

# Nvidia GPUを搭載したCisco AI UCSサーバのサポートを依頼

## 内容

---

[はじめに](#)

[バックグラウンド情報](#)

[NVIDIA AIエンタープライズソフトウェアの概要](#)

[NVIDIA AI Enterpriseとは何ですか？](#)

[主なコンポーネント](#)

[導入モデル](#)

[NVIDIA AI Enterpriseライセンスでサポートの利点](#)

[オーダー情報](#)

[サポートの利用方法](#)

[シスコサポートへのお問い合わせ](#)

[NVIDIAサポートへのお問い合わせ](#)

[その他のAI関連ソフトウェアのサポート](#)

---

## はじめに

このドキュメントでは、NVIDIA AI Enterprise Software (NVAIE)スイートと、Cisco Unified Computing System (UCS)プラットフォームに導入された場合に利用できる包括的なサポートの利点について説明します。このドキュメントでは、最適な導入と運用支援のためにシスコとNVIDIAの両方のサポートを利用する方法についても説明します。

## バックグラウンド情報

人工知能と機械学習のワークロードが企業の業務にとってミッションクリティカルになるにつれ、企業は強力なインフラストラクチャだけでなく、エンタープライズグレードのソフトウェアとサポートも必要とします。NVIDIA AI Enterprise Softwareは、Cisco UCSインフラストラクチャでの展開に最適化された、包括的で安全な、サポートされているAIプラットフォームを提供します。

この文書は次のことについて記述しています：

- NVIDIA AIエンタープライズソフトウェアの内容
- NVAIEソフトウェアスイートのライセンスのサポート上の利点
- Cisco UCS上でNVAIEのサポートチャネルを適切に連携させる方法

## NVIDIA AIエンタープライズソフトウェアの概要

NVIDIA AI Enterpriseとは何ですか？

NVIDIA AI Enterprise (NVAIE)は、仮想化されたクラウドネイティブ環境で実行されるAIワークロードに対して、エンタープライズグレードのセキュリティ、安定性、管理性、およびサポートを提供する包括的なソフトウェアスイートです。Cisco UCSなどの主要なエンタープライズプラットフォームで動作するように最適化され、認定されています。

## 主なコンポーネント

---



注：最新の包括的なリストについては、[Nvidia AI Enterprise](#)をご覧ください。

---

### AIフレームワークとツール：

- NVIDIA AI Workflows – 一般的なAIアプリケーション向けに事前構築されたエンドツーエンドのワークフロー
- NVIDIA TAO Toolkit – コンピュータビジョン、対話型AIなどの学習ツールキットを移行
- NVIDIA Riva – スピーチAIアプリケーションを構築するためのSDK
- NVIDIA Morpheus – サイバーセキュリティAIフレームワーク
- NVIDIA Metropolis – インテリジェントビデオ分析のためのVision AIプラットフォーム
- NVIDIA Clara – 医療AIアプリケーションフレームワーク

### インフラストラクチャソフトウェア：

- NVIDIA vGPUソフトウェア- GPU仮想化テクノロジー ( 別途ライセンスまたはNVAIEに含まれる )
- NVIDIA GPUオペレータ- GPU管理のためのKubernetesオペレータ
- NVIDIA Network Operator - KubernetesでNVIDIAネットワークリソースを管理
- NVIDIA Container Toolkit - GPUアクセラレーテッドコンテナを構築および実行するためのツール

### AI開発スタック：

- 最適化されたAIフレームワーク ( TensorFlow、PyTorch、JAXなど )
- データサイエンス向けNVIDIA RAPIDS
- 実稼働AI導入のためのNVIDIA Triton Inference Server
- 最適化された推論のためのNVIDIA TensorRT
- NVIDIA cuDNN、cuBLAS、およびCUDAライブラリ

### 管理とセキュリティ：

- NVIDIA Base Command Manager Essentials - AIのインフラストラクチャ管理
- NVIDIAライセンスシステム – 一元化されたライセンス管理
- セキュリティパッチと長期サポート(LTS)バージョン

