Red Hat Summit Connect

Unified AI Platform for Free World

November 2025

Fatih E. Nar
Distinguished Chief Architect
Red Hat CTO Office

Fatih E. Nar, has built a career, solving complex challenges in various domains including telecom, entertainment, media and others. With experiences at Google, Verizon Wireless, Canonical Ubuntu, Ericsson, and now Red Hat, he specializes in Cloud-Native and Data & AI-driven solutions for enterprises and service providers.

He holds an MSc in Information Technology and a BSc in Electronics Engineering, along with completed AI studies at MIT & Stanford, he is admitted to Purdue University for Doctorate Program for 2026 Spring.

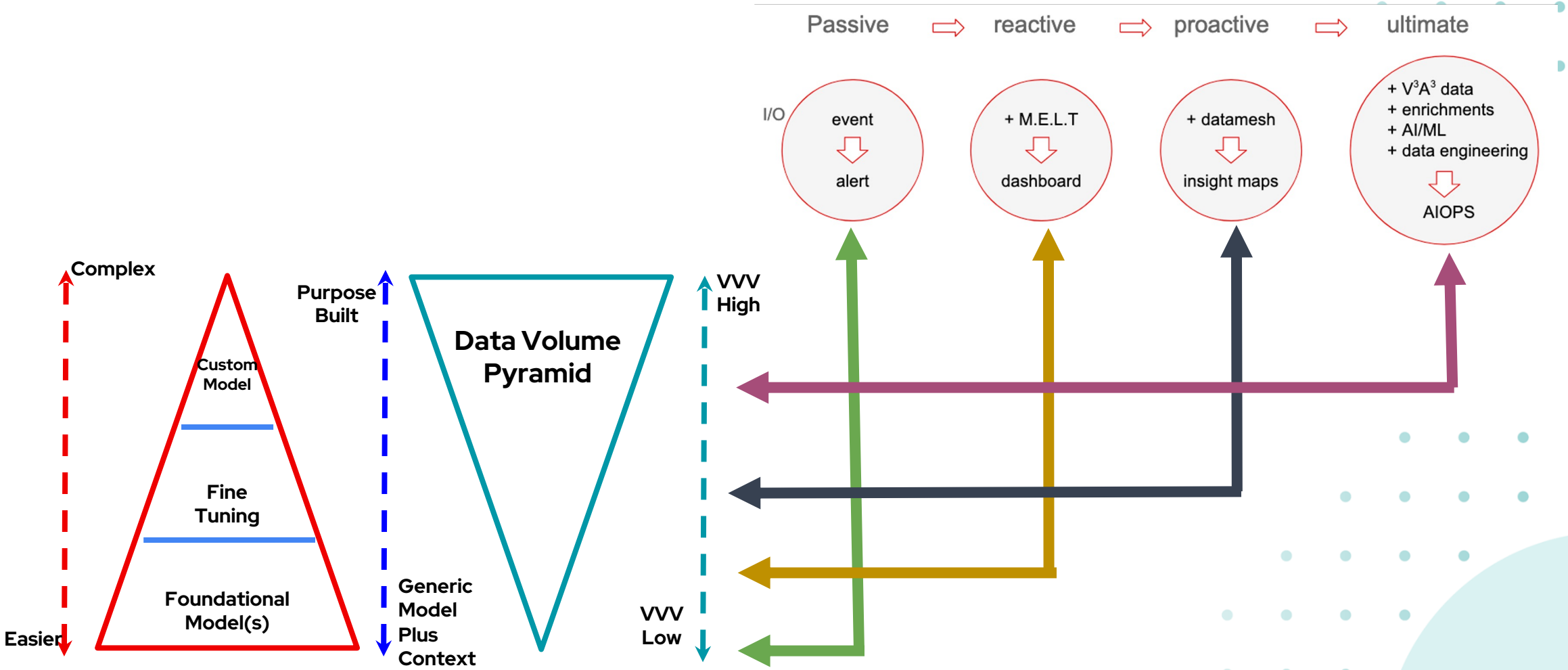
Fatih is also a recognized writer, sharing insights through his Open xG HyperCore series on Medium and contributing to AI/ML projects on GitHub and Hugging Face.

In 2025 Fatih has been elected as a subject matter expert on AI/ML within Linux Foundation Networking (LFN) organization to steer & lead AI initiatives.

