

# Phi-4

Microsoft Research

January 2025

## Introduction

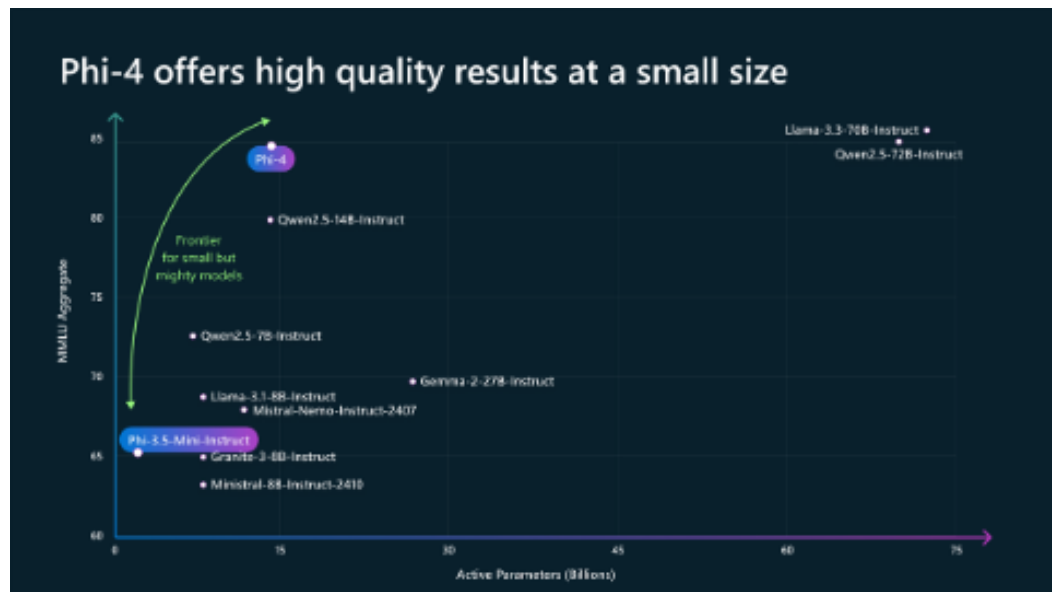
Phi-4 is Microsoft's newest small language model (SLM) specializing in complex reasoning tasks.

## New Release Highlights

We recently introduced Phi-4, our 14B parameter state-of-the-art SLM that excels at complex reasoning in areas such as math, in addition to conventional language processing. Phi-4 advances performance of SLMs by introducing innovative synthetic data generation methods for reasoning focused tasks, by optimizing the training curriculum and data mixture, and by introducing new techniques in post-training.

## Benchmark results

Phi-4 outperforms comparable and larger models on math related reasoning due to advancements throughout the processes, including the use of high-quality synthetic datasets, curation of high-quality organic data, and post-training innovations. Phi-4 continues to push the frontier of size vs quality. We highlight a couple benchmarks here. For additional results, please refer to the technical report released on arxiv [<https://arxiv.org/pdf/2412.08905>].



*Phi-4 performance on MMLU-pro: in this and many other benchmarks, phi-4 is at pareto frontier of performance vs number of parameters by a significant margin.*



*Phi-4 is particularly phenomenal at math for its size. For example, in this representative benchmark on competition math problems, Phi-4 outperforms much larger models, including Gemini Pro 1.5 (<https://maa.org/student-programs/amc/>). Note that this dataset was released after model training.*

## Scenarios Where This Tech Shines

Our model is designed for use as a small-but-mighty building block for generative AI powered features and to accelerate research on language models. The scenarios where this technology (phi-4) shines can be summarized as follows:

1. **Memory/Compute Constrained Environments:** Phi-4 can be served on single GPU which significantly broadens its usability across resource-limited settings.
2. **Latency Bound Applications:** Phi-4's small size ensures low latency, making it suitable for real-time or near-real-time applications where responsiveness is key.
3. **Tasks Requiring Strong Reasoning and Logic:** Among other features, phi-4 has been intentionally optimized for complex reasoning tasks, achieving performance on par with much larger models.

## Learn More

Phi-4 model weights are currently available on Huggingface and Azure AI Foundry under an MIT License.

- [microsoft/phi-4 · Hugging Face](#)
- [Model catalog - Azure AI Foundry](#)
- Phi-4 Blog: [Introducing Phi-4: Microsoft's Newest Small Language Model Specializing in Complex Reasoning | Microsoft Community Hub](#)
- Phi-4 Technical report: [2412.08905](#)