

EMPOWERING AI INNOVATION WITH MICROSOFT AZURE AI FOUNDRY

Naveen Kumar M



ABOUT ME

Naveen Kumar M



- Data Engineer
- 13 + Years of Experience
- 2 x C# Corner MVP
- 4 x Azure Certified
- Technical Blogger
- Speaker & Mentor



Linked 



Naveen Kumar M

Data Engineer | SQL / ETL Lead |
C# Corner MVP



AGENDA

- Introduction to Azure AI Foundry
- Architecture
- How AI Foundry Works
- What can we do with Project and Hub
- When to use AI Foundry
- Model Catalog
- Inference Tasks
- Q& A

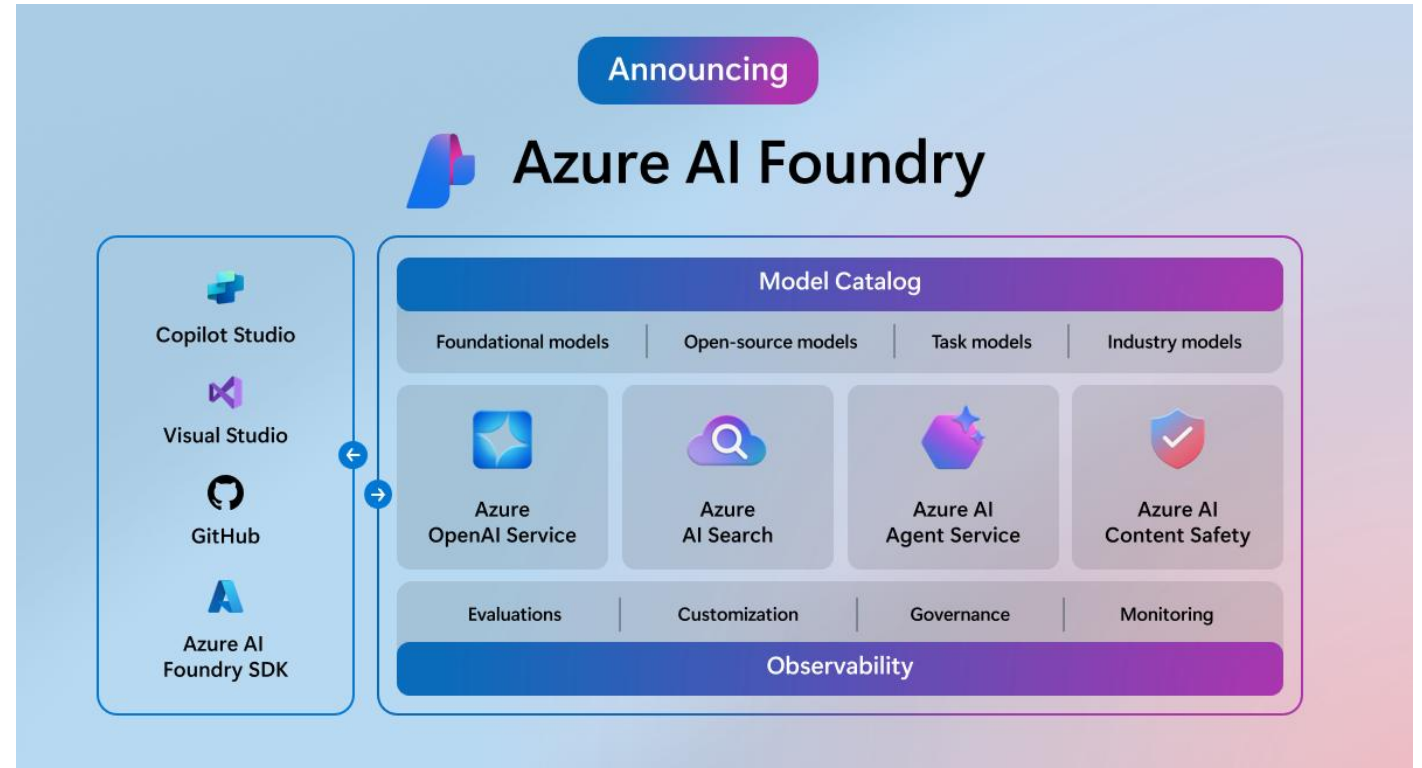
WHAT IS AZURE AI FOUNDRY

Azure AI Foundry

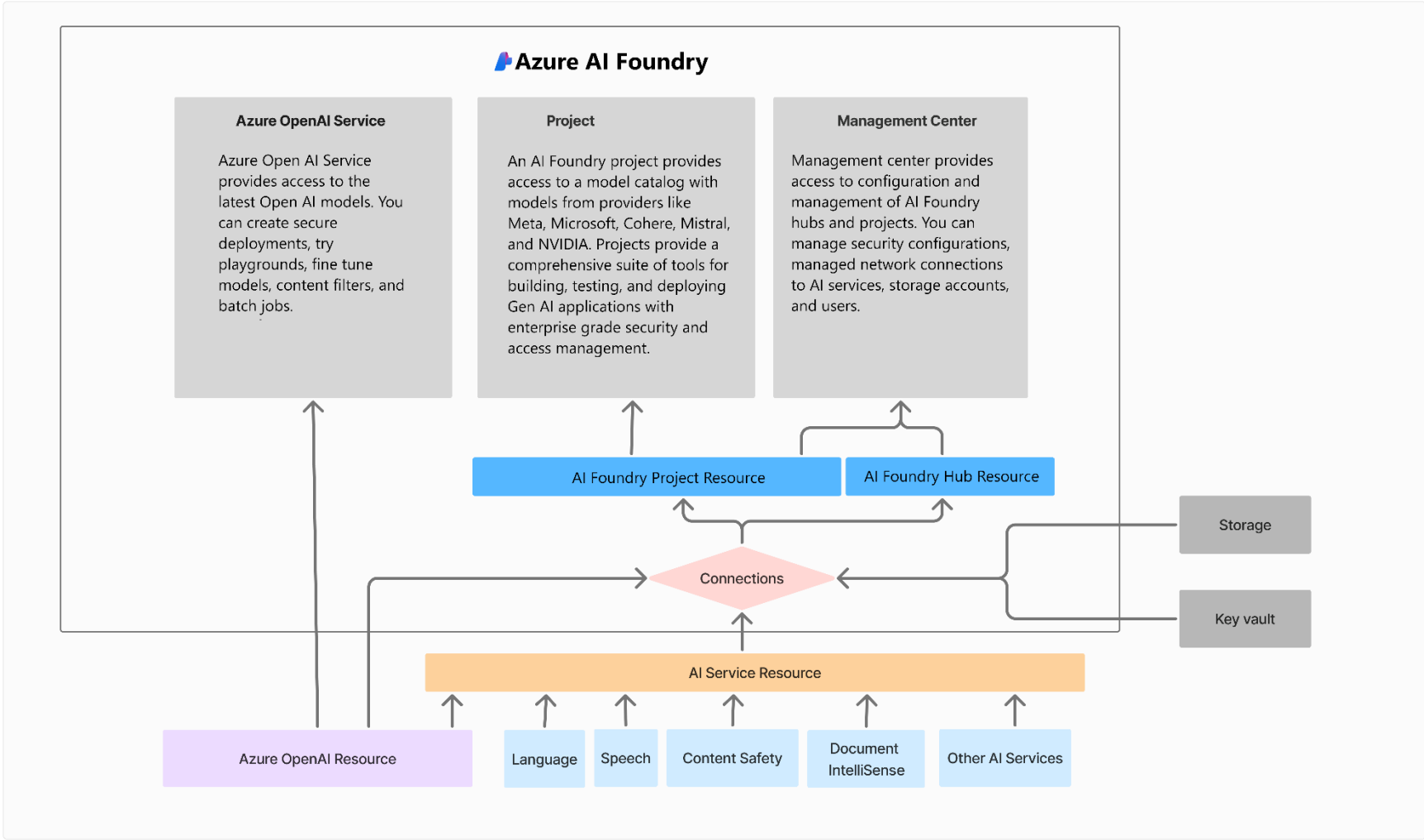
- All-in-one platform
- Formerly known as Azure AI Studio
- Designed for developers
- 1800+ pre built models

