aws marketplace

NVIDIA AI Enterprise

Reviews, tips, and advice from real users






Powered by PeerSpot

Contents

Product Recap

NVIDIA AI Enterprise

NVIDIA AI Enterprise Recap

NVIDIA AI Enterprise provides a comprehensive suite of AI tools designed for deployment across diverse industries, enabling businesses to harness the power of AI for scalable, efficient operations.

NVIDIA AI Enterprise offers a robust set of AI technologies tailored for advanced data analytics, machine learning, and neural networks. It streamlines AI deployment, optimizing workload management and facilitating rapid model training and deployment. With support for a range of frameworks and environments, it integrates seamlessly into existing IT infrastructures, reducing complexity and enhancing scalability.

What are the most important features of NVIDIA AI Enterprise?

- Pre-trained Models: Enables accelerated AI implementation with a library of models designed for common use cases.
- Containerized Environment: Simplifies deployment and management through isolated environments.
- Security Enhancements: Incorporates advanced security protocols ensuring data protection.
- Scalability: Allows for seamless scaling of resources to meet changing demand.

What benefits or ROI should users look for in reviews?

- Improved Efficiency: Users highlight reduced training times and increased workflow productivity.
- Cost Savings: Gain insights into significant reductions in overhead through streamlined processes.
- Enhanced Innovation: Leverage insights into faster deployment cycles leading to quicker time-to-market.
- Integration: Reviews often note easy alignment with existing IT structures.

Industries like healthcare, retail, and finance leverage NVIDIA AI Enterprise to implement predictive analytics, optimize supply chains, and enhance customer interactions. In healthcare, it's used for advanced diagnostics, while retailers utilize its capabilities for personalized marketing. Finance sectors harness its power for risk assessment and fraud detection.

Valuable Features

Excerpts from real customer reviews on PeerSpot:

- "Overall, it increased productivity for both development teams and end-users by making AI solutions faster, scalable, and easier to maintain."

Verified user

Operations Analyst at a non-profit with 51-200 employees

- "Additionally, there has also been productivity gains in optimizing workloads with retrieval augmented generation and running demos on the AI workstation, the laptop, leading to a 200 percent increase in productivity."

Josh Thias

AI Software Validation Engineer at a manufacturing company with 10,001+ employees

- "My experience was very positive, as it helped me start from scratch and the GPUs and also the NGC website that NVIDIA provides with all the resources are phenomenal, making it a highly enriching experience with NVIDIA AI Enterprise."

Samuel Beera

Principal Enterprise Architect at Mercy

"The impact of NVIDIA AI Enterprise on the time for my AI applications is pretty remarkable because you don't have as much downtime with this type of software."

Verified user

Solutions Architect at a manufacturing company with 10,001+ employees

"The return on investment has been substantial, and NVIDIA AI Enterprise gives my team the ability to work on multiple projects at the same time, boosting productivity by about two hours a day because we are not going to spend hours just researching products since NVIDIA AI Enterprise gives all of the resources and features available at our fingertips."

Verified user

Ai & Advanced Computing Engineer at a tech vendor with 10,001+ employees

"Now I am able to easily handle more than 60 high-resolution camera streams simultaneously on a single H100 GPU with excellent throughput and very low latency, and the development time for new vision pipelines has dramatically dropped from three to four weeks to only four to six days because of DeepStream."

Nandhavignesh Ramalingam

NVIDIA AI Architect at Lenovo

"NVIDIA AI Enterprise offers several advantages for customers. It provides faster AI performance and quicker time-to-market. Higher GPU utilization helps to accelerate work for organizations."

Anil Rahulwar

L3 Storage Engineer at Capgemini

What users had to say about valuable features:

"NVIDIA AI Enterprise offers several advantages for customers. It provides faster AI performance and quicker time-to-market. Higher GPU utilization helps to accelerate work for organizations. Organizations can develop AI solutions and deploy them across on-premises data centers or any cloud environments.

"NVIDIA AI Enterprise platform reduces the complexity of deploying and managing AI applications. Integration with AI has a significant impact on project development. We can improve quality and productivity through AI integration in development, which accelerates software quality, reduces costs, and increases team productivity. This enables faster business innovations and provides a significant competitive advantage through quick time-to-market.."

Anil Rahulwar

L3 Storage Engineer at Capgemini

Read full review

"I find that NVIDIA AI Enterprise is very useful to improve the speed of video effects and makes me work faster. I love it.

"The HP8 and HP16 are the best features for core acceleration. The HP8 and HP16 features are deep learning, so they can learn very fast and then be more efficient after the core acceleration. With a weak machine, we can achieve some really nice improvement.

"I have noticed a lot of time saved. The results can be not the best, but after the human task and after the AI, it is acceptable. Mainly, we can gain time.

