



# GPT-in-a-Box Overview

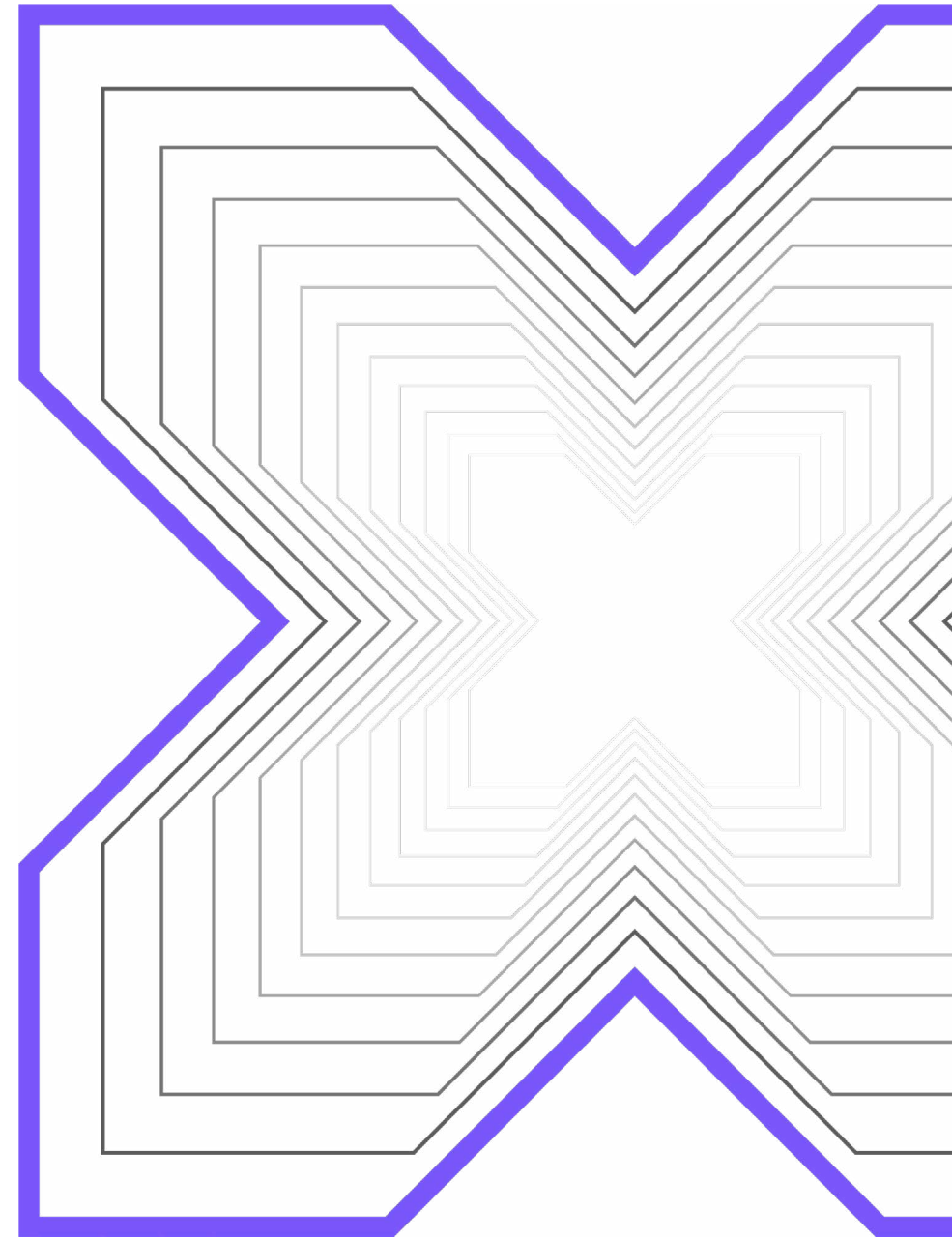
The Nutanix GPT-in-a-Box™ solution is a turnkey AI solution for organizations wanting to implement Generative Pre-trained Transformer (GPT) capabilities while maintaining control over their data and applications.

Nutanix GPT-in-a-Box addresses the complexity, scaling, and security challenges enterprises face when adopting generative AI and AI/ML applications. Nutanix GPT-in-a-Box is a full-stack software-defined AI-ready platform with services that simplify and jumpstart your initiatives from edge to core.

Nutanix GPT-in-a-Box equips you with the tools to streamline your infrastructure so it's ready for AI deployment. This includes the tools to maintain AI content exclusive to your organization to address security, privacy, and compliance requirements. Efficiently build, fine-tune, and run models, including GPTs and LLMs, while leveraging your existing teams and skills.

