

MAI-thinking-1 model card

Direct from Azure models

Direct from Azure models are a select portfolio curated for their market-differentiated capabilities:

- **Secure and managed by Microsoft:** Purchase and manage models directly through Azure with a single license, consistent support, and no third-party dependencies, backed by Azure's enterprise-grade infrastructure.
- **Streamlined operations:** Benefit from unified billing, governance, and seamless PTU portability across models hosted on Azure - all as part of one Microsoft Foundry platform.
- **Future-ready flexibility:** Access the latest models as they become available, and easily test, deploy, or switch between them within Microsoft Foundry; reducing integration effort.
- **Cost control and optimization:** Scale on demand with pay-as-you-go flexibility or reserve PTUs for predictable performance and savings.

Learn more about [Direct from Azure models](#).

Key capabilities

About this model

MAI-Thinking-1 is Microsoft AI's reasoning model. Built from scratch for serious math, coding, and real-world enterprise deployment, MAI-Thinking-1 is a medium-sized model that stands among the strongest in its class.

Key model capabilities

- **Reasoning:** Competitive with leading models in its weight class on advanced math and coding.

- **Agentic coding:** Strong software engineering performance and trained using 8M+ RLE environments.
- **Enterprise ready:** Built on top of clean data that is appropriately licensed, which allows for quality, provenance, and control.
- **Seamless migration:** Built on top of the widely used Chat Completions API, making migrations easy.

Use cases

- **Enterprise deployments:** Clean data provenance, 256K context window, function calling, complex instruction-following.
- **Coding workflows:** Reading code, editing files, running tests, bug fixing, observing failures, and recovering from intermediate mistakes.
- **Human-in-the-loop workflows:** Adapts to feedback and calibrates to user intent—helpful and safe across the full range of tasks.

Pricing

Pricing is based on a number of factors, including deployment type and tokens used. [See pricing details here.](#)

Technical specs

Please refer to the MAI-thinking-1 technical paper.

Training cut-off date

July 2025

Training time

The provider has not supplied this information.

Input formats

Text

Output formats

The provider has not supplied this information.

Supported language

The provider has not supplied this information.

Sample JSON response

Please refer to Azure documentation on how to sample JSON responses, code snippets, and more.

Model architecture

Please refer to the MAI-thinking-1 technical paper.

Long context

MAI-thinking-1 has a 256K context window.

Optimizing model performance

Please refer to the MAI-thinking-1 technical paper.

Additional assets

The provider has not supplied this information.

Distribution channels

- **Azure Foundry:** for developers interested in deploying enterprise use cases, MAI-1-Thinking is available via Azure Foundry's Private Preview. Customers can request early access through the MAI-Thinking-1 interest form [here](#).

Responsible AI considerations

Safety techniques

Built according to Humanist Superintelligence principles: advanced AI capabilities designed to serve people and organizations, not replace them. The MAI-thinking-1 technical paper dives into extensive detail on evaluations for model safety and alignment.

Acceptable use

Acceptable use policy

The provider has not supplied this information.

Quality and performance evaluations

Please refer to the MAI-thinking-1 technical paper.

Benchmarking methodology

Please refer to the MAI-thinking-1 technical paper.

Public data summary

The provider has not supplied this information.