"I gain approximately thirty percent of my time, which equals a lot of hours each month.."

Verified user

Technicien Exploitation Réseau at a outsourcing company with 201-500 employees

Read full review

The best features NVIDIA AI Enterprise offers are ease of use, deployment support, and scaling and performance optimization.

The deployment support from NVIDIA AI Enterprise helps my projects significantly. Timely support helps every team and gives us the opportunity to explore and implement solutions.

“NVIDIA AI Enterprise optimizes performance through TensorRT models, which improve the speed and throughput of the models.

“NVIDIA AI Enterprise has positively impacted my organization by improving productivity, response time, and overall GPU performance. It optimizes models and enhances their capabilities.

“The specific outcomes and metrics I have seen include faster deployment times, reduced costs, and improved model accuracy. .”

Pradipvetal Pradipvetal
Associate Architect at a tech vendor with 10,001+ employees

Read full review

"NVIDIA AI Enterprise offers a great breadth of resources. To start with, NVIDIA offers its fine-tuned models, and one of the best features are the blueprints. For enterprises who have access to NVIDIA AI Enterprise, they can kickstart their AI journey by using the models as well as the blueprints.

"As far as NVIDIA models are concerned, we did not use the NVIDIA models, but we used open source models that we downloaded and installed on the AI server. However, some of the resources that we widely used are the blueprints. For example, to start off with enterprise RAG applications, NVIDIA had everything in place that helped cut down our time to build the applications.

"The resources that are offered include the models, secondly, the blueprints, and then the training resources that NVIDIA offers, which are invaluable. For an enterprise that is trying to have AI as part of their strategy, NVIDIA AI Enterprise license provides almost everything that enterprises need to get the team trained, use the models, and get the training that is required. It is almost a one-stop shop.."

Samuel Beera
Principal Enterprise Architect at Mercy

Read full review

"NVIDIA AI Enterprise has impacted my organization positively for a number of reasons. There has been a lot of optimization when it comes to researching organizational information because we have consolidated sites such as SharePoint, and NVIDIA AI Enterprise helps us access resources much quicker without needing to search the web for article after article. That has been very helpful. Additionally, there has also been productivity gains in optimizing workloads with retrieval augmented generation and running demos on the AI workstation, the laptop, leading to a 200 percent increase in productivity.

"The accuracy of NVIDIA AI Enterprise has been exceptional, particularly when using generative AI such as retrieval augmented generation. The platform is built on reinforcement learning and model training with extensive libraries, making accuracy and reliability standout features. I believe this to be one of the best advantages of NVIDIA AI Enterprise, and the training continues to reduce errors. While models are never perfect, as humans and data curation are not perfect, I do believe that increased customer support, such as a real-time support desk, would help provide customers with the right information to support this type of platform.."

Josh Thias

AI Software Validation Engineer at a manufacturing company with 10,001+ employees

Read full review

“The best features of NVIDIA AI Enterprise are the GPU orchestration and the visualization that can happen. When you run the Enterprise software, you will have access to other features including AI Workbench.

“Many of these features you can actually view for free on build.nvidia.com, and I want to stress that because people think this is so complicated and so expensive that they will never have access to it. That is not true. NVIDIA has a number of resources and training modules that give you access to this information without necessarily needing to purchase everything because usually the companies that have AI Enterprise purchase it on a per GPU basis.

“So it is one license per GPU, and that number does add up quickly. However, it is easy because it again gives you visualization of everything. Think of it as a dashboard you can run. You don't need to know how to program or how to read code. It is very visual and it gives you metrics on how effective everything is running and you can toggle different environments. If you want to dive deeper, you can run all kinds of simulations with that.

“You kind of start high-level visual and then you can get more specific over time depending on what your job is.

“The impact of NVIDIA AI Enterprise on the time for my AI applications is pretty remarkable because you don't have as much downtime with this type of software. The downtime that has been saved is pretty much as much as possible. There is usually about six nines of availability. Six nines means it is 99.9999% up. That comes out to basically experiencing about 11 seconds out of the year, which is just basically them toggling new servers.

“What is interesting about this is that many of the GPUs are what are called hot-swappable, meaning that you can actually change them out while your data center rack is still powered on and running. They have it built in so you can pull it with a special tab, which is an orange tab. It is not going to shock you or anything. This gives you the ability to go into the facility or have the facilities team change that out if there is an issue and you need to change a GPU out or you need to change a license out, and they can make it happen without any interruption. There is really not any interruption noticed. It is still a fairly new product so there are not as many

examples, but from what I have seen, there is not really any issue with interruption. NVIDIA AI Enterprise definitely keeps the GPUs maintained, and that is why orchestration is so important. It is an electric car in nature—it charges when it is not being worked. Because if you just drive a car all day and don't maintain it, the car is going to fall apart. It is the ultimate maintenance package.."

