

Meta selects Azure as strategic cloud provider to advance AI innovation and deepen PyTorch collaboration

By Eric Boyd, CVP, Azure AI, Microsoft

Microsoft is committed to the responsible advancement of AI to enable every person and organization to achieve more. Over the last few months, we have talked about advancements in our Azure infrastructure, Cognitive Services, and Azure Machine Learning to make Azure better at supporting the AI needs of all our customers, regardless of their scale. Meanwhile, we also work closely with some of the leading research organizations around the world to empower them to build great AI.

Today we're thrilled to announce an expansion of our ongoing collaboration with Meta: Meta has selected Azure as a strategic cloud provider to accelerate AI research and development.

As part of this deeper relationship, Meta will expand its use of Azure's supercomputing power to accelerate AI research and development for its Meta AI group. Meta will utilize a dedicated Azure cluster of 5400 GPUs using the latest virtual machine (VM) series in Azure (NDm A100 v4 series, featuring NVIDIA A100 Tensor Core 80GB GPUs) for some of their large-scale AI research workloads. In 2021, Meta began using Azure VMs (NVIDIA A100 80GB GPUs) for some of its large-scale AI research after experiencing Azure's impressive performance and scale. With four times the GPU-to-GPU bandwidth between VMs compared to other public cloud offerings, the Azure platform enables faster distributed AI training. Meta used this, for example, to train their recent OPT-175B language model. The NDm A100 v4 VM series on Azure also gives customers the flexibility to configure clusters of any size automatically and dynamically from a few GPUs to thousands, and the ability to pause and resume during experimentation. Now, the Meta AI team is expanding their usage and bringing more cutting-edge ML training workloads to Azure to help further advance their leading AI research.

In addition, Meta and Microsoft will collaborate to scale [PyTorch adoption on Azure](#) and accelerate developers' journey from experimentation to production. Azure provides a comprehensive top to bottom stack for PyTorch users with best in class hardware (NDv4s and Infiniband). In the coming months, Microsoft will build new PyTorch development accelerators to facilitate rapid implementation of PyTorch-based solutions on Azure. Microsoft will also continue providing [enterprise-grade support for PyTorch](#) to enable customers and partners to deploy PyTorch models in production on both cloud and edge.

"We are excited to deepen our collaboration with Azure to advance Meta's AI research, innovation and open-source efforts in a way that benefits more developers around the world," Jerome Pesenti, VP of AI, Meta. "With Azure's compute power and 1.6 TB/s of interconnect bandwidth per VM we are able to accelerate our ever-growing training demands to better accommodate larger and more innovative AI models. Additionally, we're happy to work with Microsoft in extending our experience to their customers using PyTorch in their journey from research to production."

By scaling Azure's supercomputing power to train large AI models for the world's leading research organizations, and by expanding tools and resources for open source collaboration and experimentation, we can help unlock new opportunities for developers and the broader tech community, and further our mission to empower every person and organization around the world.