

Rome .NET Conference 2025



Sviluppo AI in .NET:
un'unica API per OpenAI,
Azure OpenAI, GitHub
models....e non solo!!!



Massimo Bonanni

Senior Technical Trainer @ Microsoft



Sponsor

Platinum Sponsor



Gold Sponsor



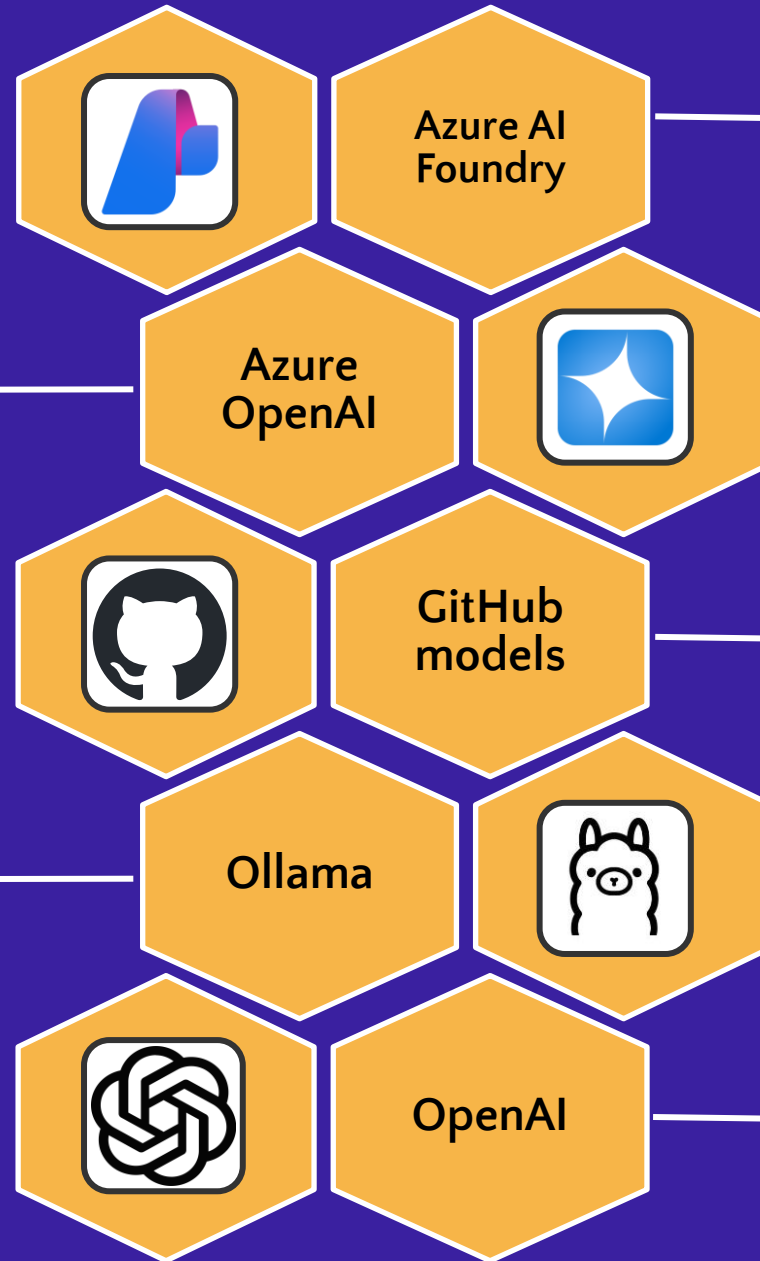
Technical Sponsor



Model Hosting

Azure OpenAI Service integrates OpenAI's advanced models, like GPT-4 and DALL-E, with Azure's enterprise-grade capabilities. It enables secure, scalable AI solutions for tasks such as content generation, coding, and data analysis, empowering businesses to innovate responsibly.

Ollama is an open-source tool for managing large language models (LLMs) locally. It ensures data privacy, reduces latency, and offers customization options for AI tasks. With features like a model library and local execution, it's ideal for developers prioritizing security and efficiency in AI projects.



Azure AI Foundry is a platform for enterprise AI solutions, offering tools, prebuilt models, and APIs for seamless AI development and deployment. Integrated with Azure services, it ensures scalability, efficiency, and responsible AI practices for transforming innovative ideas into reality.

GitHub Models empower developers to integrate AI into their workflows seamlessly. With access to cutting-edge models and tools, it enables experimentation, prototyping, and deployment directly within GitHub. This platform fosters innovation while ensuring responsible AI practices and privacy.

