# Mi-AAIMS

## Affordable AI Model Fine-Tuning System

Mi-AAIMS provides an integrated AI platform that streamlines the process of developing, fine-tuning, deploying, and monitoring large language models (LLMs) for enterprise and government use. Designed with efficiency, security, and scalability in mind, it supports on-premise deployment for full data sovereignty, offers cost-effective fine-tuning, and enables scalable, edge-ready inference. Mi-AAIMS enables organisations to fully leverage the potential of AI in a secure, affordable, and scalable manner.

### Technology Overview

Mi-AAIMS is an end-to-end AI platform designed for enterprises and government agencies to efficiently manage large language models (LLMs). It streamlines development, fine-tuning, deployment, and monitoring with a focus on cost-efficiency, performance, and scalability.

Mi-AAIMS ensures full data control while enabling flexible, secure, and scalable AI solutions for both centralised and distributed environments with on premise deployment, affordable fine-tuning, and edge-ready inference.

### Technology Benefits

- **Cost Efficiency:** NVMe memory offloading to reduce GPU reliance and cut operational costs with ideal for budget-friendly fine-tuning of LLMs.
- **Improved Data Quality:** Built-in automated data prep tools ensure clean, enriched datasets for better accuracy and training efficiency.
- **Scalability and Flexibility:** Dynamic inference nodes and edge deployment support low-latency, high-performance AI across any environment.
- **Optimised Resource Utilisation:** Maximises output with minimal infrastructure, perfect for resource-constrained organisations.
- **Faster Time-to-Insight:** End-to-end automation speeds up development, enabling quicker time-to-value.
- Adaptability Across Use Cases: Supports centralised to edge AI use cases, offering unmatched adaptability in diverse settings.

### Key Features

- **End-to-End LLMOps**
  Delivers a comprehensive suite of tools and processes to manage the entire lifecycle of Large Language Models (LLMs) from development to deployment.
- **LLM Fine Tuning with NVMe memory Offloading**
  Dramatically reduces GPU dependency and costs during the fine-tuning process.
- **Flexible and Scalable Inference Nodes**
  Supports scalable inference nodes that can grow with your needs from light workloads to enterprise-scale operations.

### Applications

- Enterprise
- Government




**MIMOS Berhad** 199501006785 (336183-H)
Taman Teknologi MRANTI, Bukit Jalil,
57000 Kuala Lumpur, Wilayah Persekutuan
Tel: +603 8995 5000 | Fax: +603 8996 2755
**www.mimos.my**


mimosmalaysia