

# Azure AI Foundry



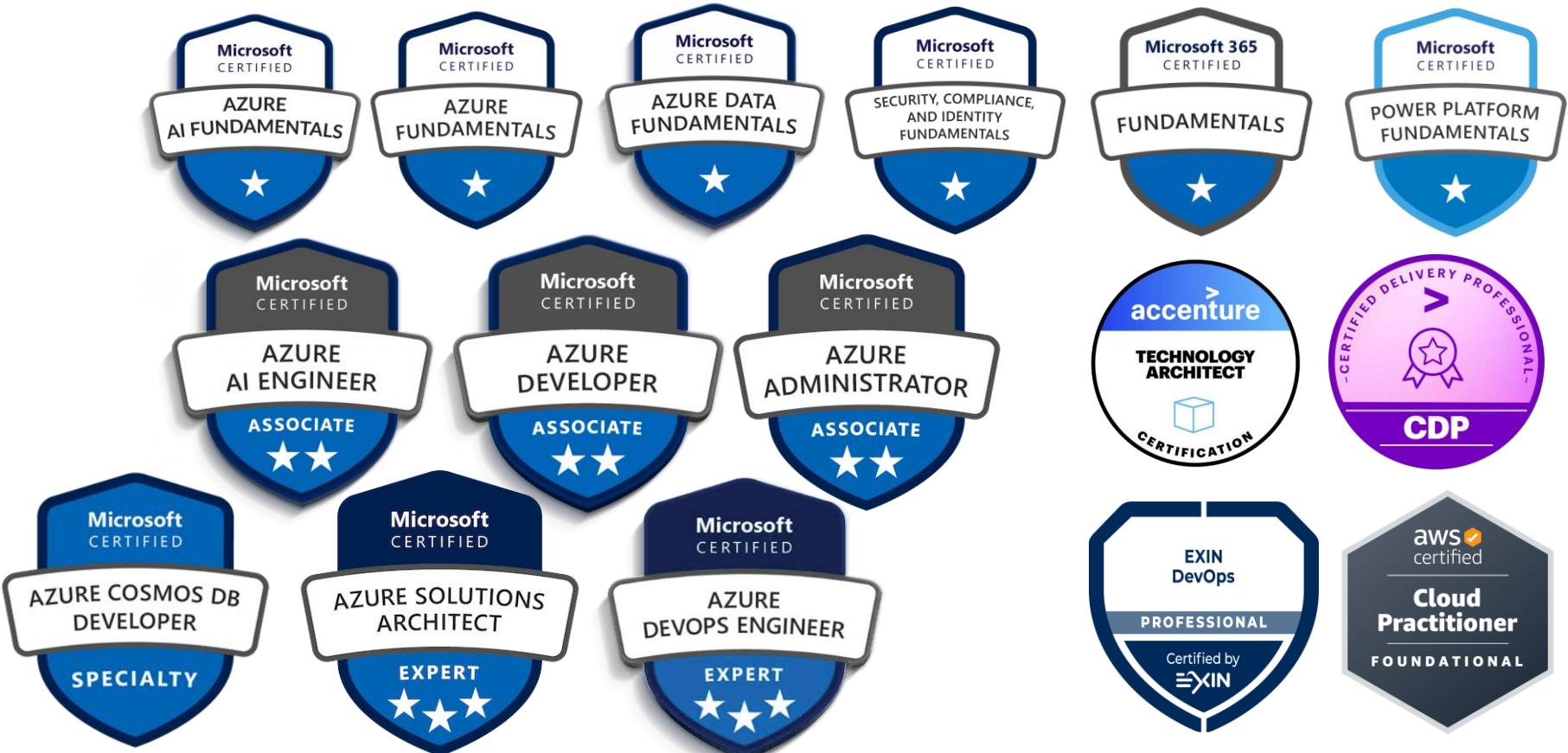
# Felipe Pimentel Augusto

- Formado em Ciências da Computação @ UGF
- Pós Graduado em MIT Arquitetura de Sistemas @ INFNET
- Aluno de Mestrado @ UTFPR
- Eterno Estudante e entusiasta de tecnologias
- +20 anos de carreira ligadas à desenvolvimento, arquitetura e liderança técnica
- Technical Architecture Manager @ Avanade
- Especialista em .NET, Azure, GitHub, Containers, EDA (Event-Driven Architecture) e outras coisas...
- Possuo algumas certificações em Azure, AWS, GitHub e outros vendors
- Apaixonado por sistemas críticos e de alta performance
- Experiência em indústrias de Health, Financial, Resources e Education
- Microsoft MVP em Developer Technology, .NET
- Host no canal DEPLOY @ Youtube



**Microsoft®**  
Most Valuable  
Professional









 **open-ai-azure-user-group-curitiba** ⋮ ☆ ... 

Azure OpenAI

◊ « »

 Overview Go to Azure AI Foundry portal  Delete

 Activity log

 Access control (IAM)

 Tags

**i Want to try the latest industry models and Agents ?**

Azure OpenAI is limited to OpenAI models. Upgrading your resource type to AI Foundry gives you access to Open AI plus third-party models like Meta , Mistral , xAI and more. Build enterprise grade AI agents using your data and tools.