OpenAI is a leading AI research organization that develops advanced models like GPT and DALL-E, aiming to ensure AI benefits humanity. Known for breakthroughs in natural language processing and creativity, OpenAI empowers users with innovative tools for diverse applications, including communication, design, and problem-solving.

Model Hosting

Azure OpenAI Service integrates OpenAI's advanced models, like GPT-4 and DALL-E, with Azure's enterprise-grade cloud infrastructure. It enables secure, scalable AI solutions for tasks such as content generation, customer support, and data analysis, empowering businesses to innovate responsibly.

Ollama is a platform for managing large language models locally. It ensures low latency, and offers a simple interface for AI tasks. Ollama includes a library and local execution, allowing developers prioritize efficiency.

Azure AI Foundry

Azure AI Foundry is a platform for enterprise AI solutions, offering tools, prebuilt models, and APIs for seamless AI development and deployment. Integrated security measures ensure scalability, and it promotes responsible AI practices for turning ideas into reality.

GitHub

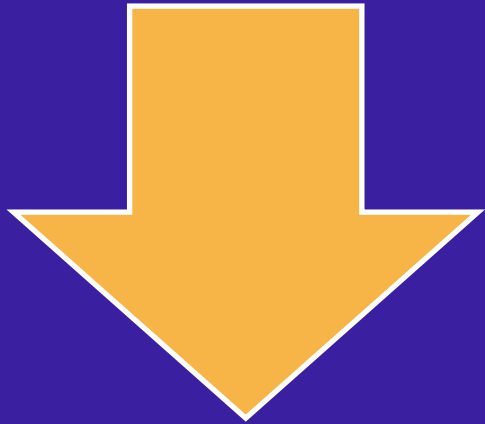
GitHub enables developers to streamline workflows and access cutting-edge AI models through AI Copilot, a code completion tool. It supports prototyping, and integration within GitHub. This facilitates collaboration while ensuring security and privacy.

Ollama

OpenAI research develops advanced models like GPT-4 and DALL-E, aiming to make AI benefits humanity. Known for breakthroughs in natural language processing, OpenAI provides innovative tools for diverse applications, including communication, design, and problem-solving.



What is Microsoft.Extensions.* ?



A modern framework **should** provide the fundamental building blocks for application development.

Microsoft.Extensions.Logging
→ ILogger

Microsoft.Extensions.Caching.*
→ IMemoryCache, IDistributedCache, ...

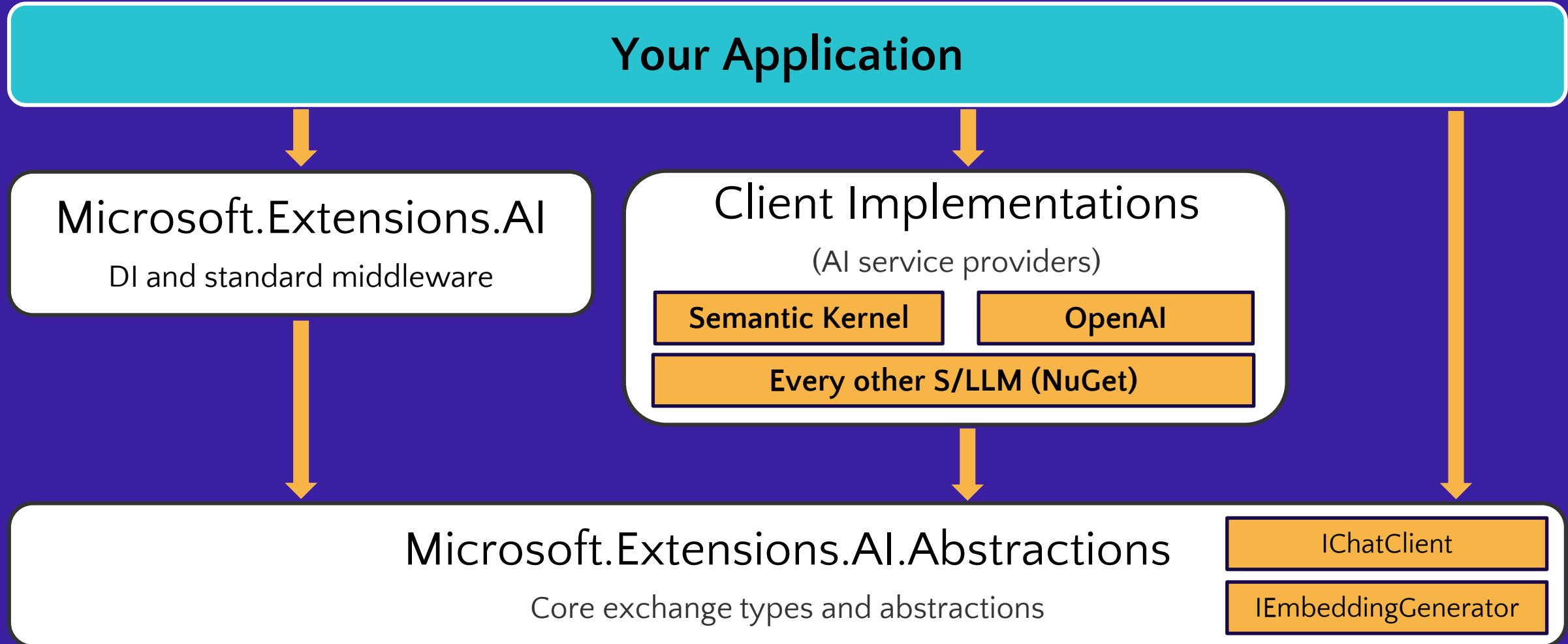
Microsoft.Extensions.DependencyInjection
→ IServiceCollection

