

PaletteAI from Spectro Cloud

The enterprise platform to design, deploy and manage AI environments at scale.

The situation

Enterprise infrastructure is undergoing its fastest transformation in decades.

AI has become the defining force reshaping infrastructure, software ecosystems, and applications — all at once.

Gartner estimates that 80% of existing applications will be rewritten to embed AI over the next few years.

To maintain competitive advantage, enterprises are racing to build AI-ready infrastructure, taking advantage of exponential advances in GPU performance, and new technologies like DPUs, which have emerged as a critical enabler of zero-trust computing and efficient data flow.

Riding this AI wave is easier said than done. ROI is hard to come by. Overprovisioning and underutilization mean scarce, valuable resources are sitting idle, resulting in millions in wasted enterprise spend.

Friction between teams is also growing.

AI and ML teams of data scientists and developers push for innovation, speed, and access to the latest tools. The AI application ecosystem is already 2.5x larger than the cloud-native ecosystem, and evolving rapidly.

At the same time, **platform teams** demand control, compliance, and cost efficiency, trying to bring order to chaos, but they struggle to control the sprawl of disconnected environments and “shadow AI” that has sprung up.

What’s the answer?

Enterprises need a unified foundation that bridges these competing needs: a solution that provides freedom and flexibility without sacrificing governance, security, or ROI.

The solution

With PaletteAI, enterprises can build repeatable, reliable, and secure AI workloads that accelerate innovation while maintaining operational discipline.

PaletteAI gives platform teams and AI practitioners a shared environment to design, deploy, and manage AI workloads from data center to edge with speed, confidence, and control.

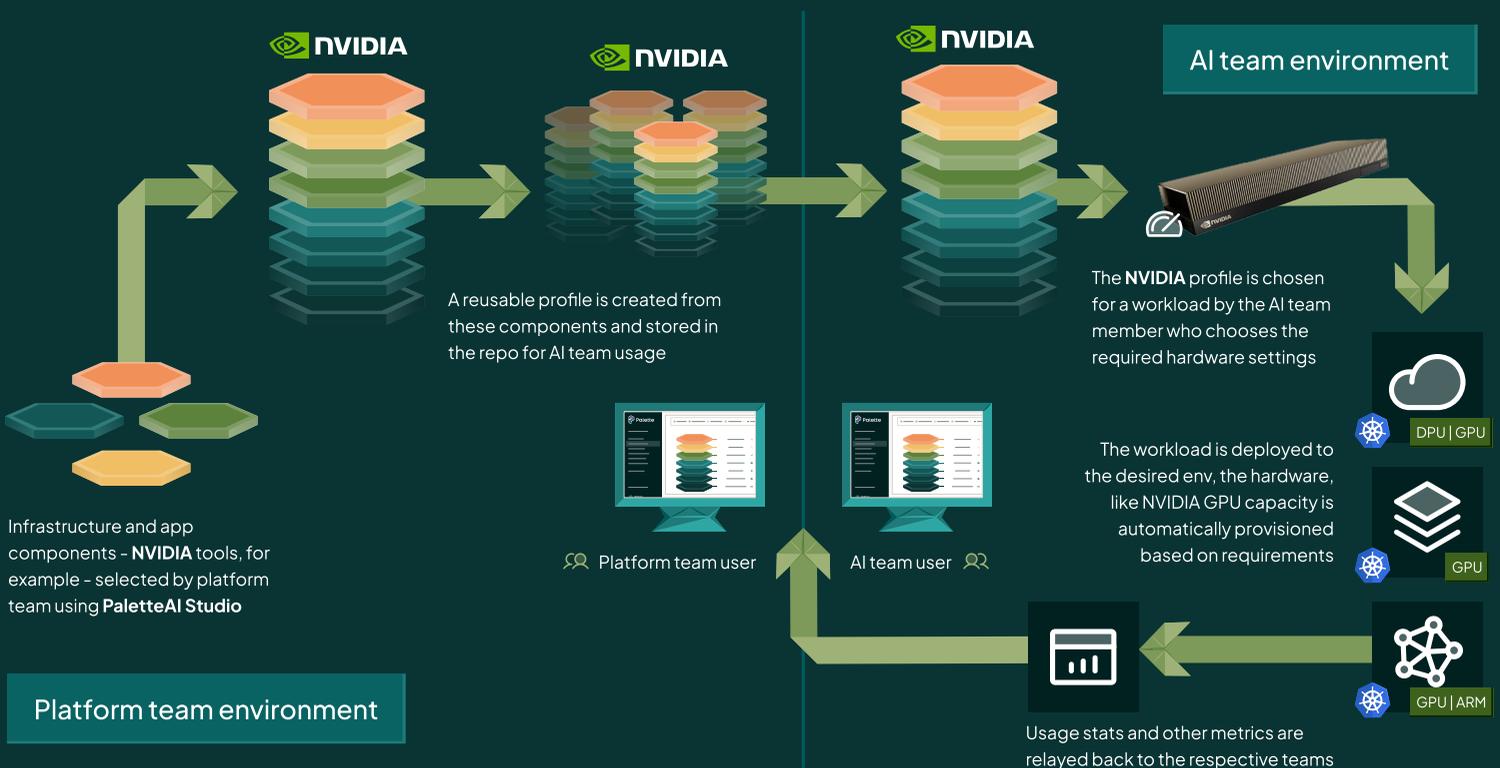
It includes distinct interfaces for platform teams and practitioners. Platform teams make available cluster templates in PaletteAI for AI teams to use. These clusters are built on standardized reference designs from NVIDIA and include all the essential infrastructure layers: operating system, Kubernetes, networking, and storage, providing a consistent and reliable foundation for AI workloads. Platform teams also curate workload profiles: preconfigured AI and ML applications enhanced with corporate IT guardrails.

