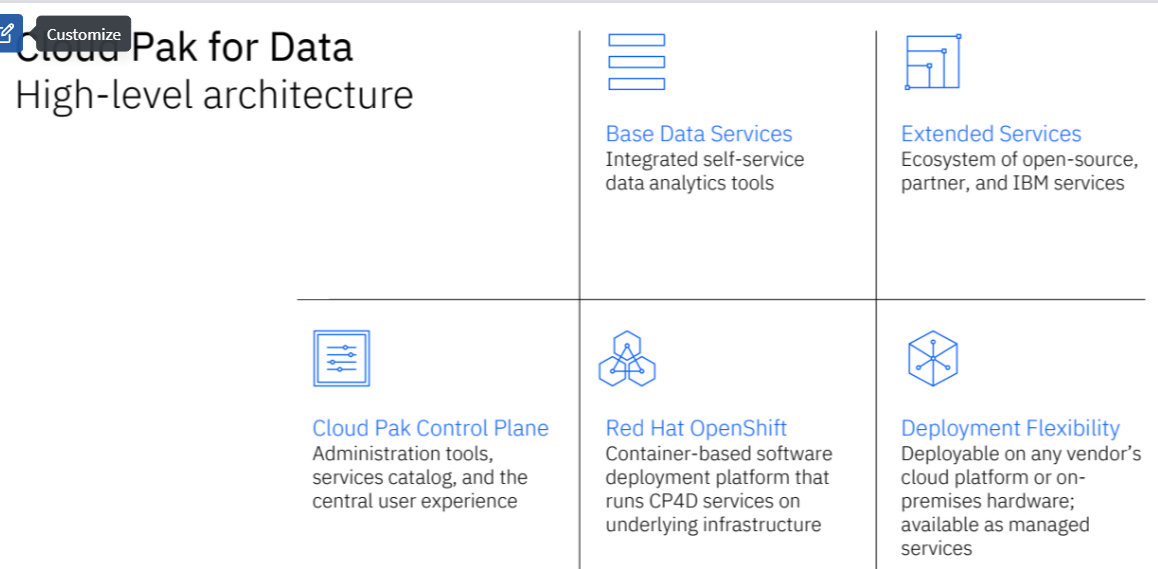


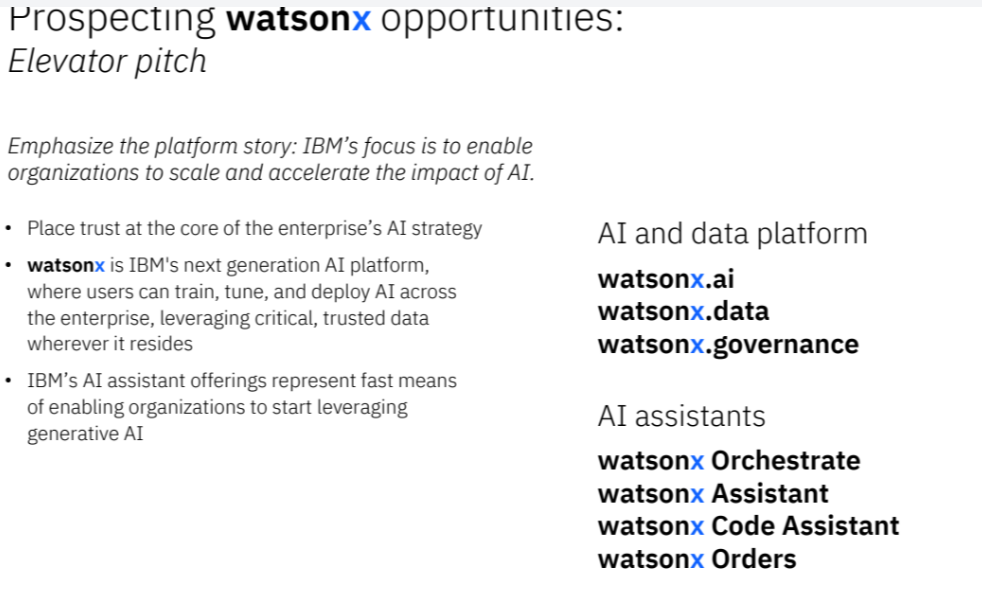
# Prompt Tuning

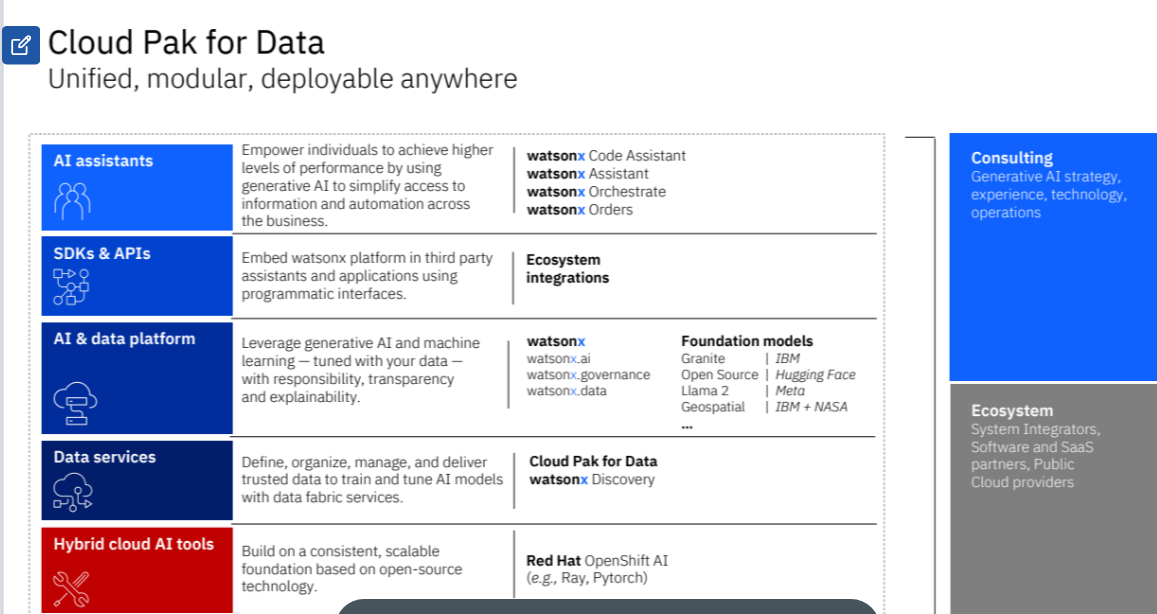
Prompt tuning is accessible via the watsonx.ai Tuning Studio. Note that prompt tuning is different from prompt engineering.

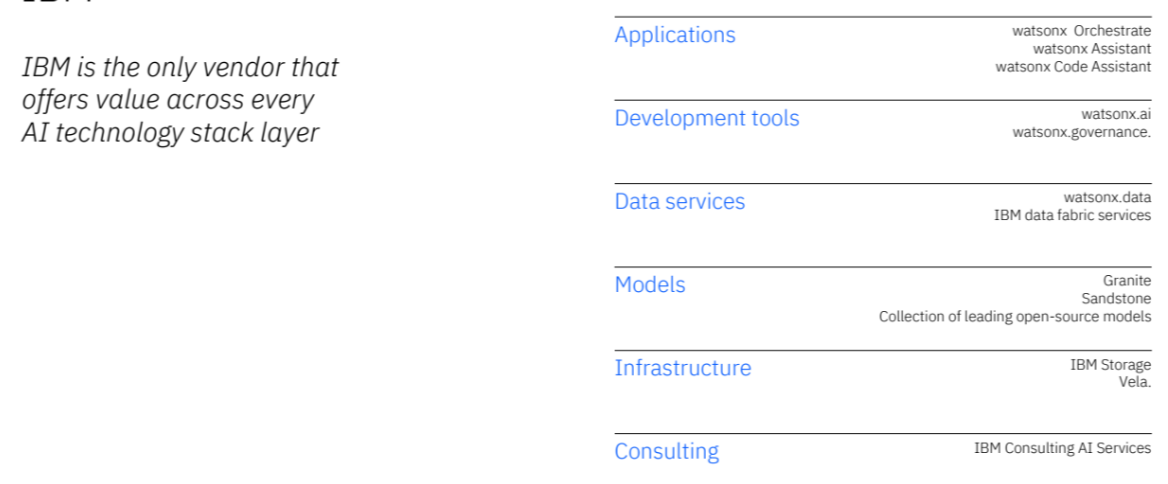
In prompt engineering, a user is modifying what is fed into the foundation model. These are “hard prompts” modified by the user. This is useful in providing context, simple examples, and suggestions for output structures to the model. However, there are limits to how much prompt engineering (even with one-shot or multi-shot prompting) can do.