A modern framework **should** allow you to use your desired implementation instead of the provided implementation.

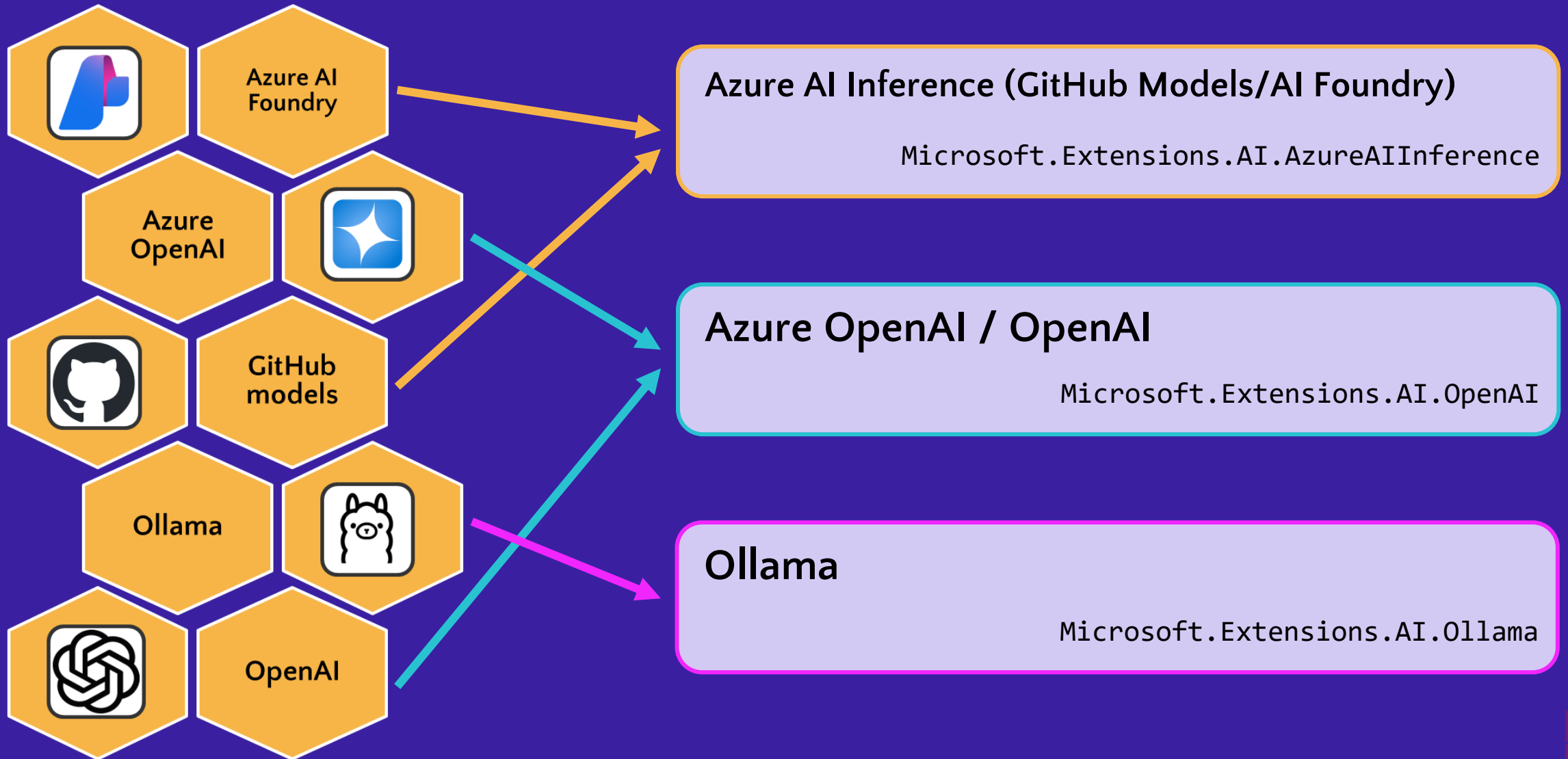


Microsoft.Extensions.AI
→ IChatClient,
IEmbeddingGenerator

Microsoft.Extensions.AI architecture



Client Implementations



Demo

Embedding
Semantic Search
Dependency Injection