Practitioners can then easily customize and deploy these workloads on the available clusters, accelerating innovation while maintaining compliance and operational consistency.

PaletteAI also supports deployment of NVIDIA-powered AI factories, enabling enterprises to bring NVIDIA's AI Factory vision to life at scale.

With native integration across the NVIDIA AI Enterprise suite including NeMo and NIM, as well as NVIDIA Run:AI, PaletteAI simplifies the rollout and management of complete AI factory environments, unifying NVIDIA Blackwell GPUs and Bluefield DPU automation with standardized, repeatable stacks across data centers, clouds, and the edge.

In short, platform teams can design guardrailed templates and policies, while AI teams can deploy freely within approved boundaries.



How it works

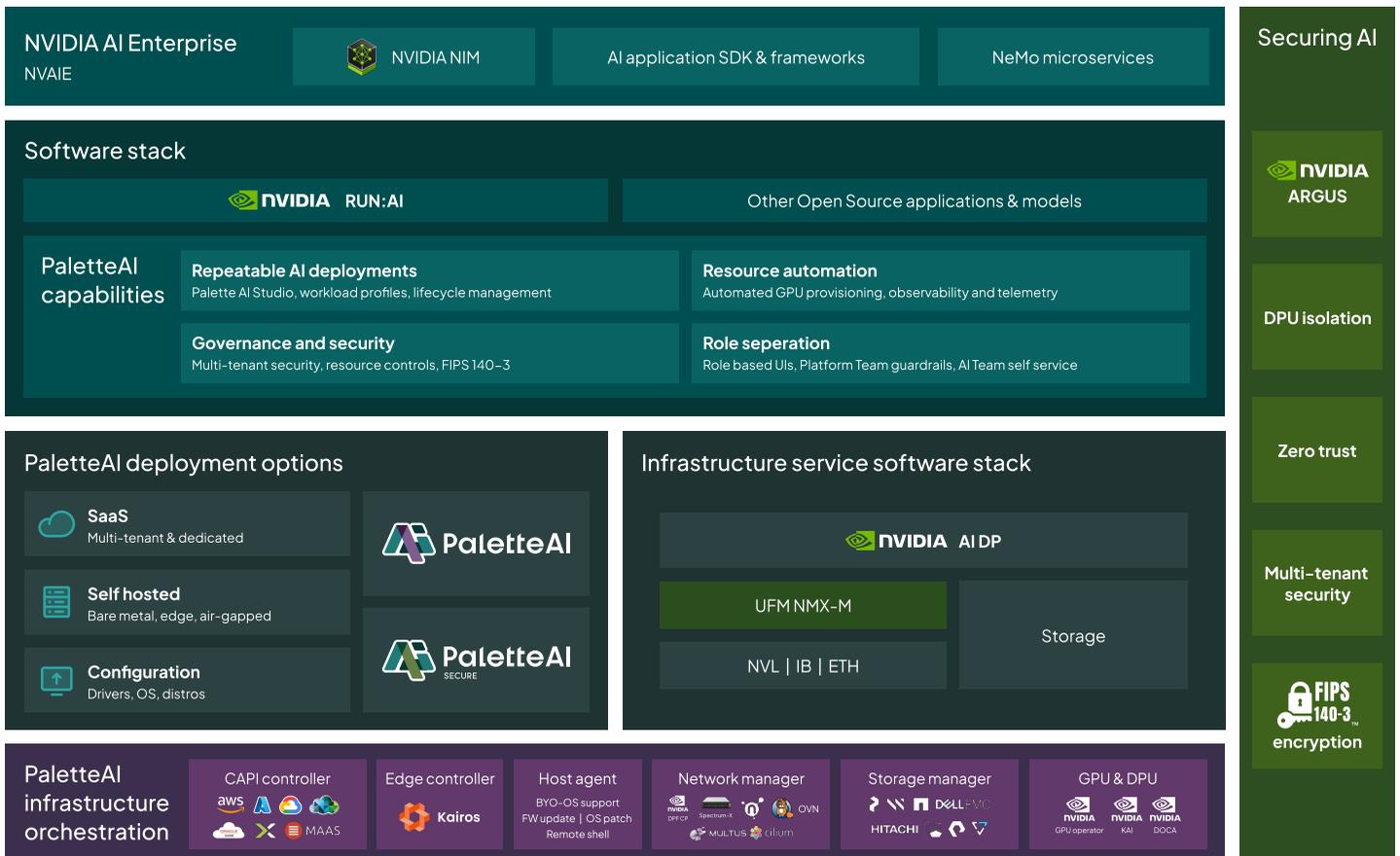
PaletteAI brings together platform and practitioner workflows into a unified system for building, deploying, and managing AI environments with consistency, security, and scale.