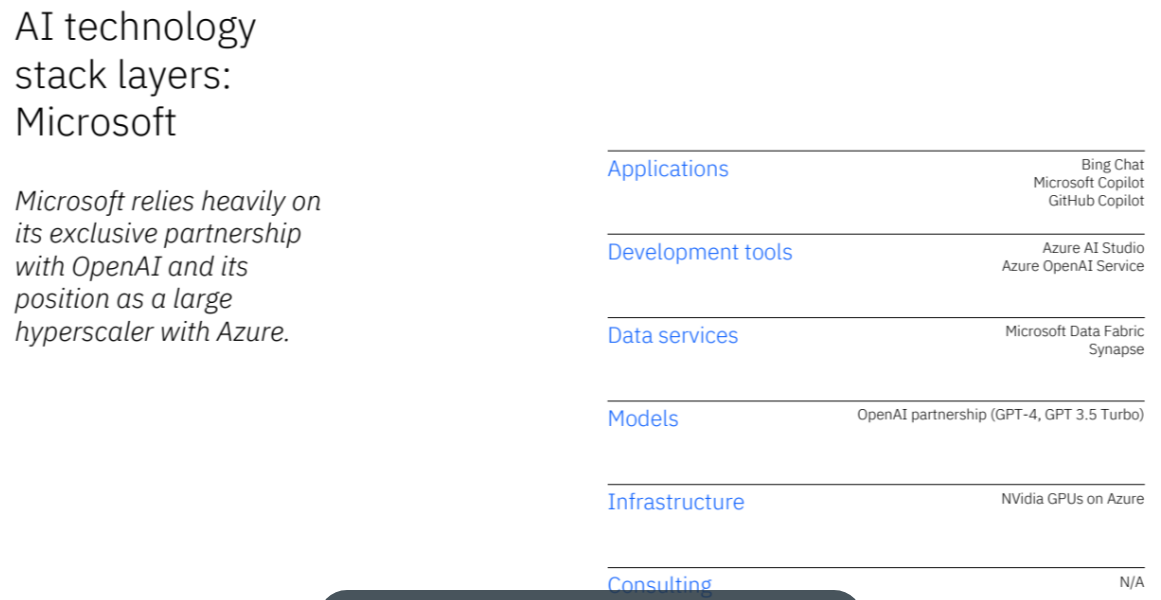
# IBM Cloud Pak for DATA

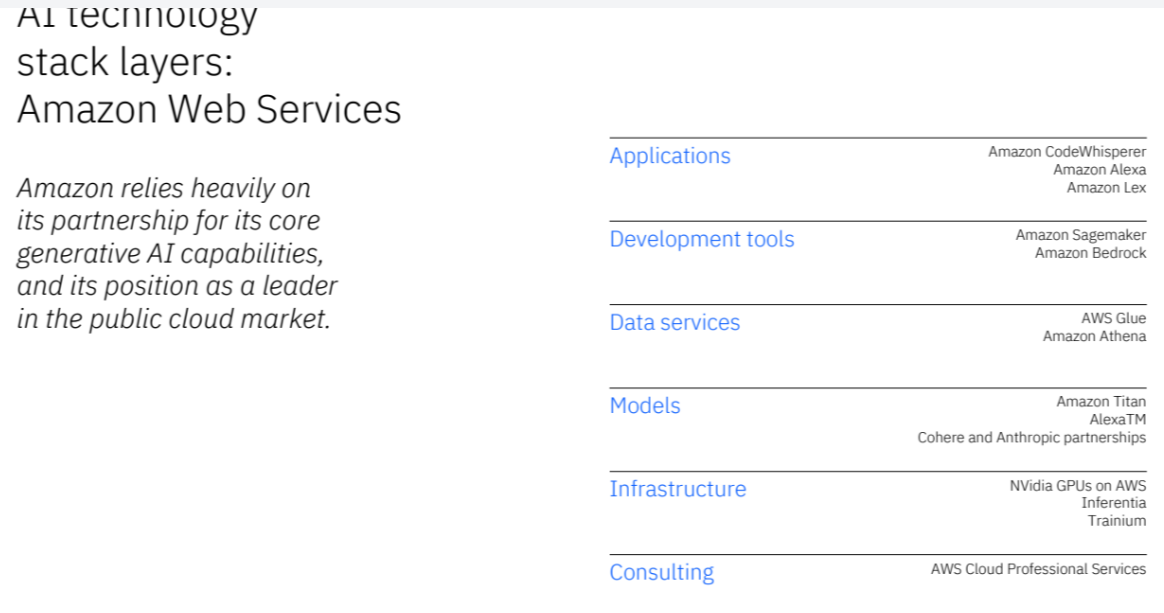


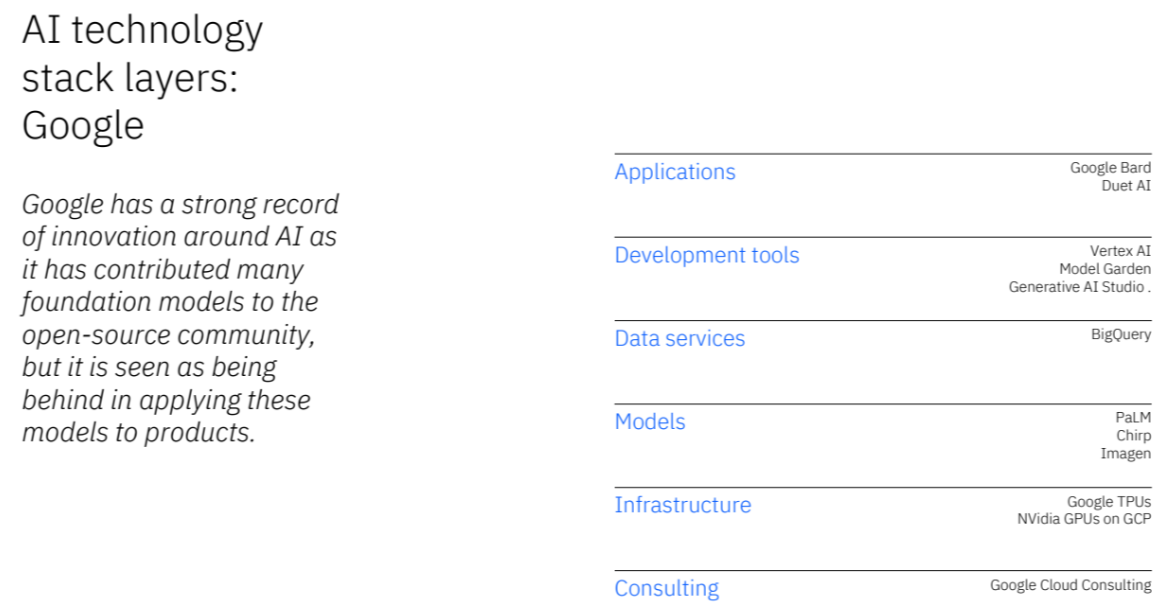












<https://youtu.be/1I6bQ12VxV0>

<https://youtu.be/hfIUstzHs9A>

<https://youtu.be/s4r5gXdSVPM>

<https://youtu.be/yu27PWzJI_Y>