Verified user

Solutions Architect at a manufacturing company with 10,001+ employees

Read full review

Other Solutions Considered

"There is no such competition for NVIDIA AI Enterprise, as they are addressing the complete AI-related space. Even if AMD has GPUs, Dell has that, and all of this, NVIDIA AI Enterprise is leading because they are addressing each and every component in the AI infrastructure.."

Bharath _Kumar
Solution Architect at a outsourcing company with 1,001-5,000 employees

Read full review

"Before adopting NVIDIA AI Enterprise, we were primarily using a combination of open-source tools and custom-built ML infrastructure. This included a standard Python-based ML stack, Docker-based deployment, and manual management of GPU environments on cloud providers like AWS.."

Verified user
Operations Analyst at a non-profit with 51-200 employees

Read full review

"We did not use a different solution. However, we were also using Microsoft Azure. The NVIDIA H100s was an internal deployment. So, we were using Azure AI Foundry, too.."

Samuel Beera
Principal Enterprise Architect at Mercy

Read full review

"I was previously using a combination of Red Hat OS and other orchestration platforms on Linux Ubuntu, which the federal government primarily utilizes. While Red Hat is crucial and works across many servers, it is not always the latest or most advanced, and its licensing costs have become expensive. The same situation applies to VMware private cloud foundations, where costs also escalated.."

Josh Thias
AI Software Validation Engineer at a manufacturing company with 10,001+ employees

Read full review

"Before choosing NVIDIA AI Enterprise, we evaluated a few other options to compare performance, cost, and ease of deployment, including AWS SageMaker, Google Vertex AI, and standard open-source MLOps stacks. NVIDIA AI Enterprise was preferred for better GPU performance optimization, lower inference latency, and tighter integration with on-premises hybrid GPU infrastructure.."

Verified user

Operations Analyst at a non-profit with 51-200 employees

Read full review

“When comparing NVIDIA with other solutions or other vendors like AWS, Google, Cerebras, I find that they are pretty much the leader in everything possible because they have the best hardware. They are not really a software company actually, because they prefer to work with their channel resellers and channel partners.

“NVIDIA is very profitable because they don't really have a huge sales team. They basically work with everybody and anybody other than AMD is obviously a competitor, but they are still partnering with every possible supplier out there to push the envelope as best they can with innovation. I would say they are vastly outperforming everybody.

“If you just look at the stock market, it has been that way for so long and now that they are as big as they are, people are excited to see where it goes, but they are wondering how much more it can grow. I would say they just need to help teach people as much as they can about how this information and technology works. They have a pretty good training and certification program, but many people don't know that it exists unless you already work there or work with someone who works there.

“People who are trying to get in the door have to think of it as a numbers game. I hope that NVIDIA would make these resources more publicly available and say, "You want to learn how to use AI Enterprise? Here are the resources." They have these classes, but unless you know somebody, you are not going to really know where to study. They have exams that are about $100, but sometimes people's companies can reimburse them for that type of thing. That would be something I would encourage NVIDIA to continue to invest in. Their solution is going to perform better than anybody out there by far. However, they are going to also be pretty expensive, so you have to compare and contrast that price with the performance..”

Verified user

Solutions Architect at a manufacturing company with 10,001+ employees

Read full review

ROI

Real user quotes about their ROI:

I have seen a return on investment through NVIDIA AI Enterprise. A significant amount of money is saved by using local models, which avoids output token costs.

Pradipvetal Pradipvetal
Associate Architect at a tech vendor with 10,001+ employees

Read full review

"Regarding return on investment, NVIDIA AI Enterprise provides financial benefits. It enables faster time-to-market, increased productivity, and reduced infrastructure costs.."

Anil Rahulwar
L3 Storage Engineer at Capgemini

Read full review

"The return on investment has been substantial. I achieved 200% time to value. It has not reduced any employees. It has just allowed more people to be more productive in their work.."

Verified user
Ai & Advanced Computing Engineer at a tech vendor with 10,001+ employees

Read full review

"The return on investment has shown significant money saved and time needed. There has not been a reduction in employees, and nobody wants their job to be replaced by AI in any capacity. However, with GPUs, especially through RunAI, the GPU orchestration platform facilitates increased effectiveness and efficiency. NVIDIA has invested in GPU orchestration by acquiring Slurm, a popular job scheduling tool for high-performance computing, providing roughly a 250 percent return on investment. Millions of dollars are being reinvested into hardware, and savings from GPU orchestration are now allocated for power and cooling operations, such as liquid-cooled and air-cooled data center GPUs.."

