

Solution Brief: Private, Zero Trust, Fully Managed, LLM for Generative AI

Take advantage of Generative AI with a Large Language Model (LLM) that is easy to install, configure, maintain and manage

Summary

Aarna Networks, Predera, and NetFoundry have partnered to offer a secure, private, fully managed large language model to help you explore the world of generative AI. Choose from a variety of foundational models that you can fine tune with your corporate data. Create value for your organization by becoming a practitioner of this new technology that has taken the world by storm by discovering new insights and revenue generating opportunities in your data.

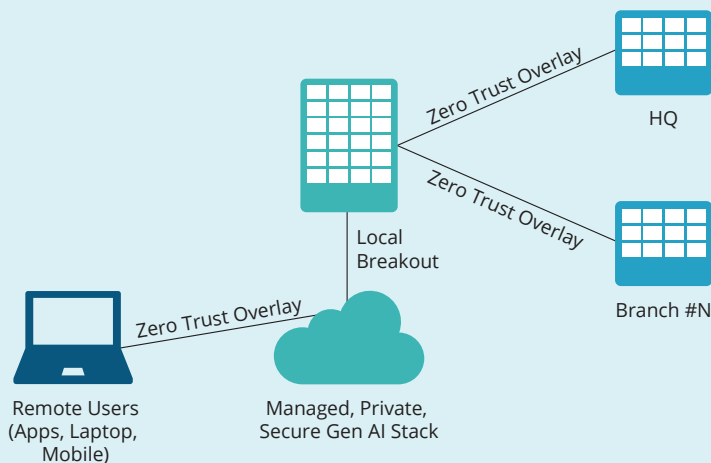
Data scientists and AI/ML architects can take advantage of:

- ▶ A zero touch, fully managed environment (installation, configuration, upgrade, management)
- ▶ Full control and customization of the model
- ▶ 0% chance of data leakage with private environment
- ▶ Potential to access sensitive data securely via a private connection
- ▶ A zero trust environment hosted in an Equinix data center with no inbound ports via a SW-only zero trust overlay ensuring easy access similar to managed commercial but only for users explicitly authenticated and authorized before connectivity
- ▶ Significant cost savings over systems integrators or commercially managed services

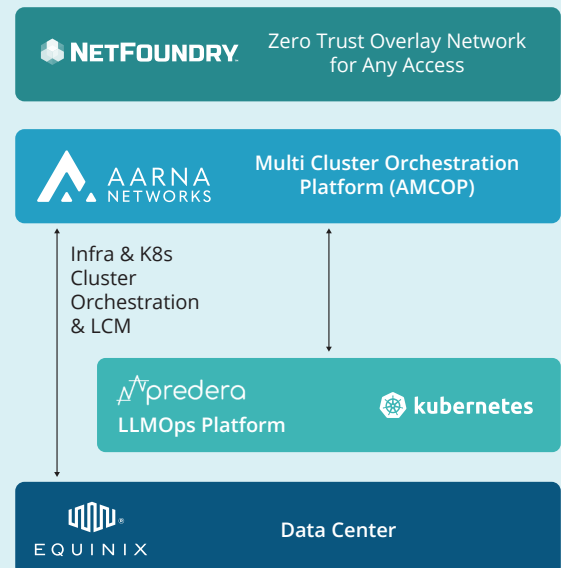
What's Included

- ▶ Compute (on Equinix Metal)
 - ▶ 2 x Intel Xeon 6338, 64 cores @ 2.0 GHz, 2 x 240 GB SSD, 1 TB, 4 x 25 Gbps, OR 2 x Intel Xeon Gold 6338, 64 cores @ 2.0 GHz, 2 x 240 GB SSD, 512 GB, 4 x 25 Gbps, NVidia A100
- ▶ A private, fully managed Large Language Model (LLM)
 - ▶ **Llama** OR **Dolly** OR **NeMo** with zero touch orchestration, configuration, management of upgrades/updates
 - ▶ Management dashboard
- ▶ 1 software-only zero trust overlay network which makes the LLM fully private with no inbound firewall ports
- ▶ Professional Services: 80 hours of for data labeling, prompt engineering, model fine tuning, RLHF, ethics/risk analysis, and more
- ▶ Operational Support: 8 hours a day x 5 days a week

Topology



Solution Stack



Timeline & Process

- ▶ **Day 1** 90 Minute Kickoff Call
- ▶ **Week 1** Package Onramp and Training¹
- ▶ **Week 2** Model Testing and Fine Tuning
- ▶ **Weeks 2+** (from 6-12 months) User Driven GenAI Innovation with Aarna/Predera support (up to 80 hours)

Why This Approach?

Deploying a GenAI foundational model from scratch is quite cumbersome and time consuming (to install, configure, upgrade, and manage the stack), and requires AI talent that few organizations have. Using a public LLM model is fraught with security and privacy risks that are not fully understood yet.

The **LLM for Generative AI** offering from Aarna Networks and Predera provides you with a secure, private, and fully managed LLM that is quick and easy to install, configure, maintain and manage with ongoing support as needed at a fraction of the cost of systems integrators or a commercially GenAI offering.

Results Expected

Create value for your organization by becoming a GenAI practitioner. You will become confident and proficient in using LLMs at a level deeper than just prompt engineering and will actually put LLMs into action by fine tuning it with your secured corporate data. You will be able to measure the impact of LLMs on your business, customize the model to your own KPIs and metrics, discover new insights, and seek new revenue generation opportunities.

How to get Started

Contact us to review the pricing models, review the services agreement, and get started with your journey with Generative AI today.

¹ A private, secure connection to your data might take additional time and effort. Contact us for more information.