## 導入モデル

1. 仮想化環境- VMware vSphere、Red Hat Virtualization、NVIDIA仮想GPU

- 2. ベアメタル：物理サーバへの直接導入
- 3. クラウドネイティブ:Kubernetesおよびコンテナ型導入
- 4. ハイブリッドとマルチクラウド：環境全体で一貫性のあるAIプラットフォーム

## NVIDIA AI Enterpriseライセンスでサポートの利点

### 包括的な企業サポート

NVIDIA AI Enterpriseのライセンスを取得することで、基本的なGPUドライバアクセスを超えるサポートの利点を得ることができます。

- 1. 実稼働グレードのソフトウェアサポート
  - ソフトウェアアップデートおよびパッチ：定期的なセキュリティアップデート、バグ修正、および機能拡張
  - 長期サポートリリース：安定したLTSバージョンと延長サポートウィンドウ（最大2年）



注: Nvidiaサポートレベル(Basic vs 24x7)は、[ここ](#)で説明するように購入された特定のライセンスに基づいています

サービスコンポーネント	標準	24時間365日
テクニカルサポートへのアクセス	現地の営業時間	24時間、週7日
重大度1の初期応答時間	四営業時間	1 時間
重大度2の初期応答時間	四営業時間	2 時間
重大度3の初期応答時間	1営業日	四営業時間
重大度4の初期応答時間	2営業日	1営業日
カスタマーポータルへのアクセス	24時間365日	24時間365日
Web	24時間365日	24時間365日
電話	現地の営業時間	24時間365日

### 2. 技術的な専門知識とガイダンス

- AI導入支援 – アーキテクチャ設計とベストプラクティスの支援
- パフォーマンス最適化 – パフォーマンスを最大化するためのAIワークロードの調整に関するガイダンス
- フレームワークサポート：スイートに含まれるTensorFlow、PyTorch、およびその他のAIフレームワークのサポート
- コンテナとKubernetesのサポート：クラウドネイティブのAI導入を支援

- ・ナレッジベースへのアクセス：包括的なドキュメント、ガイド、トラブルシューティングリソース

### 3. ライセンスおよびアクセスの利点

- ・検証済みおよび検証済みのソフトウェア – エンタープライズテスト済みおよび認定済みリリース
- ・ライセンスポータビリティ：インフラストラクチャ間でライセンスを移動できる柔軟性
- ・ライセンス管理の一元化- NVIDIAライセンスシステムによるデプロイ管理
- ・サブスクリプションモデル：年間または複数年のライセンスオプション
- ・ソフトウェアダウンロードアクセス：すべてのNVAIEコンポーネントの正式なダウンロードポータル

### 4. セキュリティとコンプライアンス

- ・セキュリティパッチ – タイムリーな脆弱性の修復
- ・CVEトラッキング：セキュリティ問題の事前通知
- ・コンプライアンス文書：規制要件のサポート
- ・セキュアソフトウェアサプライチェーン：検証済みの署名付きソフトウェアコンポーネント

#### [サポート契約 ( Support Coverage ) ]

機能	NVIDIA AIエンタープライズ
GPUドライバアクセス	✓
基本的なドキュメント	✓
コミュニティフォーラム	✓
AIソフトウェアスイート	✓
エンタープライズサポート ( 24時間 365日 )	✓
SLA応答時間	✓
セキュリティパッチ	✓
長期サポート	✓

機能	NVIDIA AIエンタープライズ
実稼動環境での認証	✓
テクニカルアカウント管理	オプションのアドオン
AIワークフローアクセス	✓
管理ツール	✓

## オーダー情報

Cisco UCSとNVIDIA AI Enterprise

ステップ1: Cisco UCSハードウェアの選択

適切なCisco UCSプラットフォームとNVIDIA GPU構成の選択：

設定例:

- NVIDIA HGX H200 GPU搭載UCS C885A M8
- NVIDIA RTX P6000 GPU搭載UCS C845A M8
- NVIDIA HGX B300 GPU搭載UCS C880A M8
- NVIDIA L40S GPU搭載UCS C240 M7
- モジュラ導入に適したNVIDIA H100 GPU搭載UCS Xシリーズ