LinkedIn

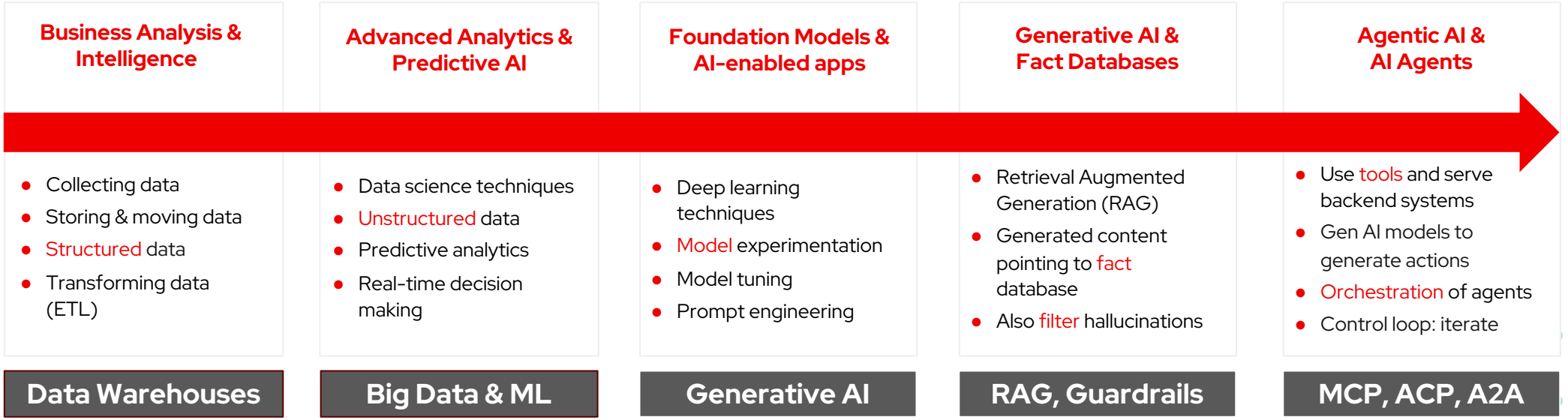
https://about.me/fenar

Better DATA -> Better Business



From Data to Agentic AI Ecosystems

From Business Intelligence to Agentic AI

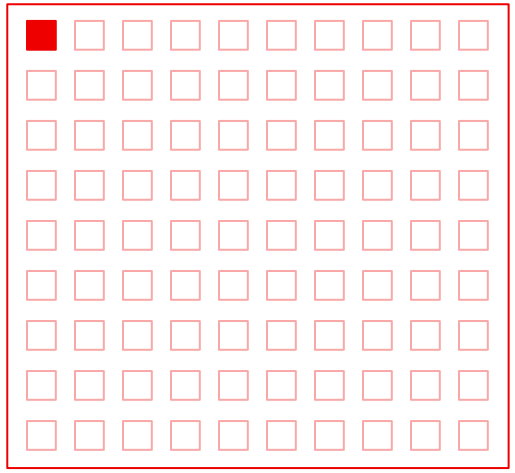


- **Predictive AI** runs businesses today
- Foundation models provide a shortcut for realizing the value of **Generative AI**
- Models are embedded in and **guarded** by applications
- Agents perform tasks

Unified Data Driven Application Framework

Enterprises need models aligned to their private data

LLMs are trained with a range of public data, not enterprise-relevant data



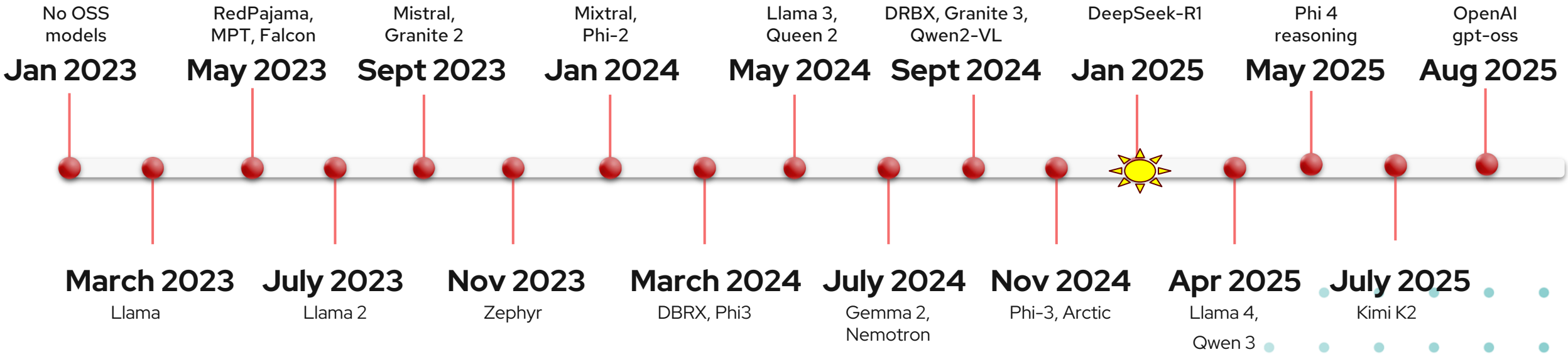
Less than 1% of all enterprise data
is represented in foundation models