Capabilities

- Build generative AI apps on a secure platform.
- Use advanced AI tools and models with responsible AI practices.
- Collaborate throughout the development lifecycle.
- Explore diverse models, services, and features.



AZURE AI FOUNDRY ARCHITECTURE



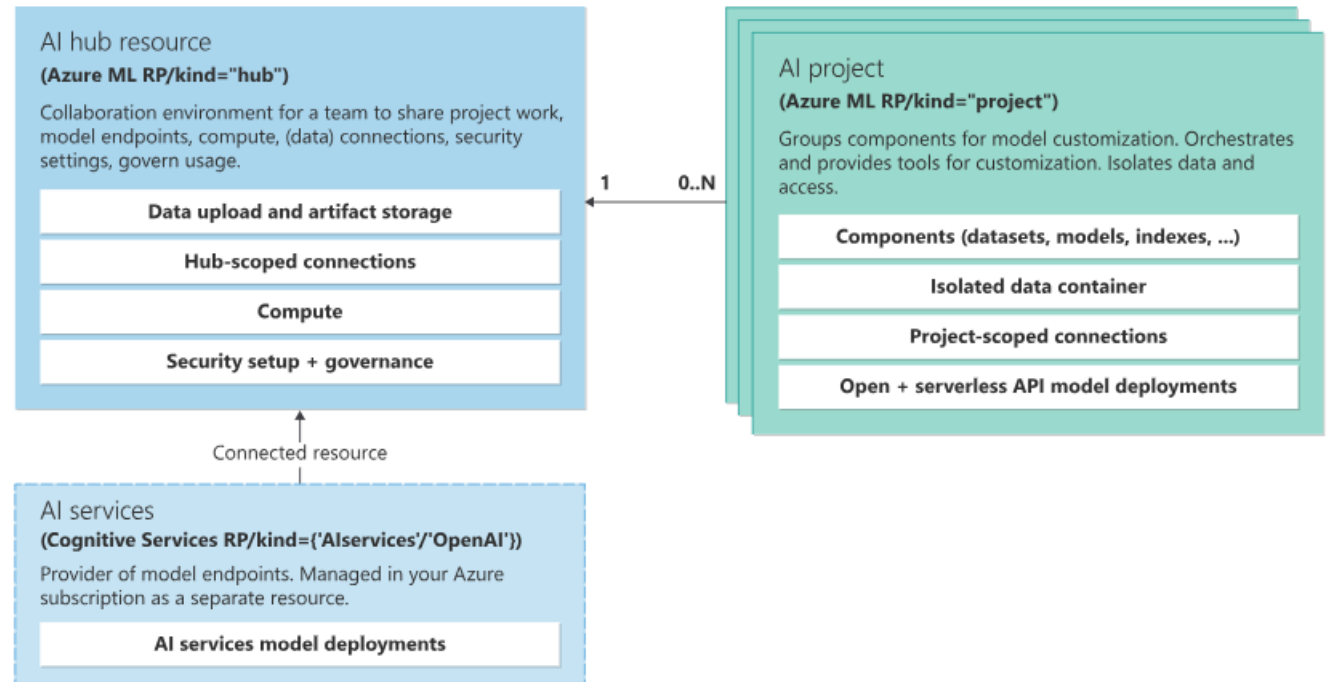
HOW AZURE AI FOUNDRY WORKS?

AI Foundry Hub

- Top-level resource in Azure AI Foundry.
- It provides the collaborative environment for a team to build and manage AI applications.

