

# Advancing AI Computing with Enterprise Software and Infrastructure Solution

邁爾凌科技 MyelinTek Inc.

技術副總VPE 李威儒Weiru Lee

# Company Profile

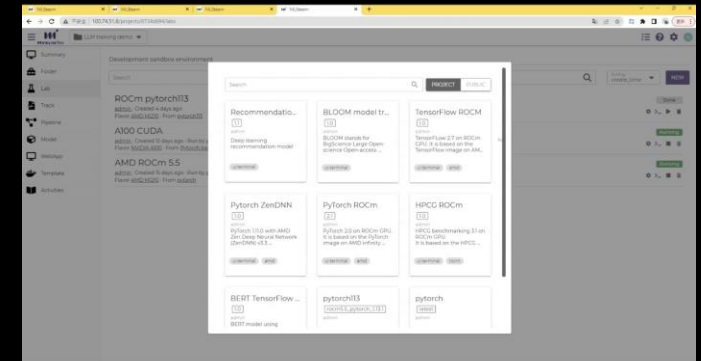


MyelinTek, an AI platform software company, was spun off from the Industrial Technology Research Institute (ITRI) in 2020.

We independently developed the MLOps platform – MLSteam, which is applied in the development, training, and inference management of AI/ML models.

MyelinTek also offers professional services, including AI/ML implementation across various industries, design and implementation of private/public cloud services, software and hardware platform integration, cluster management, and performance optimization.

Our clientele spans sectors such as enterprise, government, healthcare, and transportation.

