



How to Achieve Faster Code and Content Creation with GenAI

Enterprises are launching AI initiatives in an effort to generate code and content faster but many still remain baffled about where to begin. Start by identifying code and content creation use cases that will streamline business operations and improve the efficiency of internal teams. Some examples may include:

- **Co-pilot for content.** Deploy AI tools to help generate better content in less time.
- **Co-pilot for code.** Increase software team productivity with the latest AI tools.
- **Intelligent document processing.** Leverage natural language processing to extract, classify and interpret information from documents and speed-up business processes.

Many organizations – especially those in regulated industries and the public sector – can't or prefer not to use shared, cloud-based services from vendors such as Microsoft, OpenAI or Google for reasons of cost, security and compliance.

But by supporting code and content use cases from private infrastructure that you control, you can deliver the capabilities employees need to be successful, while avoiding the risks created when well-meaning users upload sensitive or proprietary company data to public AI services.

Considerations

Use cases for code and content creation generally rely on pretrained GenAI models. But before you begin, you'll need make sure that you have the necessary infrastructure to support:

Inference for the latest foundation models from organizations like Meta, Microsoft, Mistral, and Google. Nutanix has partnered with Hugging Face, a popular resource for open-source models.

You may also need to support:

- Fine-tuning, a form of deep learning, of your chosen models using company data.
- Retrieval augmented generation (RAG) so your model has access to selected internal data organized in a database.

This [Nutanix Validated Design](#) guide can help you get started with fine-tuning or RAG.

Challenges

The challenges associated with code and content creation use cases include:

- **Complexity.** You may need to deploy AI everywhere you operate, including datacenters, public clouds and at the edge. Technology differences between these disparate environments increase operational complexity.
- **Specialized hardware.** You'll need it to accelerate inference – GPUs from NVIDIA or CPUs optimized for inference – and GPUs to support AI training. Because this hardware is expensive and difficult to come by, it's important to keep utilization high.
- **AI lifecycle management.** Because AI apps and the models they are built on constantly evolve, containers and Kubernetes are preferred to run and scale AI models. This provides faster deployments and updates, high availability and efficient resource utilization. However, there may be a skills gap if containers and Kubernetes are new to your organization, and it might be difficult to find people with the cloud native skills.

How Nutanix Helps

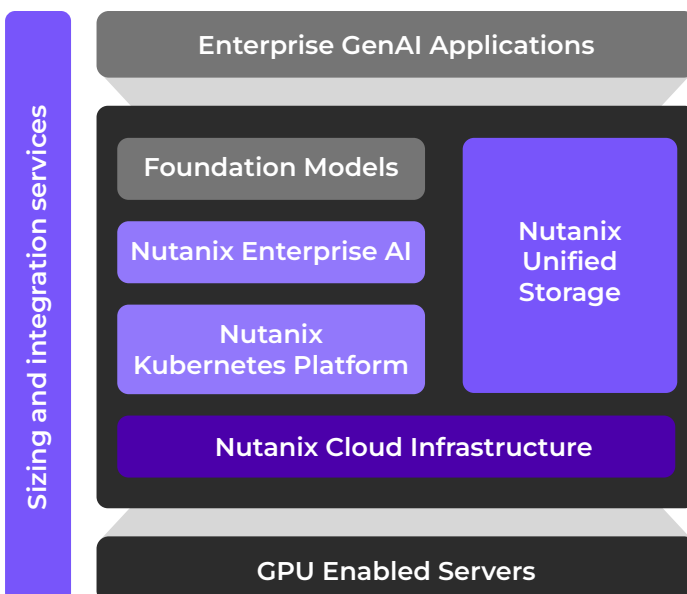
Nutanix provides the AI tools your employees need in less time with less effort using a secure, scalable and cost-effective platform. With Nutanix, you can accelerate your AI initiatives across hybrid multicloud environments while maintaining control over your data.

Nutanix addresses several key challenges surrounding use cases for code and content creation:

- **Take advantage of hybrid multicloud.** Run the latest AI models anywhere without significant operational changes in each location.
- **Efficiently utilize specialized hardware.** Get more from costly GPUs and other specialized hardware through efficient resource sharing.
- **Leverage containers and Kubernetes.** GPT-in-a-Box from Nutanix simplifies Kubernetes and makes it easy to get started with cloud native by enabling you to use your preferred tools. At the same time, NAI lets you effortlessly deploy AI models on the infrastructure and Kubernetes of your choice.

GPT-in-a-Box is a turnkey software-defined solution designed to seamlessly integrate AI models into your operations while keeping data and apps under your control. Its advantages include:

- **Full AI stack.** Deploy a curated set of large language models (LLMs) using leading open-source AI frameworks.
- **Delivered anywhere.** From datacenter to edge to public cloud.
- **Built-in data services.** Fine-tune and run LLMs on optimized storage while maintaining control of data and apps.
- **Inference at the edge.** Deploy everything you need to run AI models for inference at edge locations. A compact footprint overcomes site limitations and helps control costs.
- **Lower TCO.** Reduce TCO with automation, dynamic resource allocation and consolidation to optimize infrastructure costs.
- **Deploy and operate secure APIs for AI apps.** Easily create secure role-based APIs to connect apps to AI models.
- **Deploy and adapt with ready-to-use AI models and LLMs.** Validated AI models that can be deployed quickly from the edge to cloud and speed-up time-to-value.



Use Case: Government Agency Tackles Document Analysis

When a government agency needed to decrease the time spent finding legislative data in a repository with millions of documents in 24 languages, it turned to Nutanix. GPT in-a-box made the agency's AI solution easy to deploy and eliminated the need to build a solution from scratch.

In addition to GPT-in-a-Box with NVIDIA GPUs, key solution elements included:

- DocsGPT, an AI-powered chatbot that improves the user experience for documentation use cases.
- Red Hat OpenShift, the full-stack Red Hat Kubernetes solution fully validated for the Nutanix platform.
- A custom LLM from Mistral.

The agency projects \$1.04 million USD in savings in the first year, with a break-even point within 90 days of launch and a **10%** increase in productivity per employee.

Additional Guidance

For more guidance, download the recently published eBook, [The Essential Guide to Navigating AI and Cloud Native Deployments](#).

This indispensable guide explains what you need to build and deploy AI models that speed-up code and content creation, enrich the customer experience, and strengthen security as well as how to accelerate the deployment of cloud native apps.

NUTANIX

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

©2025 Nutanix, Inc. All rights reserved. Nutanix, the Nutanix logo and all Nutanix 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). AI-HowtoAccelerateCodeandContentCreationwithGenerativeAIModels-Infobrief-FY25Q3 03202025