AI Foundry Project

- A project is a child resource of the hub.
- It inherits the hub's connections, and compute resources.
- It serves as an organizational container with tools for AI customization and orchestration.



Resource Provider: Microsoft.MachineLearningServices/workspaces

WHAT CAN WE DO WITH HUB?

Azure Foundry Hub

- Manage members and assign roles.
- Set up and control compute instances for experiments and custom code.
- Connect to resources like data stores, GitHub, and Azure AI Search.
- Define policies, such as automatic compute shutdown.

AI hub resource

(Azure ML RP/kind="hub")

Collaboration environment for a team to share project work, model endpoints, compute, (data) connections, security settings, govern usage.

Data upload and artifact storage

Hub-scoped connections

Base model endpoints

Compute

Security setup + governance

WHAT CAN WE DO WITH PROJECT?

Azure Foundry Project

- Deploy large language models for Chatbots or copilots.
- Test models in the chat playground.
- Augment prompts with your own data.
- Use prompt flow to combine models, prompts, and code.
- Evaluate model responses.
- Manage indexes and datasets for custom data.
- Set up content filters to prevent harmful responses.
- Use Visual Studio Code in the browser for custom coding.
- Deploy solutions as web apps or containerized services.

AI project

(Azure ML RP/kind="project")

Groups components for model customization. Orchestrates and provides tools for customization. Isolates data and access.

Components (datasets, models, indexes, ...)