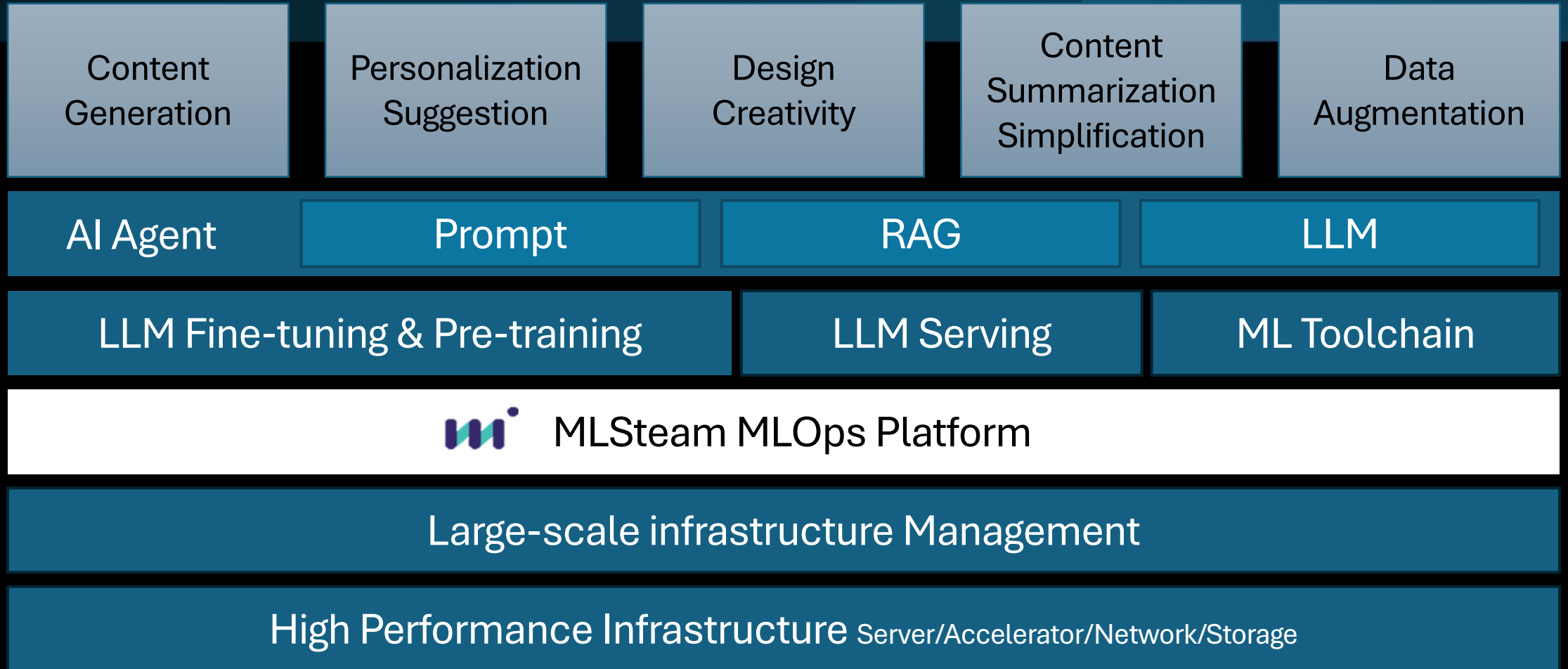


# GenAI Serving



- High Computational Requirements
- Real-Time Data Processing Challenges
- Low-Latency Networking Needs
- Integration with Existing Systems
- Data Security and Compliance Concerns
- Reliability and Fault Tolerance
- Monitoring and Managing Complex AI Systems
- Energy Consumption and Efficiency
- Support for AI Frameworks and Tools
- Data Management and Storage

# GenAI Serving



# MLSteam - Model Serving

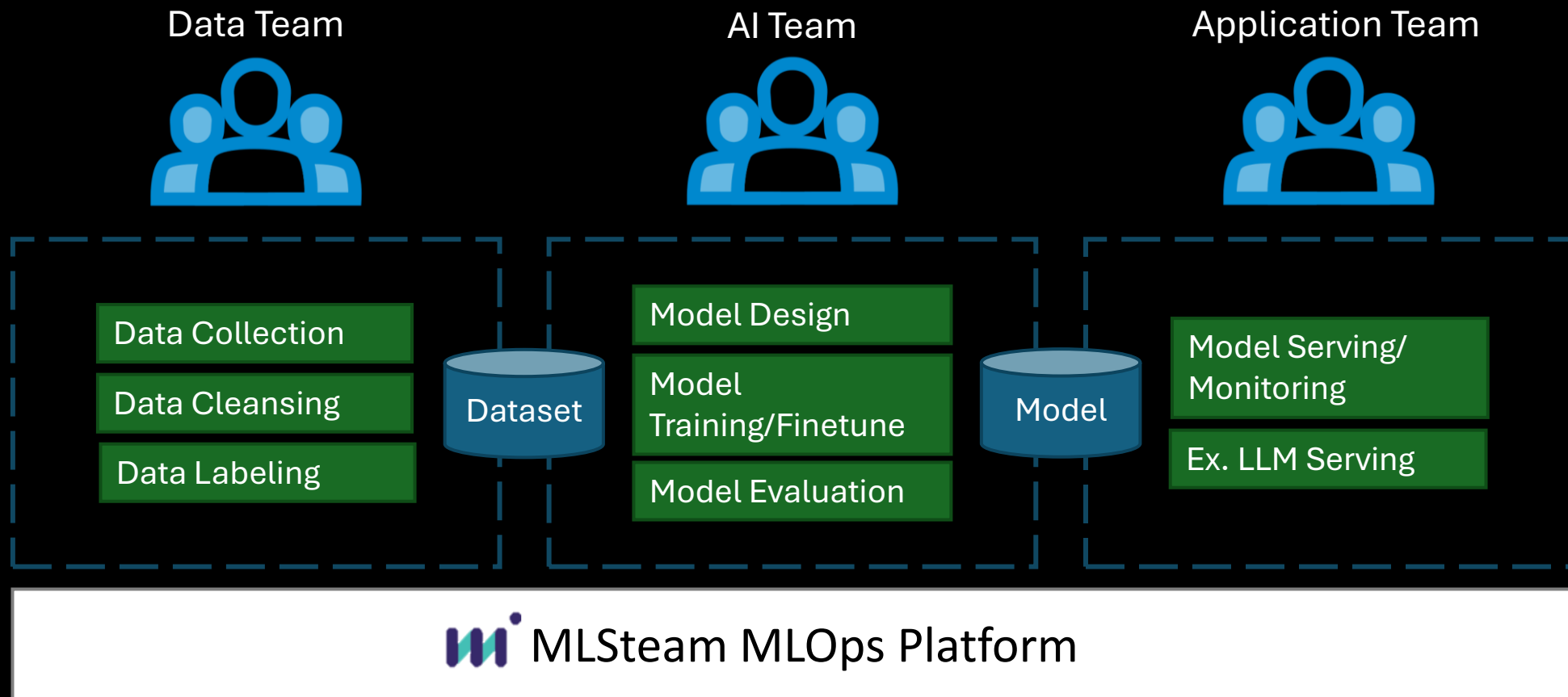
- Expose endpoint to external access
  - User can limit the request frequency
  - Easy to expose OpenAI API compatible LLM.
  - LLM Web Application Template
- Supercharge LLM performance by using vLLM on ROCm 6.2
  - Optimized LLM serving with vLLM.
  - Enhanced performance with ROCm 6.2 on AMD GPUs.
  - Achieving higher throughput (tokens/s) for AI applications.