Middleware
Functions



GitHub Models: Enhancing AI Interactions

Available in
Free Account



Experiment with AI models for free using the **Model Playground**.



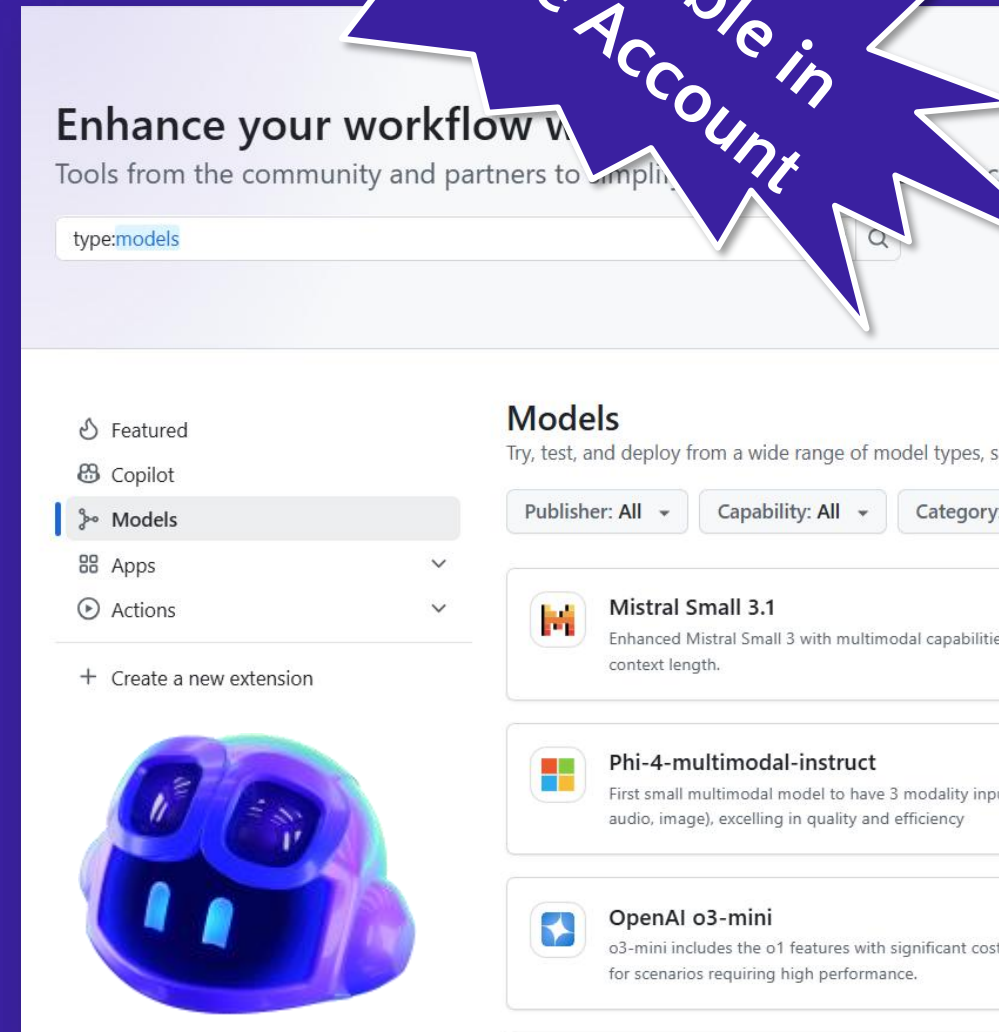
Compare responses from different models simultaneously.



Adjust parameters and test prompts to refine AI behavior.