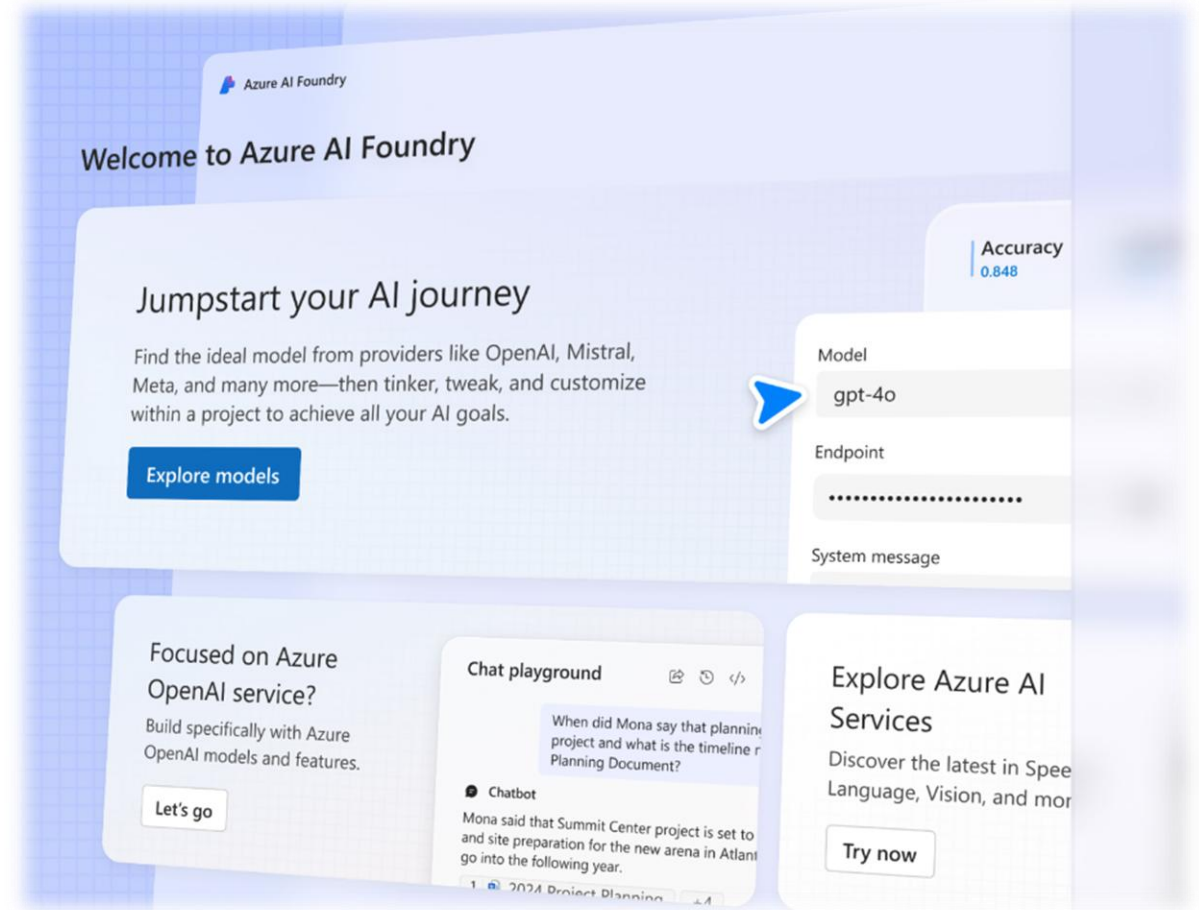
Isolated data container

Project-scoped connections

Open source + finetuned model endpoints

WHAT TO USE AZURE AI FOUNDRY?

- Create and manage AI projects
- Develop generative AI applications
- Explore available AI models
- Leverage Retrieval Augmented Generation (RAG)
- Monitor and evaluate AI models
- Integrate with Azure services
- Build responsibly





navin [redacted] ✎

? Help

Add a project description (optional) ✎

Overview

Model catalog

Playgrounds

AI Services

Build and customize ^

Agents PREVIEW

</> Code PREVIEW

Fine-tuning

Prompt flow

Assess and improve ^

Tracing PREVIEW

Evaluation

Safety + security

My assets ^

Models + endpoints

Data + indexes

Web apps

→ Management center

Endpoints and keys

[View all endpoints](#)

API Key

.....



Included capabilities

Azure AI inference

Azure OpenAI
Service

Azure AI Services

Use the following endpoint to call all your deployed base models:

Azure AI model inference endpoint PREVIEW

[redacted]