Josh Thias

AI Software Validation Engineer at a manufacturing company with 10,001+ employees

Read full review

"For organizations that want to have sovereign AI and instead of depending totally on cloud providers, the tokenomics is a huge factor. That is a huge savings that organizations can achieve by utilizing open source models, open-weight models, utilizing open source AI tooling, and the investments that go into this whole stack by doing it yourself. We are investing in NVIDIA resources that in the long run are going to pay off. So, the ROI on going for an NVIDIA AI Enterprise license is huge.

"Definitely fewer employees are needed. As far as return on investment is concerned, there are studies which show that because we got the H100s two years ago, the value of the H100s actually rose because of the increase in pricing. In a way, instead of depreciating the value of the H100 GPUs, it actually went up.."

Samuel Beera

Principal Enterprise Architect at Mercy

Read full review

"NVIDIA AI Enterprise has positively impacted our organization by improving the speed and efficiency of deploying AI solutions. It helped reduce the setup time of GPU environments, streamline model deployment, and improve performance for inference workloads. It also enabled us to build more reliable production-grade AI applications such as an internal knowledge assistant and a document automation system. Overall, it increased productivity for both development teams and end-users by making AI solutions faster, scalable, and easier to maintain.

We saw around a 30 to 40% inference performance improvement, reduced deployment time using pre-built NVIDIA AI Enterprise tools, and better GPU resource utilization for large-scale GenAI workloads.

We saw around a 25 to 30% reduction in infrastructure cost due to better GPU utilization and approximately 40% reduction in model deployment time, which improved overall delivery speed and reduced the engineering efforts needed for production release.."

Verified user

Operations Analyst at a non-profit with 51-200 employees

Read full review

Use Case

"My main use case for NVIDIA AI Enterprise is to generate videos and to improve graphics. I use NVIDIA AI Enterprise to generate videos for social media.."

Verified user

Technicien Exploitation Réseau at a outsourcing company with 201-500 employees

Read full review

"NVIDIA AI Enterprise is totally dependent on AI applications. The use cases include healthcare with medical analysis, video analytics, and generative AI with large language models. We are using Copilot-style applications for generative AI and LLMs to achieve faster customer service, improved productivity, and automation of repetitive tasks.."

Anil Rahulwar

L3 Storage Engineer at Capgemini

Read full review

"We purchased NVIDIA H100 GPUs, and as a result of that, we also got an NVIDIA AI Enterprise license.

"It was mainly used for NVIDIA NGC and to deploy the CLI on the server with the GPUs.

"As I mentioned, we have the NVIDIA NGC website and you need to register. I was the main person who did all the registrations with NVIDIA AI Enterprise. I created the organization, added users, and gave access to multiple resources that NVIDIA AI Enterprise offers through its website, NGC.."

Samuel Beera
Principal Enterprise Architect at Mercy

Read full review

My main use case for NVIDIA AI Enterprise involves deploying RAG models, LLMs, and NVIDIA Ingest. I also use audio models with NVIDIA Riva and Omniverse for digital twin applications. These use cases support retail floor assistants, research assistants, and multiple agents.

A specific example of how I am using NVIDIA AI Enterprise is a RAG-based architecture where I use NVIDIA embed models and NeMoTron embed models from NVIDIA AI Enterprise. I deploy LLMs locally, including Gemma 26 or Llama models. I use agents through agent flow from NVIDIA AI Enterprise, and I project digital humans using NVIDIA AI Enterprise software.

“I have noticed that most of my clients have unique use cases in medical fields. Sometimes for training models, I leverage NVIDIA AI Enterprise. .”

Pradipvetal Pradipvetal
Associate Architect at a tech vendor with 10,001+ employees

Read full review

“NVIDIA AI Enterprise is essentially a GPU enhancement software that takes advantage of NVIDIA's full stack built on what is called NeMo architecture and NIMs, which are the micro inferencing servers. It allows you to put a GPU into a server, oftentimes with eight or up to eight NVIDIA GPUs in a server.

“You will get additional performance that enhances whatever workloads you are running on the server, but you don't have all the right tools such as monitoring, management, and orchestration. NVIDIA AI Enterprise gives you the full software stack that gives you access to really maximize the value. The primary benefit is that you are really taking advantage of the hardware and the software working together.

“Orchestration allows you to schedule jobs to run at certain times. NVIDIA AI Enterprise software also has regular updates, so every couple of weeks there are new pushes out there so you can become more proficient and get a much better hands-on experience for achieving the goals and making the most effective GPU investment possible..”

Verified user

Solutions Architect at a manufacturing company with 10,001+ employees

Read full review

"I have been using NVIDIA AI Enterprise for two to three years and was first introduced to this product a couple of years ago through an NVIDIA sales representative that I was working with at Dell Technologies, supporting numerous large-scale AI and high-performance computing products with NVIDIA AI Enterprise.