Enterprise organizations need to

1. Start from a trusted base model
2. Create a new representation of their data
3. Deploy, scale, and create value with their AI

The power of open

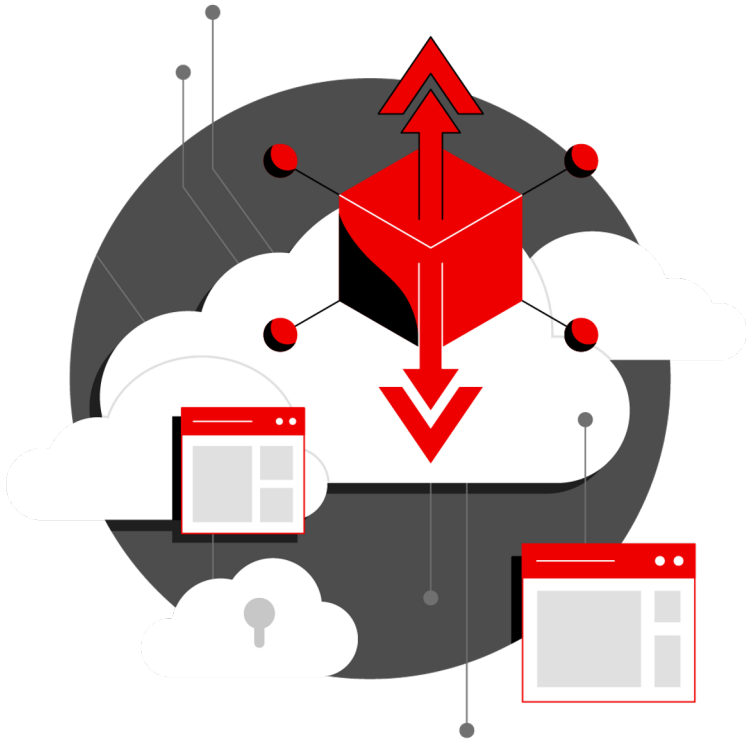
There has been an explosion of capability from open-source over the last 2 years



Meta · Ai2 · NVIDIA · MISTRAL AI_ · databricks · Hugging Face · snowflake · IBM · Google · Microsoft · deepseek

The value of open source and smaller language models

Smaller models are more efficient & customizable



- Open source AI models (Qwen, Kimi, DeepSeek, Llama, Mistral, ..) have caught up with the closed/proprietary models and surpassed for price/performance, accuracy, roi/tco.
- Smaller language models (SLMs) are orders of magnitude smaller than foundational models (<10 Billion parameters vs. >1 Trillion)
  - SLMs are cheaper and faster to run and consume less energy.
- Can be easily tuned and customized with private enterprise data for domain specific tasks.
- Customers own their own models and can create multiple instances for different use cases and deployment environments.

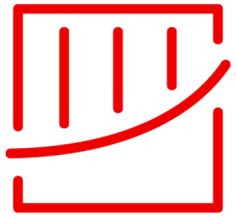
Inference is where the real value generated

Always-on intelligence

Inference powers every user interaction 24/7, making inference the constant cost and performance driver

Latency defines experience

Low-latency and fast inference ensures optimal user experiences, essential for real-time applications and user retention