{x} API documentation

Project details

Project connection string

eastus2.api [redacted] 📄

Subscription

Nav- [redacted] 📄

Subscription ID

[redacted] 📄

Location

eastus2 📄

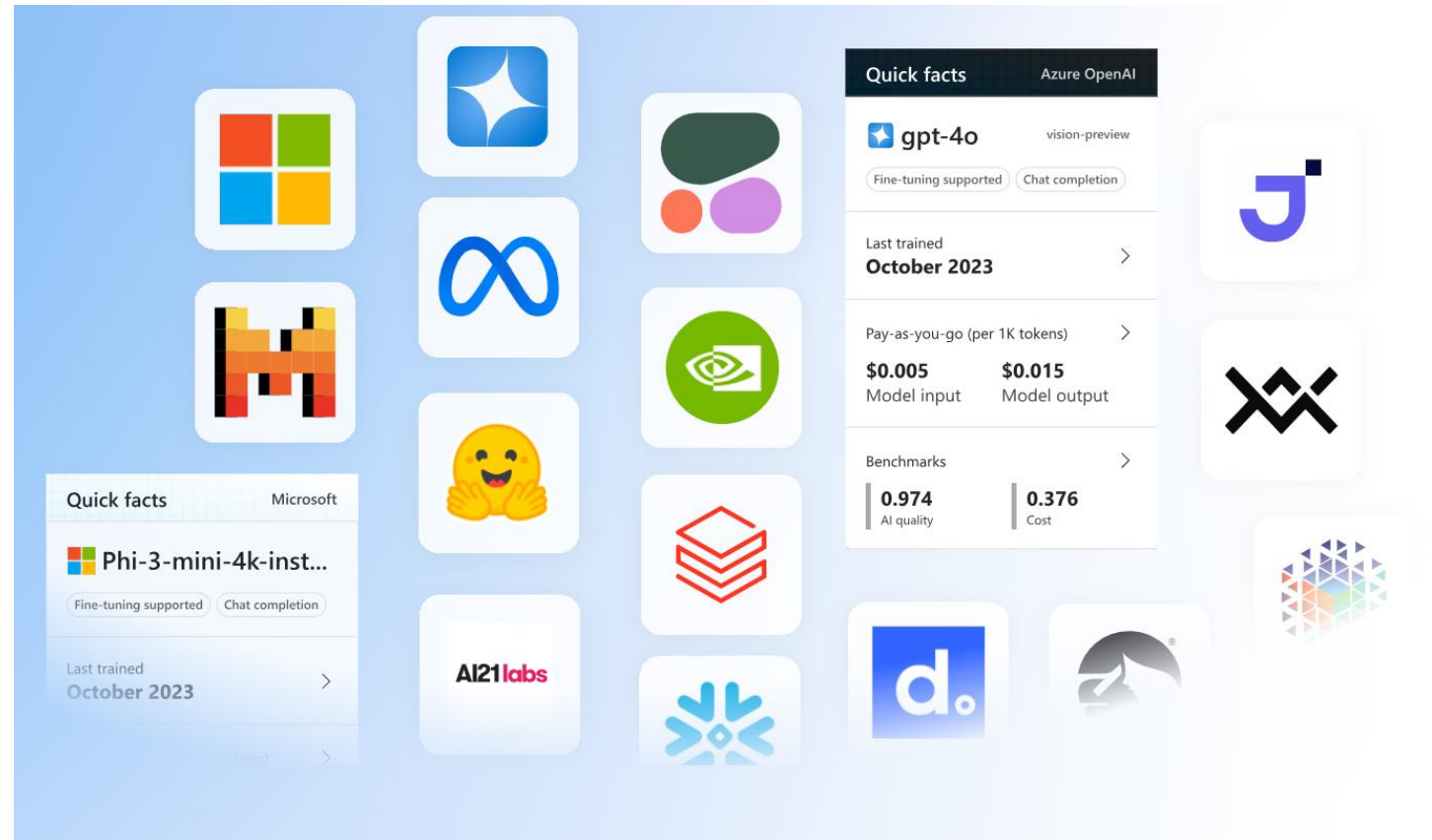
Manage project settings

- Add users
- View quota
- Connect resources
- Track costs

Open in management center

MODEL CATALOG


- Find right model to build custom solution
- 1820 Models – Jan 2025
- Models From:
 - Microsoft
 - Mistral
 - Meta
 - OpenAI
 - Hugging Face
 - NVIDIA
 - And more...




INFERENCE TASKS

- Chat completion
- Summarization
- Zero shot classification
- Conversational
- Embedding's
- Text to image
- Table question answering
- Image classification
- Text Classification
- Text to text generation
- Translation
- Token classification
- Question Answering
- Text generation
- Fill mask
- Object detection
- Image segmentation
- Speech recognition








Tech Time with Naveen

@ttwithnaveen · 204 subscribers · 115 videos

Tech Time with Naveen: Your Ultimate Destination for New Technologies and Career Growth...more


[linkedin.com/in/naveenkumarm17](#) and 2 more links

Customize channelManage videos




HomeVideosShortsPlaylistsCommunity

LatestPopularOldest




Festive Tech Calendar 18-12-2024:How to Build Personalized Chatbots Using Azure AI

19 views · 1 month ago




Create Your Own Chatbot With Azure AI Service | Hands-On Demo #chatbot #azure...

141 views · 2 months ago



Crafting Custom Copilots with Azure AI Studio | Hands-On Demo #copilot #azure #ai

71 views · 2 months ago



Introducing The DP-700 Fabric Data Engineer Certification - A Game-changer!

275 views · 3 months ago

@ttwithnaveen