It includes separate interfaces for platform teams and practitioners. Platform teams can define complete workload templates that span the full stack, from the operating system and Kubernetes distribution to storage and networking, combining these layers with powerful AI and ML tools to create ready-to-deploy AI workloads. Practitioners can then easily customize and launch these templates whenever and wherever they're needed.

PaletteAI provides a complete operational model that connects design, governance, and execution, giving enterprises a clear path from AI concept to production.

Its core capabilities create a seamless flow: platform teams design, AI teams deploy, PaletteAI automates, and organizations scale with confidence.

- **Intelligent efficiency:** Cost and resource efficiency is built in, with automated GPU and DPU allocation, right-sizing, and policy-based optimization.
- **PaletteAI Studio:** This tool lets platform teams design, publish, and govern full AI and ML stacks with prebuilt NVIDIA components such as NeMo, NIM, and DOCA.
- **Workload blueprints:** Template-driven deployments define the complete stack, from infrastructure to applications, enabling repeatable, declarative consistency.
- **Role separation:** Platform teams set guardrails, policies, and cost controls, while AI teams deploy approved templates through self-service.
- **Lifecycle management:** Provisioning, scaling, monitoring, and updates are automated across environments, reducing toil.
- **Unified governance:** observability tools deliver centralized visibility, usage insights, quotas, and compliance controls across SaaS, self-hosted, and air-gapped environments.



Why choose PaletteAI?

When you choose Spectro Cloud and PaletteAI, you're picking a best in class solution that delivers both speed and control, enabling innovation at scale with guardrails intact. We offer:

- **Accelerated ROI:** eliminating idle GPU spend, right-sizing infrastructure, and shortening pilot-to-production timelines.
- **Freedom of choice:** deploy across clouds, edge, and data center with no vendor lock-in.
- **Operational efficiency:** standardized, automated workflows and reusable blueprints reduce manual toil and human error.
- **Enterprise governance and trust:** built-in policy enforcement, auditability, and zero-trust DPU-backed security for consistent compliance at scale.
- **NVIDIA-validated architecture:** PaletteAI is built on NVIDIA AI Enterprise, supporting the latest Blackwell GPUs, Grace CPUs, BlueField DPUs, and Spectrum-X networking.
- **The fast path to NVIDIA's AI factory:** supporting evolving AI frameworks such as NIM, NeMo, and Run:AI.
- **Strong industry partnerships:** we're an NVIDIA preferred partner, and work with leaders in the AI ecosystem including HPE, Supermicro and WEKA.
- **World-class support:** We back every PaletteAI customer with global 24x7 support for your full software stack.

Start your journey with PaletteAI

Visit palette-ai.com to learn more about PaletteAI and arrange a live, hands-on demo tailored to your business and use case.