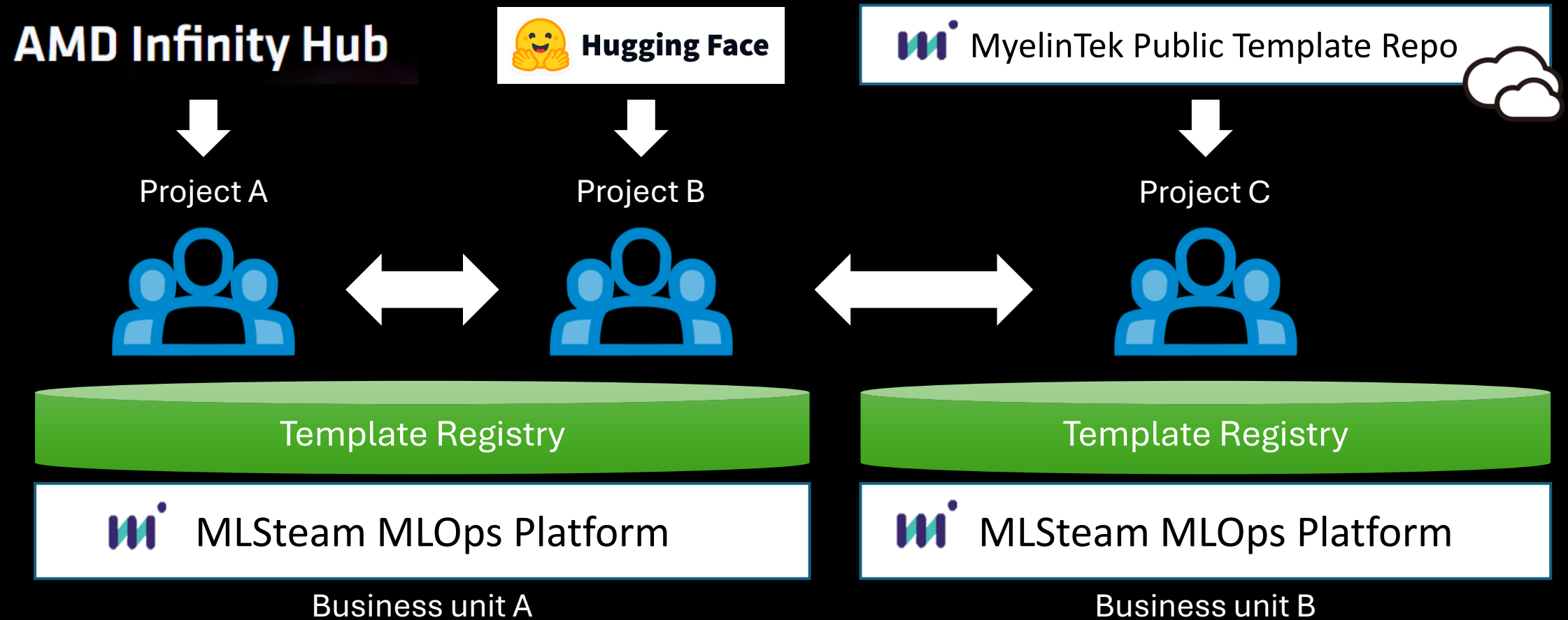
+



# MLSteam - Project Collaboration

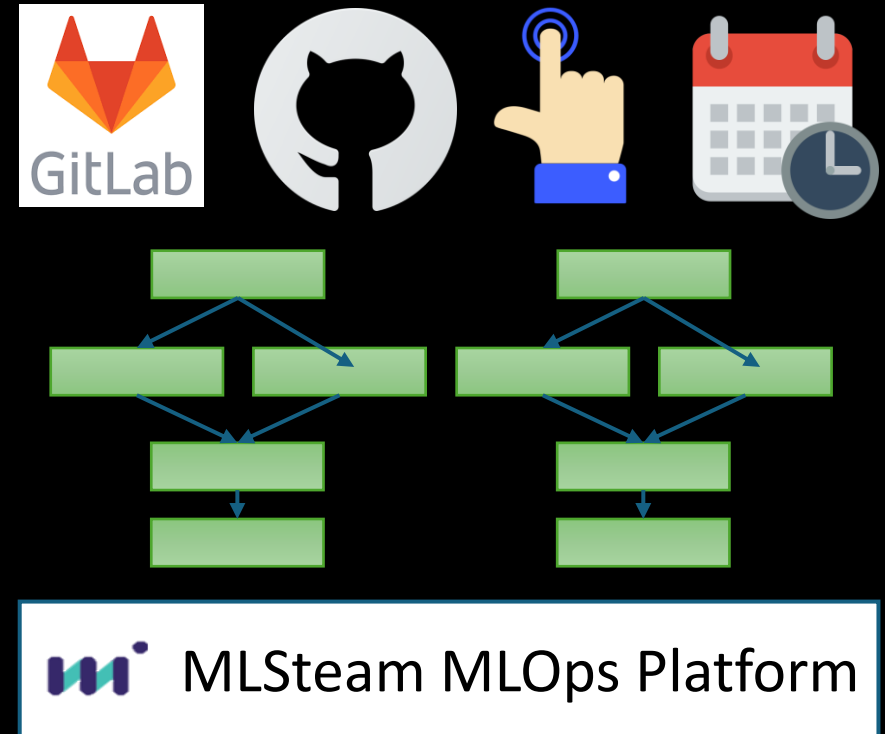


# MLSteam - AI/ML Template Sharing



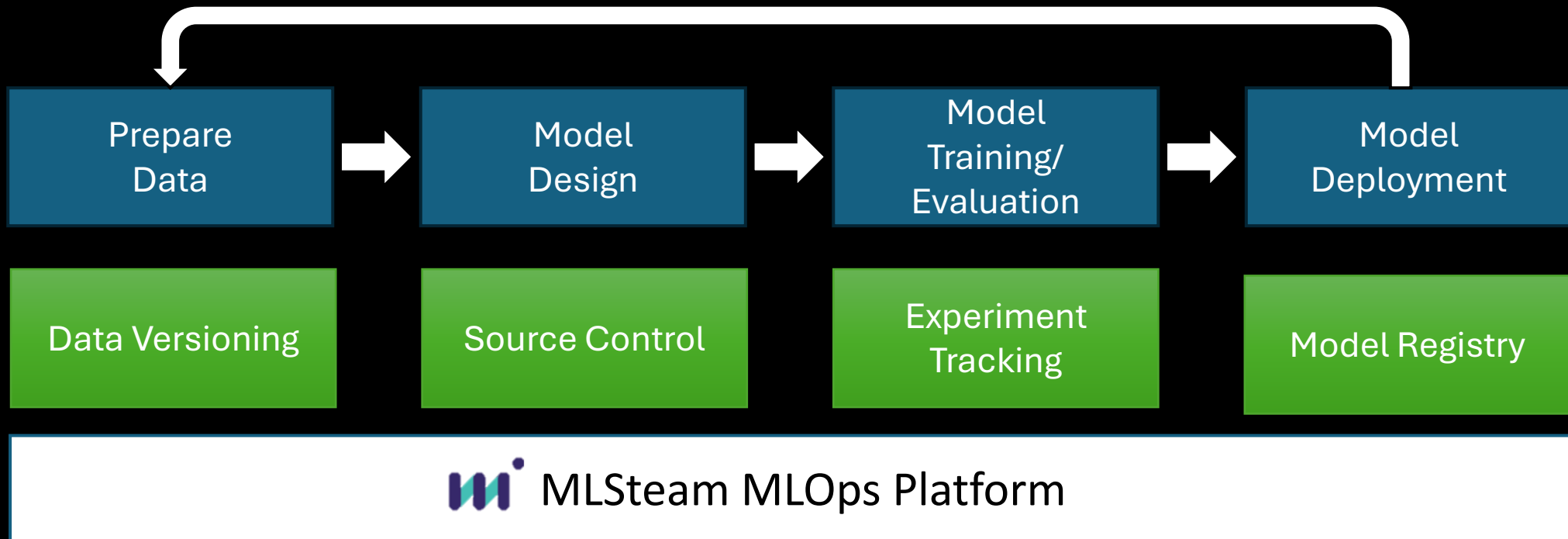
# MLSteam - Pipeline Automation

- CI/CD Integration for continues training (CT)
- Visual progress indicators and logs.
- On-demand resource scheduling to maximize resource utilization.
- Easily define execution dependencies (DAG) using a YAML file.