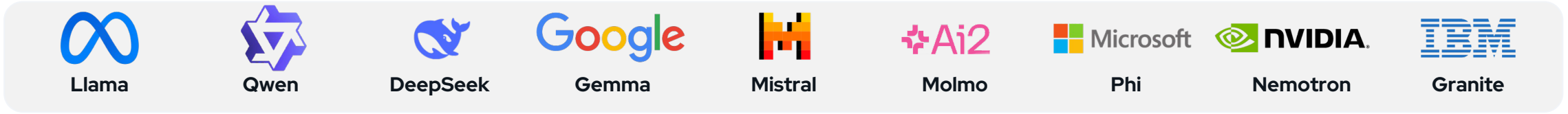


Exponential market momentum

The AI inference market is forecast to grow from USD 106.15 billion in 2025 to USD 254.98 billion by 2030, underscoring its role

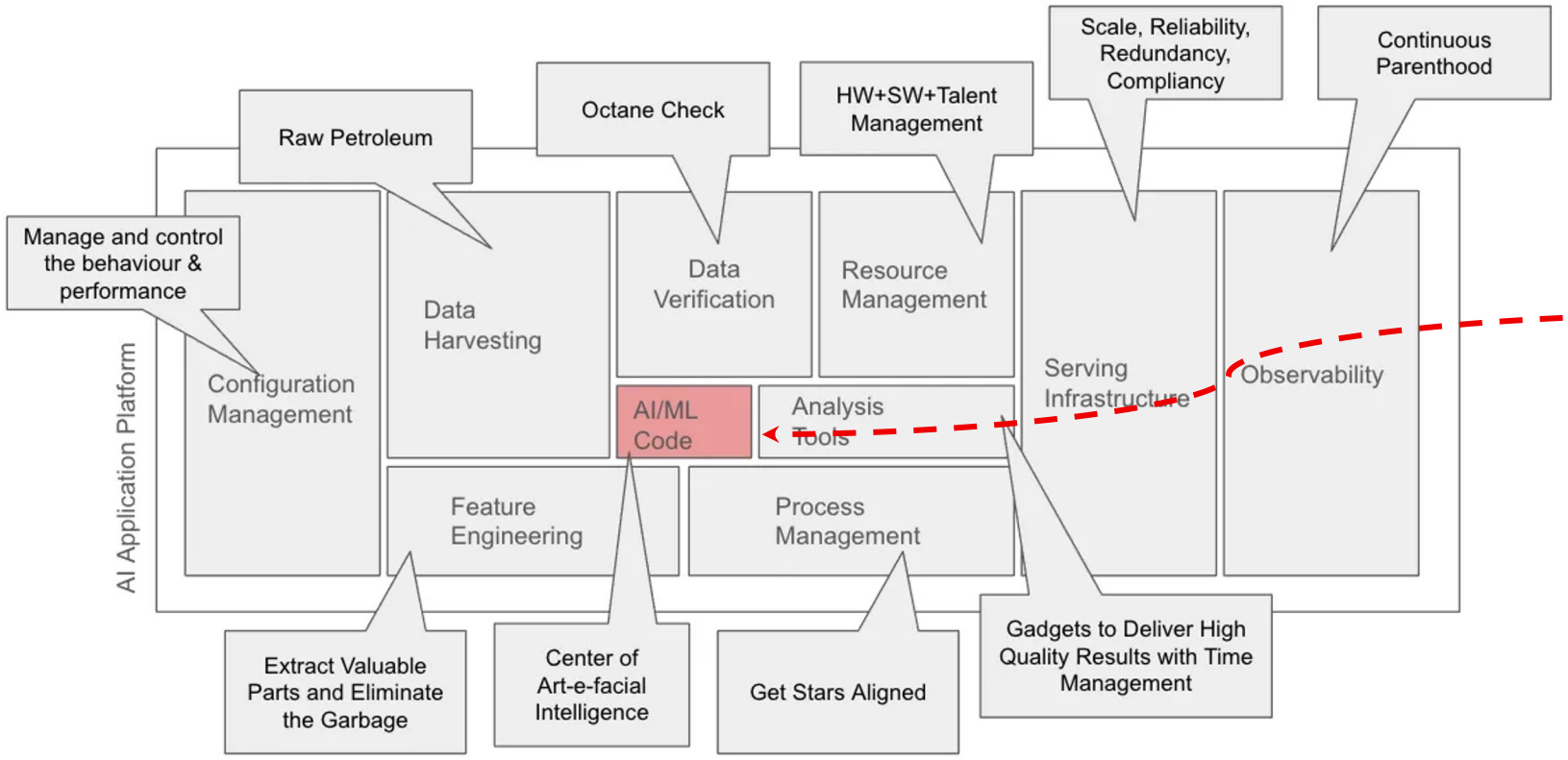
Optimized Inferencing with vLLM

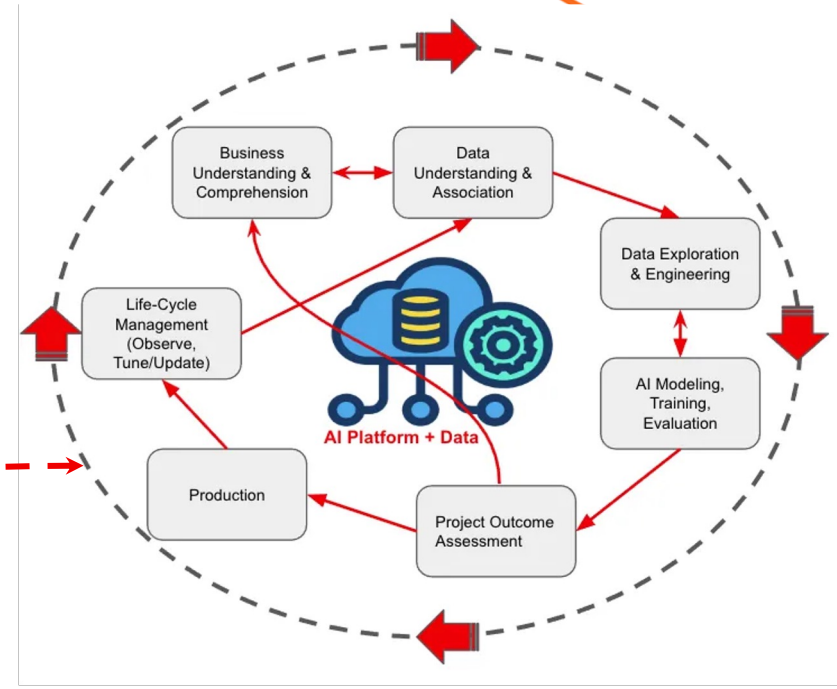
vLLM supports the key models on the key hardware accelerators