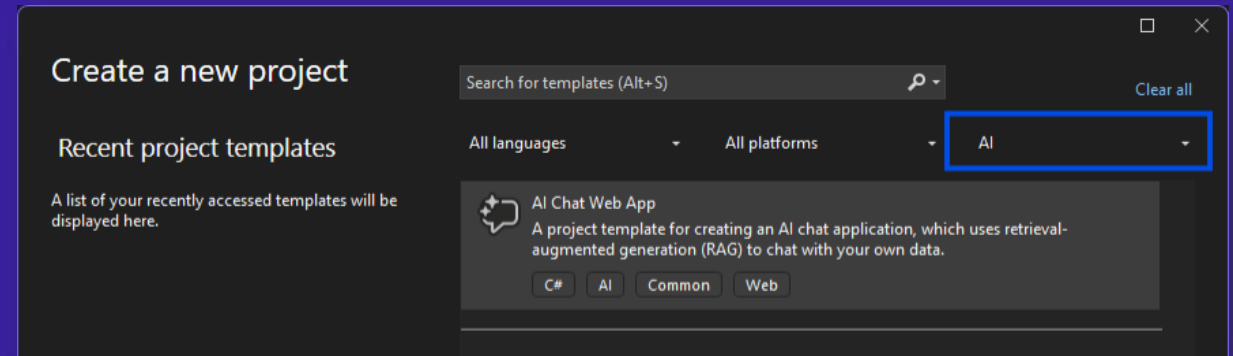


Utilize APIs to integrate AI models into applications.



.NET AI Template Now Available in Preview

- ✓ The **.NET AI Chat template** is designed to help you quickly build an AI-powered chat application to start chatting with custom data.
- ✓ This initial release focuses on a Blazor-based web app, built using the **Microsoft.Extensions.AI** and **Microsoft.Extensions.VectorData** abstractions.
- ✓ The template uses the **Retrieval Augmented Generation (RAG)** pattern commonly used for chat applications.



Start building with Microsoft.Extensions.AI

Library Developers

If you own libraries that provide clients for AI services, consider implementing the interfaces in your libraries.

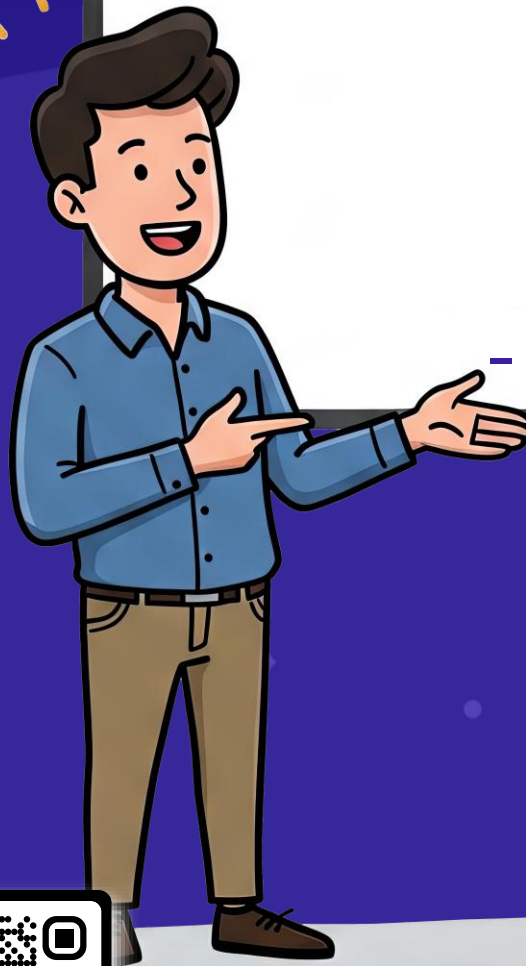
Service Consumers/ Application Developers

Use the abstractions instead of hardcoding to a specific AI service. This enables portability across models and services, facilitates testing and mocking, leverages middleware provided by the ecosystem, and maintains a consistent API throughout your app.

Ecosystem Contributors

Contribute to the ecosystem writing custom middleware components or clients.

Any Questions?



- Vote for my session -



Massimo Bonanni

Senior Technical Trainer

massimo.bonanni@microsoft.com

@massimobonanni



References



- [Introducing Microsoft.Extensions.AI Preview – Unified AI Building Blocks for .NET – .NET Blog](#)
- [Unified AI building blocks for .NET – .NET | Microsoft Learn](#)
- [Models · GitHub Marketplace](#)
- [AI Building Blocks – A new, unified AI layer \(.NET Conf 2024\)](#)
- [.NET AI Template Now Available in Preview – .NET Blog](#)