"NVIDIA VGPU was one of the compute layers that has been one of the most common main use cases for NVIDIA AI Enterprise. It enables multiple GPUs to share different virtual machines and optimizes resource utilization while condensing hardware operating costs. Because NVIDIA AI Enterprise is typically sold on a per-GPU license, it is important that customers get the best bang for their buck, and NVIDIA VGPU for compute nodes has really been helpful. I have also used this in a number of large-scale RFPs and RFIs.

"I primarily work with AI workloads that are on a hybrid cloud model because the public cloud lacks a secure posture that is required for organizations such as the Department of Defense and military organizations. The private cloud, while it is very secure, is also quite expensive. The hybrid approach is very helpful with primarily on-prem infrastructure for rack integration but also some remote connectivity options. Everything also connects via DHCP, which is a dynamic host control protocol that allows customers to use things such as PuTTY and other VS Code type platforms to essentially SSH or remote into a desktop server.

"I have also been using a couple of other software development kit libraries including NVIDIA NeMo, which is one of our data curation tools that helps clean the data and allows for model training and fine-tuning, and NVIDIA AI Blueprints are very important, allowing for retrieval augmented generation or RAG. Model training and data curation are very important as well.

"There is a large range of libraries offered by NVIDIA AI Enterprise. These catalogs give you all of the information necessary to securely run AI workloads. That has been a very important use case, such as NeMo for the data curation engine for retrieval augmented automated generation, and there are a couple of other use cases such as TensorRT, which is a built-in library for Jupyter notebooks, providing resources for developing the code and the programming. There are also other options available such as NVIDIA for Digital Twins that gets you interested in

building a virtual layer to a physical data center, with various APIs available such as NVIDIA Base Command Manager, and many libraries available. The vast majority of these libraries are open source and can be found on tools such as GitHub and GitLab.."

Josh Thias

AI Software Validation Engineer at a manufacturing company with 10,001+ employees

Read full review

Setup

The setup process involves configuring and preparing the product or service for use, which may include tasks such as installation, account creation, initial configuration, and troubleshooting any issues that may arise. Below you can find real user quotes about the setup process.

"Installation is easy if you have an experienced person who has installation expertise. For new users, there is a learning curve to understand how to open and manage multi-GPU servers, AI environments, Kubernetes containers, and related components.."

Anil Rahulwar

L3 Storage Engineer at Capgemini

Read full review

"As for the installation part, to be honest, I have not installed NVIDIA AI Enterprise right now. We had done only an eight-GPU deployment in our CoE. Eight built servers with eight GPUs were deployed for our lab setup. For the customers, I think there is another team who generally takes care of that.."

Bharath _Kumar

Solution Architect at a outsourcing company with 1,001-5,000 employees

Read full review

"I actually took a new role with my current employer, Mercy. However, the experience with NVIDIA is at Point32 Health, where I was the enterprise architect. We had signed a contract with NVIDIA. It is a three-year license to purchase the H100 GPUs. We did it internally. I was talking about sovereign AI, so we designed the server and we installed the GPUs with the help of HPE.."

Samuel Beera

Principal Enterprise Architect at Mercy

Read full review

"Before choosing NVIDIA AI Enterprise, we evaluated a few other options to compare performance, cost, and ease of deployment, including AWS SageMaker, Google Vertex AI, and standard open-source MLOps stacks. NVIDIA AI Enterprise was preferred for better GPU performance optimization, lower inference latency, and tighter integration with on-premises hybrid GPU infrastructure.."

Verified user

Operations Analyst at a non-profit with 51-200 employees

Read full review

"In terms of onboarding, it depends on the deployment model. For basic experimentation, it is manageable because NVIDIA AI Enterprise provides containers, NVIDIA NIM microservices, and deployment resources. For production enterprise deployment, a setup requires careful planning around GPU infrastructure because it can cost a lot if I do not carefully select the configuration.

"The platform can be complex. Cost planning requires discipline, and successful deployments need stronger platform engineering and governance. It is very powerful, but it works best when the organization has the maturity to operate enterprise AI infrastructure properly. Special training is required to use this.."

Singh Aman

Generative AI Engineer at Tata Consultancy

Read full review

“The deployment of NVIDIA AI Enterprise is very easy because they handle all of it for you basically. You are just getting the software to install on the GPUs.

“There is a pretty useful manual you get, and you get support. Pretty much everybody who has this doesn't do it alone. They have NVIDIA services or professional services that they would purchase as well, and it is all bundled.

“There can be NVIDIA team members that can handle this on-site deployment installation or you can buy that remotely or you can get training credits. There is always another resource available to help out. You just figure it out, but it is pretty easy. My approach is to try to learn as much as I can beyond just as much as I am allowed to learn because there is never an end to that..”

Verified user

Read full review

Solutions Architect at a manufacturing company with 10,001+ employees

Customer Service and Support

"As I said, I did not have much interaction with the customer support. It was mostly through email. I did not have any issues, so I did not find the need for customer support much.."