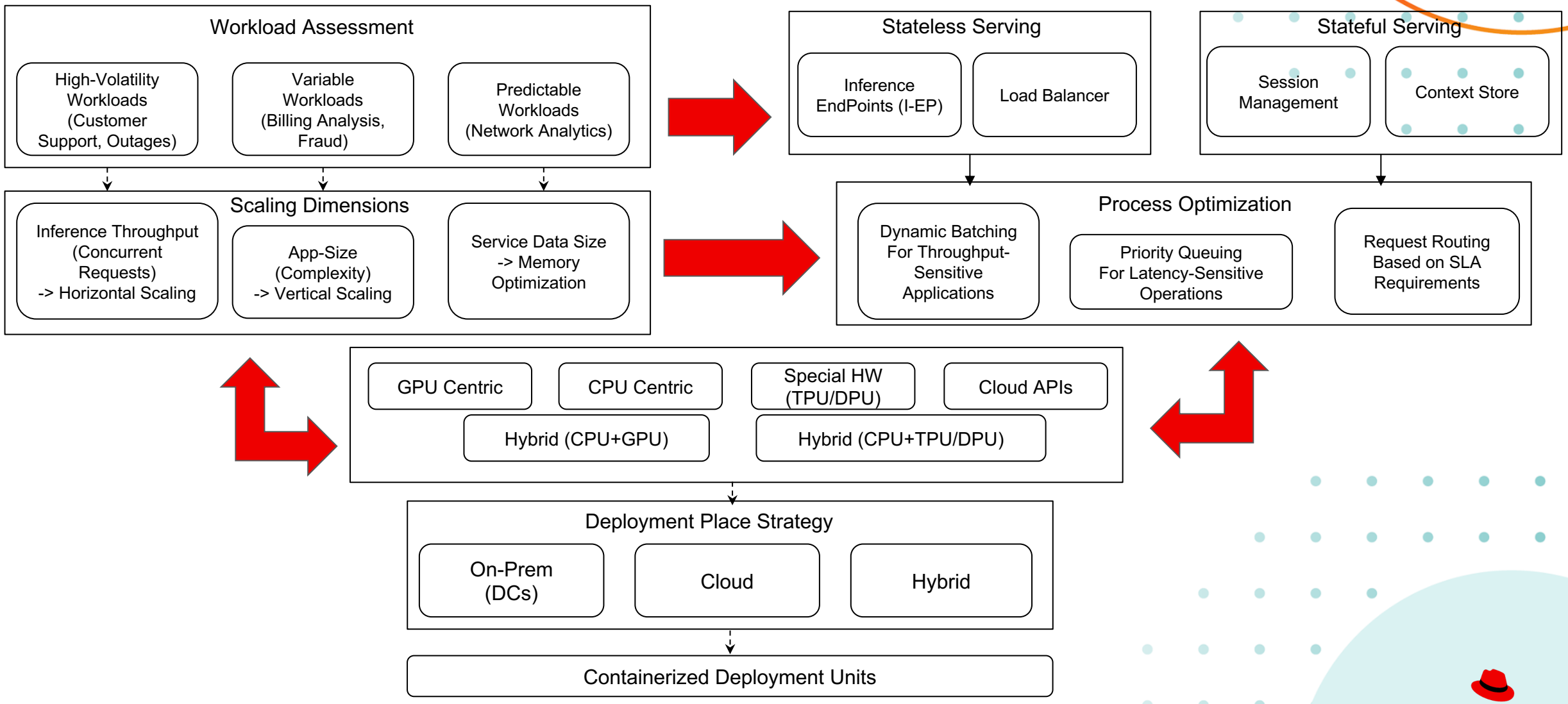
vLLM

AI -Inside-SW Stack