[Get Started](#)





Se você acha que IA é coisa do Futuro ...

Eu te afirmo com toda certeza que o futuro é agora!



# Categorias de modelos no Azure AI Foundry

O catálogo do Azure AI Foundry é organizado em duas grandes categorias:

## **Modelos vendidos diretamente pelo Azure**

Hospedados e suportados pela Microsoft, com integração profunda ao ecossistema Azure e SLAs empresariais.

Incluem modelos do **Azure OpenAI** (como GPT-4, GPT-4o, GPT-5, Codex) e modelos proprietários da Microsoft.

Garantem conformidade e segurança para ambientes corporativos e nuvens soberanas.

## **Modelos de parceiros e comunidade**

Modelos de provedores como **Meta (Llama)**, **Mistral**, **Cohere**, **DeepSeek**, **Hugging Face**, **NVIDIA**, entre outros.

Oferecem diversidade e inovação, mas podem ter restrições para ambientes com requisitos de compliance rigorosos



# Principais tipos de modelos no catálogo

O Azure AI Foundry disponibiliza mais de **11.000 modelos** (em constante expansão), agrupados por finalidade:

## 1. Voz

### TTS (Text-to-Speech)

Converte texto em fala natural.

#### Exemplos:

gpt-audio (Azure OpenAI)

Modelos de parceiros como Whisper TTS

**Casos de uso:** Assistentes virtuais, leitura de conteúdo, acessibilidade.

### STT (Speech-to-Text)

Transcreve áudio para texto.

#### Exemplos:

Whisper STT

Modelos multimodais com suporte a áudio

**Casos de uso:** Transcrição de reuniões, legendas automáticas.



# Principais tipos de modelos no catálogo

## 2. Embeddings

**Função:** Representar texto ou outros dados em vetores numéricos para busca semântica, RAG e classificação.

**Exemplos:**

text-embedding-ada-002 (Azure OpenAI)

Modelos de parceiros como Cohere embeddings

**Casos de uso:** Pesquisa semântica, recomendação, recuperação de contexto.



# **Principais tipos de modelos no catálogo**

## **3. Chat / Modelos Generativos**

### **Chat Completions**

Modelos de linguagem para diálogo, geração de texto e raciocínio.

#### **Exemplos:**

GPT-4, GPT-4o, GPT-5 (Azure OpenAI)

Llama 2, Mistral

**Casos de uso:** Atendimento ao cliente, copilots internos, automação de processos.



# **Principais tipos de modelos no catálogo**

## **4. Multimodais**

**Função:** Processar texto, imagem, áudio e vídeo.

**Exemplos:**

GPT-4o multimodal

Sora-2 (geração de vídeo)

**Casos de uso:** Criação de conteúdo, análise de imagens, geração de mídia.



# Tipos de Prompts

System: Como sua IA vai se comportar

User: Sua pergunta

Assistent: A resposta da IA

**Histórico: Todas as mensagems de todos envolvidos.**



## **Model lifecycle stages**

Os modelos no catálogo de modelos pertencem a uma dessas etapas :

- Preview
- Generally available
- Legacy
- Deprecated
- Retired



# Preview

Os modelos com a etiqueta "Preview" são experimentais.

Os pesos, o tempo de execução e o esquema da API de um modelo podem mudar enquanto ele estiver em pré-visualização.

Não há garantia de que os modelos em pré-visualização se tornarão disponíveis para o público em geral. Os modelos em pré-visualização têm a etiqueta "Prévia" ao lado do nome no catálogo de modelos.



## Generally available

Esta etapa é a etapa padrão do modelo. Modelos que não incluem um rótulo de ciclo de vida ao lado do nome geralmente estão disponíveis e adequados para uso em ambientes de produção. Nesta etapa, os pesos e APIs do modelo são fixos. No entanto, contêineres de modelo ou tempos de execução com vulnerabilidades podem ser corrigidos, mas as correções não afetam as saídas do modelo.



# Legacy

Os modelos marcados como **Legacy** (Legado) destinam-se à descontinuação. Você deve planejar a migração para um modelo diferente, como um modelo novo e aprimorado que possa estar disponível na mesma família de modelos. Enquanto um modelo estiver no estágio "Legado", as implantações existentes do modelo continuarão funcionando e você poderá criar novas implantações do modelo até a data de descontinuação.



# **Deprecated**

Os modelos marcados como **Deprecated** (Obsoleto) não estão mais disponíveis para novas implantações. Você não pode criar novas implantações para o modelo; no entanto, as implantações existentes continuarão funcionando até a data de desativação.



## **Retired**

Os modelos marcados como **Retired** (Desativados) não estão mais disponíveis para uso. Você não pode criar novas implantações e as tentativas de usar implantações existentes retornam erros <código de retorno> .



