

# NAI GPT Deployment

Product Code: CNS-NAI-A-GPT-SVC-DEP-ONP

## At-a-Glance

### Stage: Deploy

The Nutanix Enterprise Artificial Intelligence (NAI) Deployment service accelerates the deployment of a comprehensive inference endpoint management product designed to streamline and optimize your AI model orchestration experience. NAI allows you to select, deploy, and manage large language models (LLMs) on a Kubernetes cluster. This offer is ideal for the Deploy stage of the Nutanix Enterprise AI solution journey.

### Related Services

- AI/ML Planning
- AI/ML Design
- AI/ML Strategy & Optimization Series
- Infrastructure Deployment
- NKP Deployment
- NUS Deployment

## Service Scope

Highly skilled consultants with strong domain expertise and rich experience deploy NAI on a supported Kubernetes platform. After the deployment of the solution, the consultant demonstrates the LLM with sample data.

This service includes the following activities:

- Review the NCI cluster configuration that runs a supported Kubernetes Platform, including:
  - GPU Support
  - GPU operator installed
  - NKP version
  - NUS files installed with CSI Setup in NKP
  - NKP load balancer set up with FQDN
  - SSL certification (secure)
- Install NAI on the supported Kubernetes cluster, including:
  - Add and update Nutanix helm repository
  - Set up Hugging face
  - Importing LLM
  - Configuring Endpoint
  - Demonstrating the LLM with sample application

## Project Management

Nutanix Project Management (PM) oversees Nutanix resources and aligns execution with your goals, scope, and timelines.

Core project management activities may include the following:

- Serve as a single point of contact for all project communication
- End-to-end Nutanix resource management
- Coordinate change window(s) and implementation schedules with customer
- Track and facilitate readiness and prerequisite completion
- Conduct project kickoff/technical readiness meeting(s)
- Integrate customer resources into the high-level project timeline
- Send status update(s)
- Manage timeline(s)
- Deliver created artifacts to the customer
- Facilitate project closeout activities

## Limitations

- For each quantity purchased, deployment is limited to 1 on-premises NCI cluster
- Excludes training a new LLM
- Excludes creation or updates to existing Design Document
- Excludes NCI cluster, NKP, and NUS deployment

**Note:** For AI/ML workloads running on Bare Metal or Public Cloud, a custom SOW is required.

## Supported Hypervisors

- Nutanix AHV

## Supported Kubernetes Platforms

- Nutanix Kubernetes Platform (NKP)

## Project Management

- Excludes scheduling customer resources and activities
- Excludes detailed project plan (schedule) development and management
- Excludes responsibility for creating, managing, or delivering change management communications

## Prerequisites

- Fully supported and functional on-premises NCI cluster that meets all product requirements for NCI, NKP, NUS, and a supported GPU

**Note:** For information on the requirements for NCI Clusters, see Field Installation Overview in the *Field Installation Guide* on the Nutanix Support Portal.

For information on the requirements for deploying NKP see Basic Installations by Infrastructure in the *Nutanix Kubernetes Platform Guide* on the Nutanix Support Portal.

For information on the requirements for NAI, see Nutanix Enterprise AI Requirements in the *Nutanix Enterprise AI Guide*.

For information on NUS Files Prerequisites, see Prerequisites in *Nutanix Files User's Guide* on the Nutanix Support Portal.

- Completed Pre-delivery Questionnaire

## Required Product Licenses

- Nutanix Cloud Infrastructure (NCI) Ultimate Edition
- Nutanix Enterprise AI (NAI) GPT Pro or Ultimate Edition Only
- Nutanix Kubernetes Platform (NKP) Pro or Ultimate Edition
- Nutanix Unified Storage (NUS) Pro Edition

## Delivered Artifacts

Delivered Artifact	Description
As-built Guide	Captures the final, deployed configuration of the solution, detailing how the environment was actually built and configured in comparison to the customer provided design.

## Level of Effort

Typically up to 2 days

## Delivery Type

Delivery Type	Delivery Activities
Virtual	<ul style="list-style-type: none"> <li>• Virtual deployment</li> <li>• Virtual project management</li> </ul> <p><b>Note:</b> Any in-person project management activities provided solely at Nutanix's discretion</p>
In-person	<ul style="list-style-type: none"> <li>• In-person deployment</li> <li>• Virtual project management</li> </ul> <p><b>Note:</b> Any in-person project management activities provided solely at Nutanix's discretion</p>

## Related Products

- Nutanix Cloud Infrastructure (NCI)
- Nutanix AI (NAI)
- Nutanix Kubernetes Platform (NKP)
- Nutanix Unified Storage (NUS)

## Terms and Conditions

This document contains the entire scope of the service offer. Anything not explicitly included above is out of scope. This service offer is subject to the Nutanix Services General Terms and Conditions that can be viewed at <https://www.nutanix.com/support-services/consulting-services/terms-and-conditions>

©2026 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. Nutanix, Inc. is not affiliated with VMware by Broadcom or Broadcom. VMware and the various VMware product names recited herein are registered or unregistered trademarks of Broadcom in the United States and/or other countries. All other brand names mentioned herein are for identification purposes only and may be the trademarks of their respective holder(s).