



## Nutanix Simplifies Adoption of Generative AI with New Nutanix GPT-in-a-Box Solution

August 15, 2023

*AI infrastructure solution accelerates customers' path to GPT and LLMs while keeping organizations in control of their data*

SAN JOSE, Calif.--(BUSINESS WIRE)--Aug. 15, 2023-- [Nutanix](#) (NASDAQ: [NTNX](#)), a leader in hybrid multicloud computing, today announced the Nutanix GPT-in-a-Box™ solution for customers looking to jump-start their artificial intelligence (AI) and machine learning (ML) innovation, while maintaining control over their data. The new offering, available today, is a full-stack software-defined AI-ready platform, along with services to help organizations size and configure hardware and software infrastructure suitable to deploy a curated set of large language models (LLMs) using the leading open source AI and MLOps frameworks on the Nutanix Cloud Platform™. It allows customers to easily procure AI-ready infrastructure to fine-tune and run generative pre-trained transformers (GPT), including LLMs at the edge or in their datacenter.

Many enterprises are grappling with how to quickly, efficiently and securely take advantage of the power of generative AI and AI/ML applications, especially for use cases that cannot be run in the public cloud because of data sovereignty, governance and privacy concerns. New use cases emerge every day as organizations look to leverage generative AI to improve customer service, developer productivity, operational efficiency and more. From automated transcription of internal documents, to high-speed search of multimedia contents, and automated analysis, many organizations see the opportunity with AI but are struggling with growing concerns regarding intellectual property leakage, compliance and privacy. Additionally, organizations looking to build an AI-ready stack often struggle with how to best support ML administrators and data scientists, while the prospect of large AI investment costs has enterprises stalled in their AI and ML strategy.

"As customers look to design and deploy generative AI solutions, they find themselves struggling with balancing the deep expertise required to install, configure, and run these workloads with concerns around their data security and protecting company IP – all while controlling costs," said Greg Macatee, Senior Research Analyst, Infrastructure Systems, Platforms and Technologies Group at IDC. "With GPT-in-a-Box, Nutanix offers customers a turnkey, easy-to-use solution for their AI use cases, offering enterprises struggling with generative AI adoption an easier on-ramp to deployment."

The Nutanix GPT-in-a-Box solution delivers ready-to-use customer-controlled AI infrastructure for the edge or the core data center and allows customers to run and fine-tune AI and GPT models while maintaining control over their data. Nutanix provides a full complement of security and data protection offerings ideal for AI data protection.

"Helping customers tackle the biggest challenges they face in IT is at the core of what we do, from managing increasing multicloud complexity, to data protection challenges, and now adoption of generative AI solutions while keeping control over data privacy and compliance," said Thomas Cornely, SVP, Product Management at Nutanix. "Nutanix GPT-in-a-Box is an opinionated AI-ready stack that aims to solve the key challenges with generative AI adoption and help jump-start AI innovation."

This new solution includes:

- The Industry-leading Nutanix Cloud Infrastructure™ platform, with the Nutanix Files Storage™ and Objects Storage™ solutions, the Nutanix AHV® hypervisor and Kubernetes, along with NVIDIA GPU acceleration, which can be sized for large to small scale.
- Nutanix services to help customers size their cluster and deploy an opinionated stack with the leading open source deep learning and MLOps frameworks, inference server, and a curated set of large language models such as Llama2, Falcon and MPT.
- Ability for data scientists and ML administrators to immediately consume these models with their choice of applications, enhanced terminal UI, or standard CLI.
- The platform can also be leveraged to run other GPT models, as well as fine tune these models leveraging internal data, hosted on included Nutanix Files or Objects Storage services.

"Leveraging AI to more efficiently and effectively help our customers is a top priority for us but, as a regulated financial services organisation, maintaining full control over our data is necessary," said Jon Cosson, CISO at JM Finn. "The Nutanix Cloud Platform delivers the performance, flexibility and security required to safely deploy AI workloads."

The Nutanix GPT-In-a-Box solution builds on the full stack scalability, performance, resilience and ease of use that the Nutanix Cloud Platform is known for. Nutanix's expertise with scalable infrastructure across public cloud, datacenter and edge use cases delivers the ideal environment to fine-tune and run AI applications while maintaining control over the data. In fact, in a recent survey, 78% of Nutanix customers indicated they were likely to run their AI/ML workloads on the Nutanix Cloud Platform.