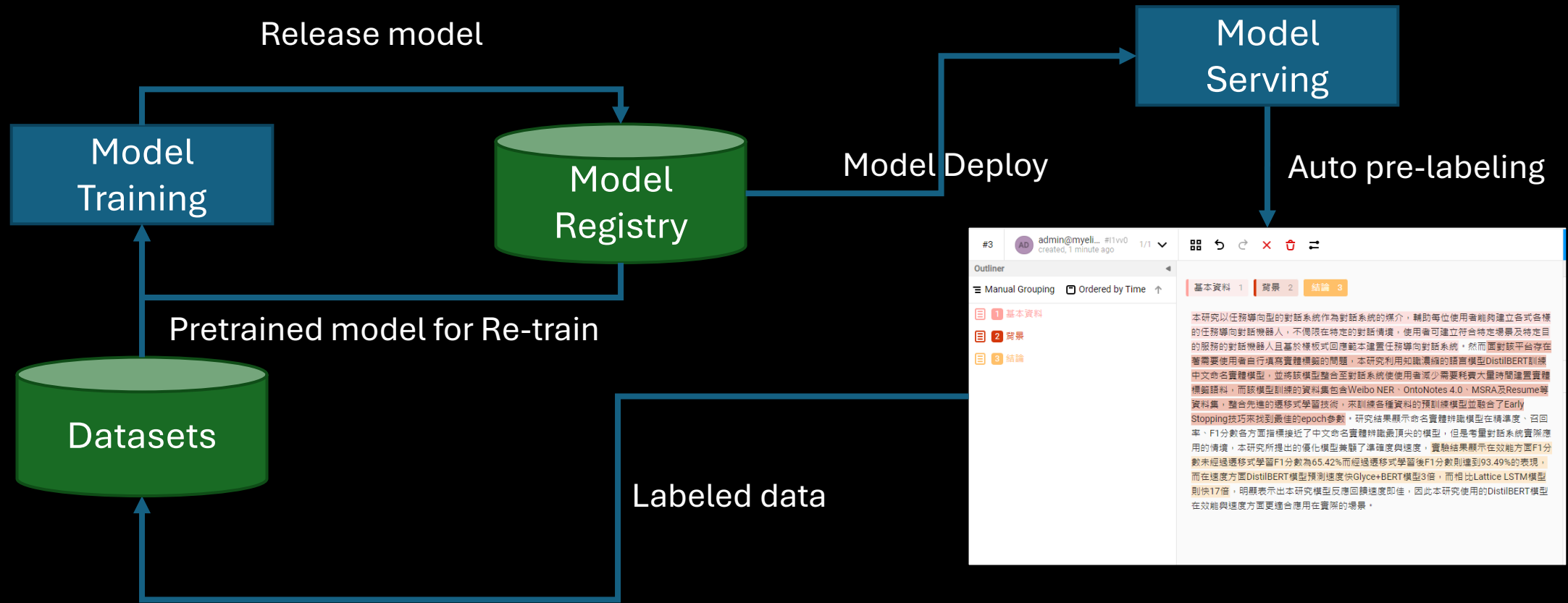




# MLSteam - Version Control Everything



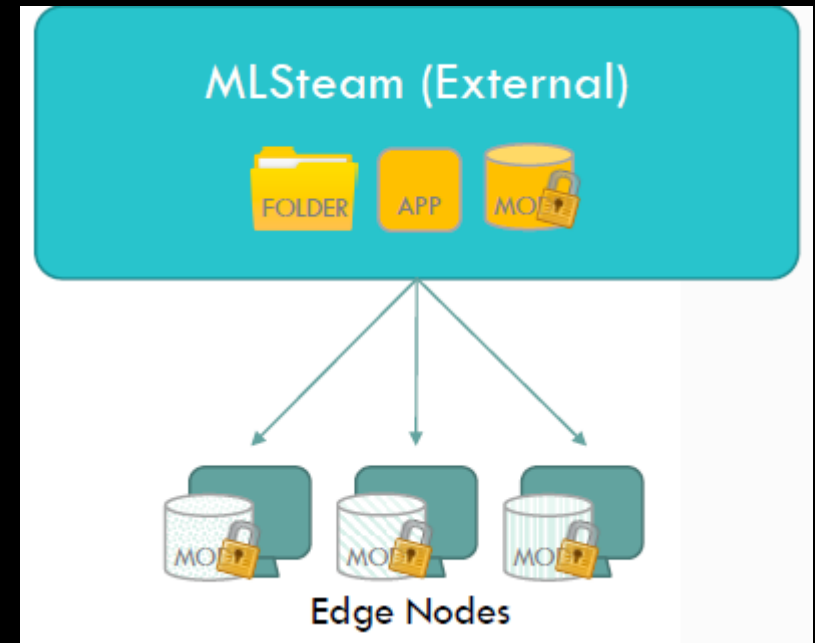
# MLSteam - End-to-End Model Lifecycle Management