Samuel Beera
Principal Enterprise Architect at Mercy

Read full review

"If the government and other users use NVIDIA AI Enterprise, they can be sure that everything is very safe, and they do not have to worry about their data or their application being vulnerable to a hack.."

Steven Yueh
Sr solution Engineer at ADVANTECH CO.LTD

Read full review

"Customer support for NVIDIA AI Enterprise has been generally good, especially for enterprise-level issues. We have had 24/7 enterprise support for fast response times through NVIDIA Enterprise support portals and access to dedicated technical accounts and managers for critical issues. Most production issues are resolved quickly with clear guidance and regular updates.."

Verified user
Operations Analyst at a non-profit with 51-200 employees

Read full review

"I think I did not deal with the TAC and all of this, but the way the solution team provides design-level queries and answers questions about sizing is valuable. If you have any challenges in terms of sizing and you reach out to them, that kind of proactive support is always there. That means I can say that it is good. Based on my observations and experience with support, I can give it eight points from zero to ten, where ten is the best.."

Bharath _Kumar
Solution Architect at a outsourcing company with 1,001-5,000 employees

Read full review

"We receive substantial benefits from NVIDIA AI Enterprise support. It provides access to NVIDIA AI experts and faster resolution times. NVIDIA AI Enterprise includes monthly patches for vulnerabilities and bug fixes, which helps reduce security risk and downtime.

"I am satisfied with NVIDIA AI Enterprise customer service because support is available 24/7. Whenever we raise a ticket, they respond immediately without prioritizing by severity or priority level.."

Anil Rahulwar
L3 Storage Engineer at Capgemini

Read full review

"Customer support varies based on the support level purchased, whether it is ProSupport Plus with a mission-critical four-hour response. While this level guarantees quick access, sometimes there are delays as support can bounce between Dell, NVIDIA, and other involved partners and vendors. I believe there is room for improvement regarding transparency and communication in customer support.

"I would rate customer support a seven, as there are metrics assessing effectiveness, time to value, and return on investment for customers. However, there have been delays in communication and responsibilities between companies such as Dell and NVIDIA, creating confusion regarding who owns specific responsibilities. I would like better communication between both parties, which would require investing in highly skilled AI services departments and customer support, including the online chat I previously mentioned.."

Josh Thias
AI Software Validation Engineer at a manufacturing company with 10,001+ employees

Read full review

Other Advice

Regarding NVIDIA AI Enterprise's AI capabilities, I believe its governance and security are strong.

The accuracy and reliability of output from NVIDIA AI Enterprise are excellent.

“The scalability of NVIDIA AI Enterprise is impressive.

“I would rate this review 8 out of 10. .”

Pradipvetal Pradipvetal
Associate Architect at a tech vendor with 10,001+ employees

Read full review

“Resource optimization helps minimize downtime by ensuring proper allocations, monitoring resource utilization, and balancing workloads. This reduces stress on servers and optimizes resource usage to maintain application performance and service levels during traffic spikes.

“I recommend measuring a 30% faster performance improvement through these optimizations. One feature I would suggest is auto-scaling AI infrastructure, which I have documented in a white paper. This feature allows you to add or remove GPUs based on workload demands, improving resource utilizations and lowering costs. My overall review rating for NVIDIA AI Enterprise is 9.5..”

Anil Rahulwar
L3 Storage Engineer at Capgemini

Read full review

"My advice for others considering NVIDIA AI Enterprise is to conduct thorough research and discuss with their facility team. Understanding the rack layout, data center size, floor height, and humidity or CFM in the room is essential. You must determine whether you have the plumbing for AI data center needs, the capacity to support the weight of heavy racks (typically two to 3,000 pounds), and essential infrastructure components such as shock pallets, doors, heat exchangers, and chillers. Once these components are solidified, you can have conversations regarding the appropriate type of NVIDIA AI Enterprise support based on your GPUs.

"NVIDIA AI Enterprise platform continues to evolve over time, and the more often customers are able to go online and teach themselves about these platforms the better. NVIDIA Omniverse Enterprise is a collaborative environment for 3D workflows. When you are making a digital twin, you are basically creating a 3D layer that virtualizes a hardware infrastructure platform, bringing the ideas to life.

"NVIDIA AI Enterprise is primarily deployed in my organization through a hybrid cloud, which I have discussed earlier. Hybrid cloud combines both private and public on-prem solutions, offering the best of both worlds. Data that needs to stay on-prem can live in a secure environment while allowing for archival or secondary storage in the cloud, which can reduce costs. Working with a company such as Equinix for colocation of data back and forth plays a crucial role in the deployment as it provides a scalable, flexible approach, with private cloud environments making the most sense for the customers I work with. I am providing this review with an overall rating of 9.."