Nutanix's expertise and involvement in the open source AI community provide customers with a strong foundation on which to build their AI strategy. Key contributions include: participation in the MLCommons (AI standards) advisory board; co-founding and technical leadership in defining the ML Storage Benchmarks and Medicine Benchmarks; serving as a co-chair of the Kubeflow (MLOps) Training and AutoML working groups at the Cloud Native Computing Foundation (CNCF).

The Nutanix GPT-in-a-Box solution is available to customers today. Customers looking to get started can find more information [here](#).

Supporting Quotes:

- "Red Hat is helping organizations across various industries to accelerate business and mission-critical initiatives through the development and deployment of AI-enabled applications in the hybrid cloud. With this launch, joint customers can take

advantage of Red Hat OpenShift AI along with the Nutanix AI-ready stack to jump-start innovation on a consistent, scalable open source foundation while maintaining control over their data.”

- Sarwar Raza, Vice President, Cloud Services, Red Hat

- “Enterprises everywhere are looking to leverage the generative AI opportunity but many, especially in regulated industries, are concerned with potential IP and data security issues. TCS has a long track record of embracing technology innovations in a secured manner to satisfy customer demand and enable them to focus on business outcomes. Nutanix GPT-in-a-Box along with TCS Cognix, a future-ready service delivery platform, allows customers to realize the value of artificial intelligence and machine learning by providing service resiliency, business agility, and enhanced customer experience to organizations.”

- Dinanath Kholkar, SVP and Global Head of Partner Ecosystems & Alliances at TCS

- “Taking advantage of the potential of generative AI is one of the most significant opportunities available to global enterprises today, and finding the right partner to guide this journey is crucial. Nutanix GPT-in-a-Box, along with consulting and solutions from Hexaware, can help jump-start this AI journey by addressing some of the key talent, security, compliance, deployment and scale challenges they’re facing today.”

- Vinod Chandran, COO, Hexaware

- “Generative AI is top of mind for many enterprises, but most don't have the in house resources to fully take advantage of this opportunity. Nutanix GPT-in-a-Box, along with solutions from DKube, simplifies enterprise AI adoption with an AI-ready stack to jumpstart innovation.”

- Prasad Vellanki, CEO, DKube

- “Generative AI, including Chat-GPT and LLMs, are revolutionizing the way enterprises operate but many require support to best take advantage of this opportunity. Lamini makes enterprise LLMs easier while Nutanix delivers AI-ready infrastructure to help organizations simplify operations and maintain control over their data significantly simplifying adoption of generative AI.”

- Greg Diamos, Co-Founder, Lamini and ML Commons

#### **Additional Resources:**

- Blog: [The AI-Ready Stack: Nutanix Simplifies Your AI Innovation Learning Curve](#)
- [Nutanix AI Solution](#)
- [Kubeflow brings MLOps to the CNCF Incubator](#)
- [Announcing MedPerf Open Benchmarking Platform for Medical AI](#)
- [Debojyoti Dutta, Board Advisor, ML Commons](#)

#### **About Nutanix**

Nutanix is a global leader in cloud software, offering organizations a single platform for running apps and data across clouds. With Nutanix, companies can reduce complexity and simplify operations, freeing them to focus on their business outcomes. Building on its legacy as the pioneer of hyperconverged infrastructure, Nutanix is trusted by companies worldwide to power hybrid multicloud environments consistently, simply, and cost-effectively. Learn more at [www.nutanix.com](http://www.nutanix.com) or follow us on social media @nutanix.

© 2023 Nutanix, Inc. All rights reserved. Nutanix, the Nutanix logo, and all Nutanix product and service names mentioned herein are registered trademarks or unregistered trademarks of Nutanix, Inc. (“Nutanix”) in the United States and other countries. Other brand names or marks mentioned herein are for identification purposes only and may be the trademarks of their respective holder(s). This release may contain express and implied forward-looking statements, which are not historical facts and are instead based on Nutanix’s current expectations, estimates and beliefs. The accuracy of such statements involves risks and uncertainties and depends upon future events, including those that may be beyond Nutanix’s control, and actual results may differ materially and adversely from those anticipated or implied by such statements. Any forward-looking statements included herein speak only as of the date hereof and, except as required by law, Nutanix assumes no obligation to update or otherwise revise any of such forward-looking statements to reflect subsequent events or circumstances.

Red Hat, the Red Hat logo and OpenShift are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the U.S. and other countries.

View source version on [businesswire.com](https://www.businesswire.com/news/home/20230815076667/en/): <https://www.businesswire.com/news/home/20230815076667/en/>

#### **Media Contact:**

[pr@nutanix.com](mailto:pr@nutanix.com)

Source: Nutanix