注: NVIDIA H200 NVL、H100 PCIe、およびH100 NVL GPUには、それぞれNVIDIA AI Enterpriseを含む5年間のNVIDIA Enterpriseサブスクリプションが付属しています。ソフトウェアアクティベーションが必要です。ライセンスを有効にし、必要に応じてNVIDIAでソフトウェアサポートケースを開く必要があります。

詳細については、[ここ](#)をクリックしてください。

たとえば、NVIDIA H200 GPU搭載のUCS C845A M8、UCSX-210C-M8、UCS C240 M7には、5年間のNVIDIA Enterpriseサブスクリプションが含まれています。

ステップ2: NVIDIA AI Enterpriseライセンス

NVIDIA AI Enterpriseライセンスは別売りで、以下の方法で購入できます。

- Cisco Commerce - Cisco UCS購入にバンドル



注:NVIDIAライセンスは、シスコ以外で購入することもできます。

## ライセンスのタイプ:

- NVAIE-Standard – 標準の年間契約 ( GPUソケットまたはユーザーごと )
- NVAIE-Educational – 教育機関向け年間サブスクリプション ( GPUソケット単位またはユーザー単位 )
- NVAIE-vGPU - NVAIE + vGPUライセンスの組み合わせ

詳細については、[NVIDIA注文ガイド](#)を参照してください



注 : 特定の製品番号と価格については、シスコアカウントチームまたはNVIDIAの営業担当者にお問い合わせください。

## ステップ3 : サポートオプション

次の点を確認してください。

- Cisco UCSサポート契約 – シスコのサポート ( 標準、拡張、シグニチャ )
- NVIDIA AIエンタープライズサブスクリプション- NVIDIAサポートを含む

## リファレンス構成

### AIインフラの発注例 :

1. ハードウェア : UCS C885A M8 ( NVIDIA H200 GPU搭載 ) (H200 X 8)
1. ソフトウェア : NVIDIA AI Enterprise ( GPUライセンスx 8 )
2. サポート:
  - Ciscoサポート – Enhanced for UCSハードウェア
  - NVIDIA AI Enterpriseサブスクリプション ( 24時間365日サポートを含む )
3. サービス : 導入のためのオプションのプロフェッショナルサービス

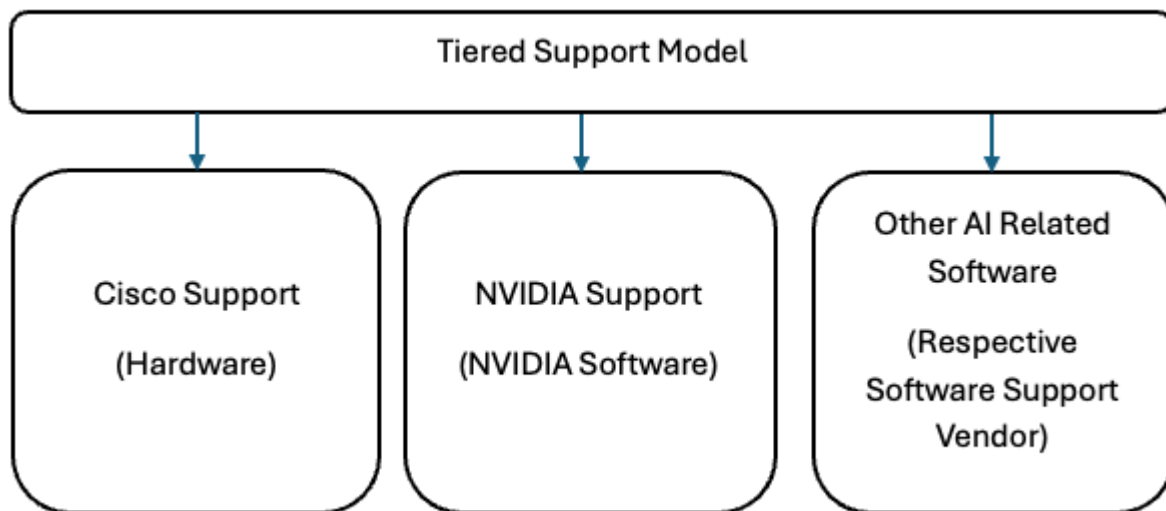
発注ガイドの詳細については、次を参照してください。

- [Cisco UCS用NVIDIA AI Enterprise発注ガイド](#)