 MLSteam MLOps Platform

# MLSteam - Security

- Isolated execution environment
- Dataset permission control
- Model deployment protection Mechanism



# Resource Management and Sharing



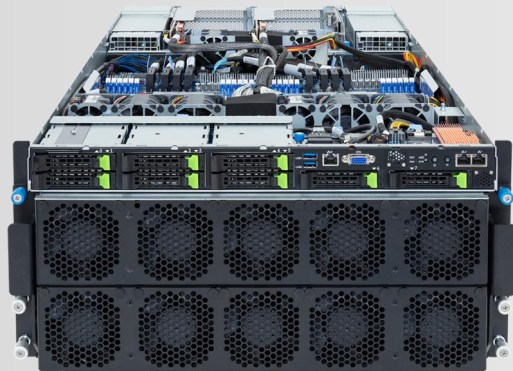
## Policy Base Resource Management

- Reserved/OnDemand GPU allocation
- GPU fraction management
- GPU-Time based quota

## Monitoring

- Resource consumption table
- Task Resource allocation table
- Global/Task Resource Utilization

 MLSteam MLOps Platform + Pod Management



# GIGABYTE Server Powering the Future with Full AMD GPU Support

AMD MI325X

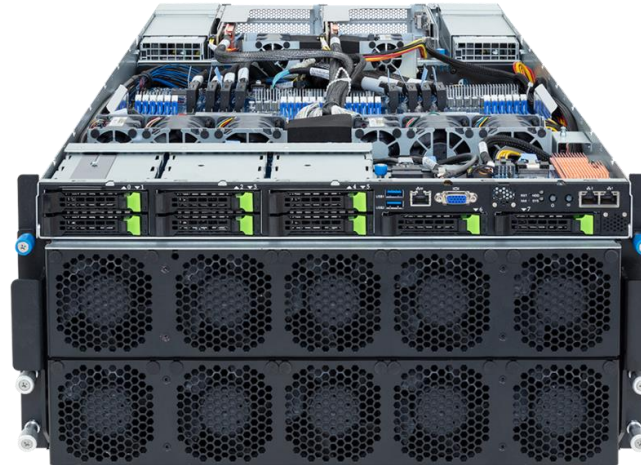


**GIGABYTE G893-ZX1**

**Now at Booth**

Height: 8U  
CPU: 2 x AMD Turin/Genoa platform  
Storage: 8 x 2.5 NVMe  
Expansion slot: 12 x FHHL PCIe5.0 Slots  
**GPU: 8 x AMD MI325X**  
PSU: 6+6 x 3000W & 4+4 x 5200W

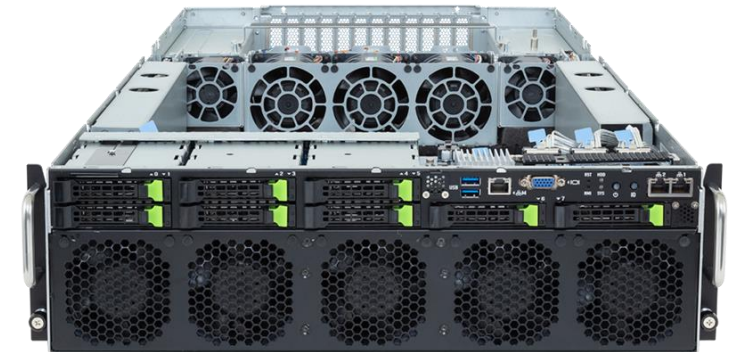
AMD MI300X



**GIGABYTE G593-ZX1**

Height: 5U  
CPU: 2 x AMD Turin/Genoa platform  
Storage: 8 x 2.5 NVMe  
Expansion slot: 8 x Low-profile PCIe5.0 Slots  
Expansion slot: 4 x FHHL PCIe5.0 Slots  
**GPU: 8 x AMD MI300X**  
PSU: 4+2 x 3000W

AMD MI300A



**GIGABYTE G383-R80**

Height: 3U  
**APU: 4 x AMD MI300A**  
Storage: 8 x 2.5 NVMe  
Expansion slot: 8 x FHFL PCIe5.0 Slots  
PSU: 3+1 x 3000W

Action Now



[contact@myelintek.com](mailto:contact@myelintek.com)