Josh Thias
AI Software Validation Engineer at a manufacturing company with 10,001+ employees

Read full review

"Absolutely, most enterprises depend on the CSP platforms, such as Azure or AWS or GCP. For regulated domains, such as healthcare, security, or other domains like financial, if they have a strategy of having sovereign AI, then NVIDIA AI Enterprise

is the way to go because you will get the resources and you will get the training that is required. Of course, it is a matter of skills in order to understand the AI architecture and implement them. However, I believe that sovereign AI is the way to go for the regulated industries and having an NVIDIA AI Enterprise license is definitely a way to cut down the time for enterprises to learn things and help them get on their journey and implement the strategies.

"I have not really delved into the governance aspect of NVIDIA, but security is pretty robust because everything that NVIDIA provides is very clear. If you are utilizing it for your enterprise applications, then they provide what are the guardrails and the best practices. If it is a POC, they say that it is okay to do this, but not for production. So, they have it all covered as far as security is concerned. Governance has definitely some blueprints for governance, but I have not used the governance aspect, so I cannot comment on that.

"As far as reliability and accuracy are concerned, that is something that the enterprise which is deploying the models and building apps and agents is responsible for it. NVIDIA provides the models and the microservices, everything, but again, everything depends on how they are implemented by the enterprise. So it is the responsibility of the enterprise or the AI architects of the enterprise in order to understand and lay it out as per the specs that they have.

"I would say definitely consider it. Look at the pros and cons and the needs of your organization. As far as my experience goes, it is a great solution. It also depends on what your infrastructure needs are, whether you are deploying it on-premises or you want to actually deploy it on the cloud yourself. So, it all depends.

"My experience was very positive. As I mentioned, it helped me start from scratch and the GPUs and also the NGC website that NVIDIA provides with all the resources are phenomenal. It was a highly enriching experience that I had with NVIDIA AI Enterprise. My rating for this solution is 9 out of 10.."

Samuel Beera
Principal Enterprise Architect at Mercy

Read full review

“My advice to others looking into NVIDIA AI Enterprise is to learn as much as they can and ask the right questions and be as open-minded as they can because this is a pretty new product, but it has a lot of upscale potential and it is going to create value for just about anybody. You have to be open-minded to that.

“The integration of NVIDIA AI Enterprise with AI frameworks on my project is basically the most important piece of the AI framework because it takes the AI possibilities and actually brings them to reality. It gives you the full capabilities that you would not have access to if you were not running this software because the GPUs alone are just there to help run parallel processor workloads. They are there to basically be resources to handle very intense streams of information running on the server. They are not really built by themselves to be completely optimized and customized without software on top of it. You have to run NVIDIA software to get the full benefit of the APIs, which are the application programmable interfaces, and all the specific use cases that you want, whether it is modeling a digital twin, which is what Omniverse does, and Omniverse is also a bundle.

“The thing about AI Enterprise is that it is an overarching term. There are several other AI softwares that NVIDIA has that they are actually running promotions with. When you buy AI Enterprise, you also get access. It changes on a somewhat regular basis, but there are promotions going on. Those promotions are additional packages such as SDKs or software development kits that have the ability to run more things. You might have heard of NVIDIA Omniverse or Run AI—those are two of the most common ones. They have had these deals where depending on what kind of GPU model you are getting, if you get the AI Enterprise software license, you are actually going to get the software, but you are also going to get additional software that is all tied together with AI Enterprise.

“I have NVIDIA AI Enterprise deployed in a hybrid model because when I was working with the federal government, security is a top priority. If I did do cloud, it would be a hybrid cloud because it would require some on-site presence with a little bit of remote or cloud orchestration. Hybrid is definitely the approach, and also SaaS because that gives the customer the ability to use it as they go with a consumption model without needing to pay for large upfront costs. Cloud can be expensive because if you don't keep track of your cloud resources, you can spend a

lot more money than you anticipate. People are moving to the cloud.

“My direct team using NVIDIA AI Enterprise is between 10 to 15 people, but we were supporting an entire sales organization that is 10,000 or more people. The actual team that was the specific sales team was about 10 to 15 and pretty much everybody is using it as best they can.

“I would rate this product a 9 out of 10..”

Verified user
Solutions Architect at a manufacturing company with 10,001+ employees

Read full review

“NVIDIA AI Enterprise platform typically rolls out quarterly updates where NVIDIA is pushing SDKs or software development kits that are going to be allowing customers for even more models and training. There is also a ton of free resources available. Even if customers do not want to purchase NVIDIA AI Enterprise, which is going to have a ton of bundles and save money on ROI and things regarding optimization of resources, they can also visit build.nvidia.com as a platform that is going to have API keys, and some of them are free. Customers can take care of some of those models and actually get some hands-on experience. There is a lot to learn online, and that is something that I really encourage everybody to do.

