TATA COMMUNICATIONS

TATA

# TATA COMMUNICATIONS
# VAYU AI CLOUD

## UNIFIED | EFFORTLESS | TRUSTED

In today's fast-evolving digital landscape, artificial intelligence is no longer a luxury but a necessity. With the explosion of data and the rise of generative AI (Gen-AI) applications, surges the demand for advanced AI models capable of processing vast amounts of information, generating insights, and driving innovation at scale.


vayu AI Cloud

# BREAK THE SILOS, SIMPLIFY THE PROCESS, AND TRUST EVERY OUTCOME

With Tata Communications Vayu AI Cloud, you have the power to build your own AI "super factory," unlocking the full potential of artificial intelligence. Our platform offers scalable GPU resources and state-of-the-art AI tools, empowering you to develop, train, and deploy AI models at a scale.

## DRIVE INNOVATION WITH A UNIFIED, EFFORTLESS, AND TRUSTED SOLUTION

| UNIFIED | EFFORTLESS | TRUSTED |
|---|---|---|
| End-to-end platform with tools for data management, experimentation, training, fine-tunning, and deployment. | Rapid deployment with minimal setup. | Built with core principles of national data control. |
| Harmoniously connects data, tools, and AI workflows. | Effortlessly expand resources to match your evolving needs. | Meets global security and regulatory standards. |
| Consistent performance across all environments, from centralised cloud to decentralised edge. | Fully automated and managed, removing complexity from your operations. | Defense-grade protection woven into every layer. |

Accelerate innovation, streamline operations, and bring AI-driven solutions to market faster by leveraging this cutting-edge cloud infrastructure, whether you're tackling complex machine learning algorithms, developing generative AI (Gen-AI) large language models (LLMs), exploring multi-modal use cases, or advancing computer vision applications, **Tata Communications Vayu AI Cloud** provides the flexibility, performance, and security to transform your boldest AI visions into reality.

We offer NVIDIA GPU instances and clusters specifically designed for AI training, fine-tuning, and inference, enabling our customers to harness powerful computing resources tailored to their needs. Access the latest NVIDIA GPUs, equipped with InfiniBand GPU-to-GPU non-blocking networks and high-speed parallel file systems, optimised for deep learning and machine learning tasks.

This ensures high performance and efficiency during model training and inference, while giving you the flexibility to scale computing resources up or down based on project requirements—minimising costs and maximising productivity.

There's no need to invest in expensive hardware upfront; simply spin up GPU instances as needed, whether for short-term projects or long-term initiatives. This flexibility is especially valuable for enterprises, researchers, and developers who need the ability to experiment with different models, run complex simulations, or process large datasets without the limitations of traditional infrastructure.

Additionally, with streamlined access to popular AI frameworks and tools, you can avoid the constraints of proprietary systems and vendor lock-in, empowering you to accelerate your AI development and deployment cycles.

## TATA COMMUNICATIONS VAYU AI CLOUD – STACK VIEW

**AI APPLICATIONS**

Industry leading AI solutions and usecases

**AI STUDIO**

- AI workbench
- AI supermarket
- Serverless AI
- Data management
- MLOps/GenAIOps
- Responsible AI

**TCˣ: CLOUD MANAGEMENT PLATFORM**

Orchestration, AI tooling, PaaS, Security

**SOVEREIGN INFRASTRUCTURE**

**Tata Communications Vayu Cloud**

- Hybrid cloud
- Private cloud

**Tata Communications Vayu Edge**

**Cloud-to-edge continuum**

# UNIFIED

Harness AI's potential through our comprehensive platform that delivers a powerful suite of capabilities featuring large-scale GPU computing, LLM Ops, and serverless functions, enabling enterprises for building, training, deploying, and running Generative AI (Gen AI) applications and use cases.

Get access to popular AI frameworks and tools, advanced data management, and a dynamic AI ecosystem that fosters innovation and collaboration.

We provide a diverse range of models, including leading options like Cohere, Mistral, and Llama 2, with real-time benchmarking tools for seamless performance comparisons. This enables data-driven decisions, ensuring you select the optimal model for your specific needs with confidence and flexibility.

Additionally, with seamless integration to Hugging Face, you gain access to an extensive library of pre-trained models and datasets, expanding your options for pow ering AI applications. This broad selection enables you to experiment, innovate, and deploy AI solutions quickly and efficiently, all within a single platform that takes care of the heavy lifting.

Our platform enables scalable AI model inference, whether in the cloud or at the edge, ensuring optimal deployment wherever needed. With the ability to handle large volumes of data and deliver real-time insights, our solution is perfect for high-demand, low-latency applications across industries. Whether you're processing data in a centralised cloud environment or require fast responses at the edge—such as in IoT devices or remote locations—our platform guarantees seamless, efficient performance.

# EFFORTLESS

Tata Communications Vayu AI Cloud is a fully managed service that provides effortless access to cutting-edge AI models, eliminating the need for infrastructure management and enabling you to focus on innovation.

Our platform streamlines working with Gen AI models, enabling you to build advanced AI applications without deep technical expertise. From development to deployment, every step is optimised for ease, allowing businesses to quickly implement AI solutions. Whether automating content, enhancing customer experiences, or developing new AI products, our platform supports a wide range of use cases, so you can focus on delivering value, not managing complexity.

Our Serverless AI platform makes deploying AI models as APIs effortless. It transforms complex models into scalable, accessible services that integrate seamlessly into any application, without the need to manage or scale the underlying infrastructure.

Deploying your models as APIs enables real-time interaction with other applications or users, whether for generating insights, automating tasks, or enhancing user experiences.

This API-driven approach simplifies sharing AI capabilities, making your models easily accessible across various platforms without requiring deep infrastructure knowledge. It accelerates time to market, ensuring your AI solutions are efficient, scalable, and ready for any environment.

Perform live evaluations of AI models to gain real-time insights into their performance and suitability for your use cases. This feature allows you to assess models under real-world conditions, ensuring they meet your operational standards before full deployment. You can also fine-tune models with your own data, customising them to your specific business needs and optimising their accuracy and relevance.

Beyond evaluation and fine-tuning, you can easily set up Retrieval-Augmented Generation (RAG) workflows that integrate seamlessly with your enterprise data sources. This powerful combination enables AI models to retrieve relevant information from large datasets and generate more accurate and contextually relevant outputs. RAG workflows are especially useful for applications like customer support automation, knowledge management, and personalised content generation, helping you get the most out of your AI-driven solutions by connecting them directly to your proprietary data.

# TRUSTED

Our platform is secure by design and sovereignty at its core, with built-in safeguards to protect against harmful content, including racial, gender, or explicit bias, ensuring responsible AI development. Additionally, our explainability tools offer tracing capabilities, enhancing transparency and giving you greater control over the decision-making process of your AI models. Experience superior performance without compromising security for infrastructure, data and AI Models.

Our platform combines integrated data, model, and prompt security with zero-trust policies and trusted governance to ensure robust protection at every layer. With features like a comprehensive data catalog and sandboxed environments, you can manage and experiment with data safely.  We also provide social guardrails that promote responsible AI usage, protecting against harmful biases and content. Additionally, our platform includes integrated tools for understanding AI model decisions, offering transparency and control over the decision-making process, so you can confidently deploy ethical and accountable AI solutions.

Our solution is designed to deliver fast, reliable results while ensuring that all outputs are secure and compliant with your enterprise's data protection standards. This combination of scalability, performance, and security gives you the confidence to deploy AI across a wide range of environments, from large-scale cloud systems to localised edge devices, without sacrificing control or protection.

**For more information, visit us at www.tatacommunications.com**

CONTACT