## Key Features

- **Freedom of Choice:** Bring-your-own-AI model lets you choose from popular models, including the Meta Llama 2, Falcon LLM and MosaicML MPT models.
- **Simplified Operations:** Bridge developer and data scientist self-service and governance needs.
- **Enterprise Data Services:** Protect AI models with the proven value of Nutanix data services.
- **Lower TCO:** Dynamically allocate resources from cloud to core to edge to deliver efficient and timely value.



## GPT-in-a-Box Summary

Nutanix Cloud Platform Software	Planning, Design, and Deployment Services*	MLOps Stack
<p>Includes</p> <ul style="list-style-type: none"><li>Nutanix Cloud Infrastructure Ultimate</li><li>Nutanix Unified Storage Pro for 10 TB Objects and Files</li></ul> <p>Rich data services, integrated virtualization with the Nutanix AHV® hypervisor, resilience and management features with infrastructure AIOps: Monitoring, planning, rightsizing, and low-code automation.</p> <p>Visit our software options page for a full list of features included with Nutanix Cloud Infrastructure Ultimate and Nutanix Unified Storage Pro.</p>	<p>Includes</p> <ul style="list-style-type: none"><li><b>Planning Workshop:</b> The Nutanix Artificial Intelligence/Machine Learning Planning Workshop offers IT teams in-depth and practical guidance to plan and implement AI/ML workloads based on a GPT model for on-premises Nutanix Cloud Infrastructure™ clusters. This workshop is beneficial during the planning stage of the Nutanix GPT-in-a-Box solution.</li><li><b>Stack Design Workshop:</b> The Nutanix Artificial Intelligence/Machine Learning Stack Design Workshop offers IT teams in-depth and practical guidance to create a comprehensive design for AI/ML workloads. This is based on a GPT model for on-premises Nutanix Cloud Infrastructure clusters. It covers various aspects such as performance, scalability, flexibility, integration, and operational needs. This workshop is beneficial during the design stage of the Nutanix GPT-in-a-Box solution.</li><li><b>Deployment:</b> The Nutanix Artificial Intelligence Machine Learning Deployment service accelerates the deployment of an AI/ML solution based on a GPT workload on the Nutanix Cloud Platform™ solution. A highly-skilled consultant deploys the AI compute instances and associated prerequisites to use the Nutanix Cloud Infrastructure cluster for inference. This offer is ideal for the deployment stage of the Nutanix GPT-in-a-Box solution.</li></ul>	<p>Includes</p> <ul style="list-style-type: none"><li><b>PyTorch® software:</b> a fully-featured framework for building generative AI models.</li><li><b>Kubeflow® MLOps toolset:</b> allows for deployments of ML workflows on Kubernetes® containers to provide simplicity, portability and scale.</li><li><b>Bring-your-own-AI model:</b> Choose from popular open-source models, including the Meta Llama 2™, Falcon LLM™, and MosaicML® MPT models.</li></ul>

\* Partners may offer equivalent deployment and migration services in place of Nutanix-offered services.



## GPT-in-a-Box Professional Services Options

Product Code	Service Description
CNS-DAY0-WRK-PLN-GPT/AI/ML	Planning Workshop
CNS-INF-A-WRK-DES-GPT/AI/ML	Stack Design Workshop
CNS-INF-A-SVC-DEP-GPT/AI/ML	Deployment

For more information, visit our [Professional Services: Service Descriptions Catalog](#).

Suggested Use Cases/Workload
<p><b>Common Across All Verticals:</b></p> <ul style="list-style-type: none"><li>• AI bots for customer service and technical support</li><li>• Document search and analysis, such as loans, contracts, research</li><li>• Code co-piloting</li><li>• Fraud and theft protection</li></ul> <p><b>Pharma/Healthcare:</b> Generative antibodies design, processing of OMICs data for drug discovery</p> <p><b>Financials:</b> Portfolio optimization, customer upsell analysis</p> <p><b>Retail/Edge:</b> Visual quality inspections, fraud detection</p> <p><b>Federal:</b> Mission co-pilot, simulations – digital twins, multi-agent systems</p>

For additional information, visit us at [Nutanix.com](https://www.nutanix.com) or contact your nearest [Nutanix partner](#).

**NUTANIX**

[info@nutanix.com](mailto:info@nutanix.com) | [www.nutanix.com](https://www.nutanix.com) | [@nutanix](https://twitter.com/nutanix)

©2023 Nutanix, Inc. All rights reserved. Nutanix, the Nutanix logo and all product and service names mentioned herein are registered trademarks or trademarks of Nutanix, Inc. in the United States and other countries. All other brand names mentioned herein are for identification purposes only and may be the trademarks of their respective holder(s). GPT-Kit-Nutanix-HCI-Starter-Kit-Customer-Version-2-FY24Q1-11082023