“NVIDIA AI Enterprise has a number of platforms that it can run on. It is built to be more or less vendor agnostic, which is very helpful. I know a couple of examples of that, such as Penguin Solutions, which is a big-time AI computing platform. They were working closely with NVIDIA because their hardware and their solutions can be layered with NVIDIA AI Enterprise, so that gives them a flexible approach and there is no vendor lock-in. Customers can simply take this approach and create their solution and make it custom. The flexibility and customization are incredible, especially with how quickly the market is moving. NVIDIA is making as many approaches to invest in these platforms early and often, so it is ahead of the curve. That gives customers the advantage of scalability as well as flexibility.

"NVIDIA AI Enterprise gives my team the ability to work on multiple projects at the same time. I would say it boosts productivity by two hours a day because one is not going to spend hours just researching products because NVIDIA AI Enterprise gives all of the resources and features available at fingertips. The productivity gains with time are significant, especially when making decisions about products, where I have saved about two hours a day.

"NVIDIA AI Enterprise is built on zero-trust security, so it has a lot of secured features, but it is not efficient for the government because it does not have all the right certifications like FIPS 140-2, 140-3, STIG hardening, data-at-rest encryption, and other types of cryptography keys. I would encourage NVIDIA to continue to invest in the governance world, especially with the US government, such as the Department of Defense.

"NVIDIA AI Enterprise is very accurate, and it continues to be trained on models that are highly effective and efficient, such as TensorFlow and TensorRT and different open-source models from Hugging Face. I expect it to continue to improve itself with reinforcement learning.

"For those considering implementation, I would make sure to provide as much detail as possible for their use case and actually be able to understand how their data center facility, floor layout, rack layout, and power and cooling requirements are set up. I gave this product a review rating of 9 out of 10.."

Verified user
Ai & Advanced Computing Engineer at a tech vendor with 10,001+ employees

Read full review

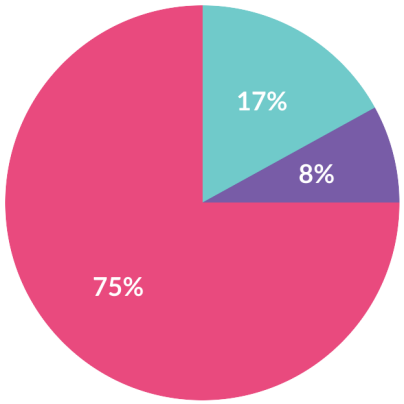
Top Industries

by visitors reading reviews

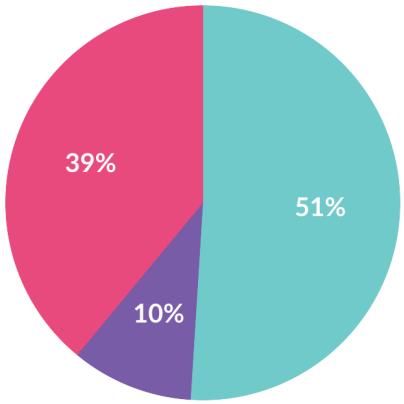
Outsourcing Company	16%
Comms Service Provider	9%
University	9%
Wholesaler/Distributor	9%

Company Size

by reviewers



by visitors reading reviews



Large Enterprise Midsize Enterprise Small Business

About this buyer's guide

Thanks for downloading this PeerSpot report.

The summaries, overviews and recaps in this report are all based on real user feedback and reviews collected by PeerSpot's team. Every reviewer on PeerSpot has been authenticated with our triple authentication process. This is done to ensure that every review provided is an unbiased review from a real user.

Get a custom version of this report... Personalized for you!

Please note that this is a generic report based on reviews and opinions from the collective PeerSpot community. We offer a customized report of solutions recommended for you based on:

- Your industry
- Company size
- Which solutions you're already considering

The customized report will include recommendations for you based on what other people like you are using and researching.

Answer a few questions in our short wizard to get your customized report.
Get your personalized report here

About PeerSpot

PeerSpot is the leading review site for cloud, AI, and business software. We created PeerSpot to provide a trusted platform to share information about software, applications, and services. Since 2012, over 22 million people have used PeerSpot to choose the right software for their business.

PeerSpot helps tech professionals by providing:

- A list of products recommended by real users
- In-depth reviews, including pros and cons
- Specific information to help you choose the best vendor for your needs

Use PeerSpot to:

- Read and post reviews of products
- Access over 30,000 buyer's guides and comparison reports
- Request or share information about functionality, quality, and pricing

Join PeerSpot to connect with peers to help you:

- Get immediate answers to questions
- Validate vendor claims
- Exchange tips for getting the best deals with vendor

Visit PeerSpot: www.peerspot.com

PeerSpot

244 5th Avenue, Suite R-230 • New York, NY 10001

reports@peerspot.com

+1 646.328.1944