## サポートの利用方法

### サポートモデルの概要

Cisco UCS上のNVIDIA AI Enterpriseは、ハードウェアとソフトウェアのサポートを明確に区別する階層型サポートモデルに準拠しています。



注:NVIDIAソフトウェアの問題については、NVIDIAで直接サポートケースを作成する必要があります。必要に応じて、共同のトラブルシューティング作業に関するシスコサポートも含めてください。

## シスコサポートへの問い合わせ方法

- ハードウェアの問題:
  - Cisco UCSサーバの障害またはエラー
  - ハードウェア診断と交換
  - UCSコンポーネントのファームウェアアップデート
  - UCS ManagerまたはIntersightの設定の問題
  - 物理的なGPUのインストールまたは検出の問題
  - UCSインフラストラクチャのネットワーク接続の問題
- シスコソフトウェア:
  - UCS ManagerまたはCisco Intersightの問題
  - AIインフラ向けシスコネットワーク構成

シスコサポートケースのオープン方法：

1. 次のサイトを参照してください。 [各国のシスコサポートの連絡先](#)
2. シスコのサポートに直接お問い合わせください（24時間365日対応のホットライン）。
3. Cisco Support Case Managerの使用
4. Cisco Intersightを介した連絡（クラウド接続デバイス用）

サポート・リクエストを作成する前に必要な情報：

- Cisco UCSのモデルとシリアル番号
- サービス契約番号

- 問題の詳細な説明
- テクニカルサポートファイル(show tech-support)
- GPUインベントリと構成の詳細

## NVIDIAサポートへのお問い合わせ

### NVIDIA AIエンタープライズソフトウェア：

- AIフレームワークのインストールまたは設定の問題
- NVIDIA AIワークフローに関する質問
- Triton Inference Serverの導入
- NVIDIA vGPUソフトウェアの問題
- GPUオペレータまたはネットワークオペレータの問題
- NVIDIA Container Toolkitでのコンテナランタイムの問題
- ライセンスの有効化または管理
- AIワークロードのパフォーマンス調整
- ソフトウェアのバグや機能の要求
- セキュリティ脆弱性に関する質問

### ドライバおよびソフトウェアの問題：

- NVIDIAドライバのインストールまたは互換性
- CUDAツールキットの問題
- cuDNNまたはTensorRTの問題
- GPU使用率またはパフォーマンスの問題

### サービスリクエストをオープンする方法：

1. NVIDIA Enterprise [Support Portal](#)にアクセス
2. NVIDIA Enterpriseアカウントでログインします
3. 「Create Case」を選択し、次の情報を入力します。
  - NVIDIA AI Enterpriseライセンス情報
  - サーバーとGPUの詳細
  - ソフトウェアのバージョン
  - 問題の詳細な説明
  - ログと診断出力

### サポートポータルアクセス：

- アクティブなNVIDIA AIエンタープライズサブスクリプションが必要
- ライセンスプロビジョニング時のアカウント設定
- ナレッジベース、ダウンロード、ケース管理へのアクセス

### その他のAI関連ソフトウェアのサポート



該当するAIソフトウェア/アプリベンダーが提供するサポート手順に従う

このドキュメントの情報は、特定のラボ環境にあるデバイスに基づいて作成されたものです。このドキュメントで使用するすべてのデバイスは、クリアな（デフォルト）設定で作業を開始しています。本稼働中のネットワークでは、各コマンドによって起こる可能性がある影響を十分確認してください。

## 翻訳について

シスコは世界中のユーザにそれぞれの言語でサポート コンテンツを提供するために、機械と人による翻訳を組み合わせて、本ドキュメントを翻訳しています。ただし、最高度の機械翻訳であっても、専門家による翻訳のような正確性は確保されません。シスコは、これら翻訳の正確性について法的責任を負いません。原典である英語版（リンクからアクセス可能）もあわせて参照することを推奨します。