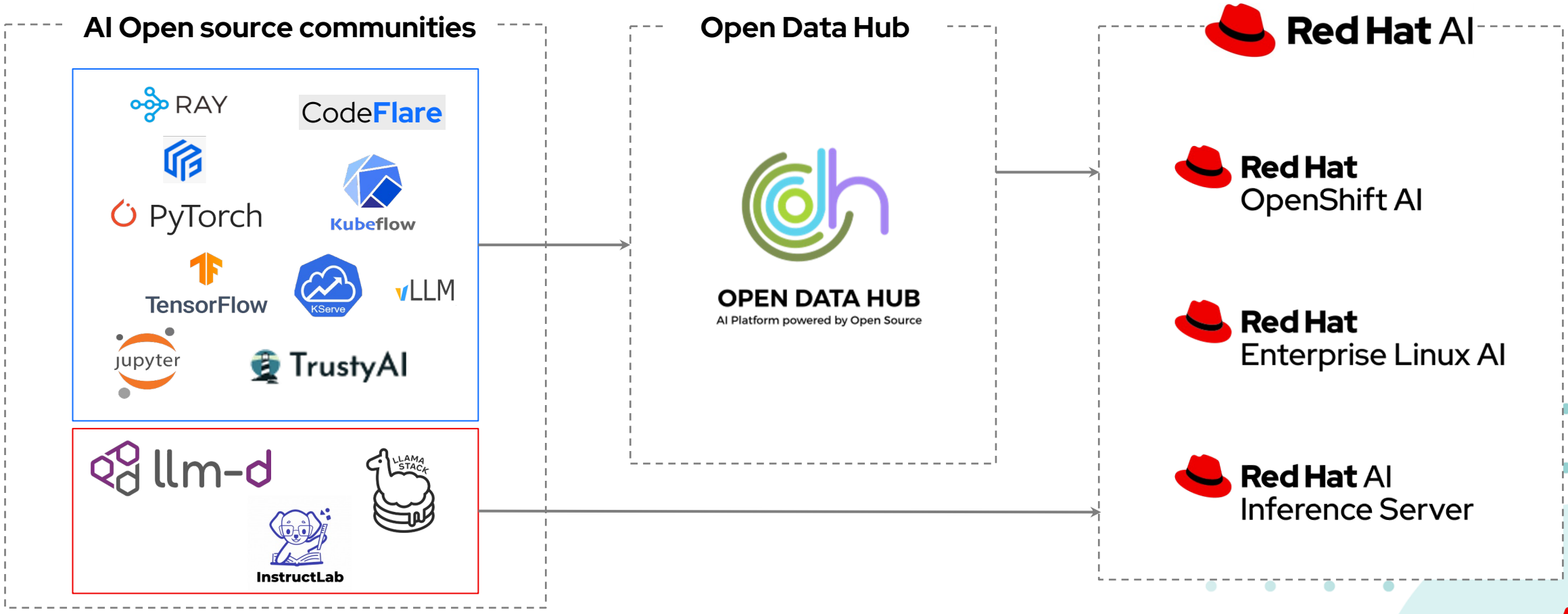


Workload Driven Architectures



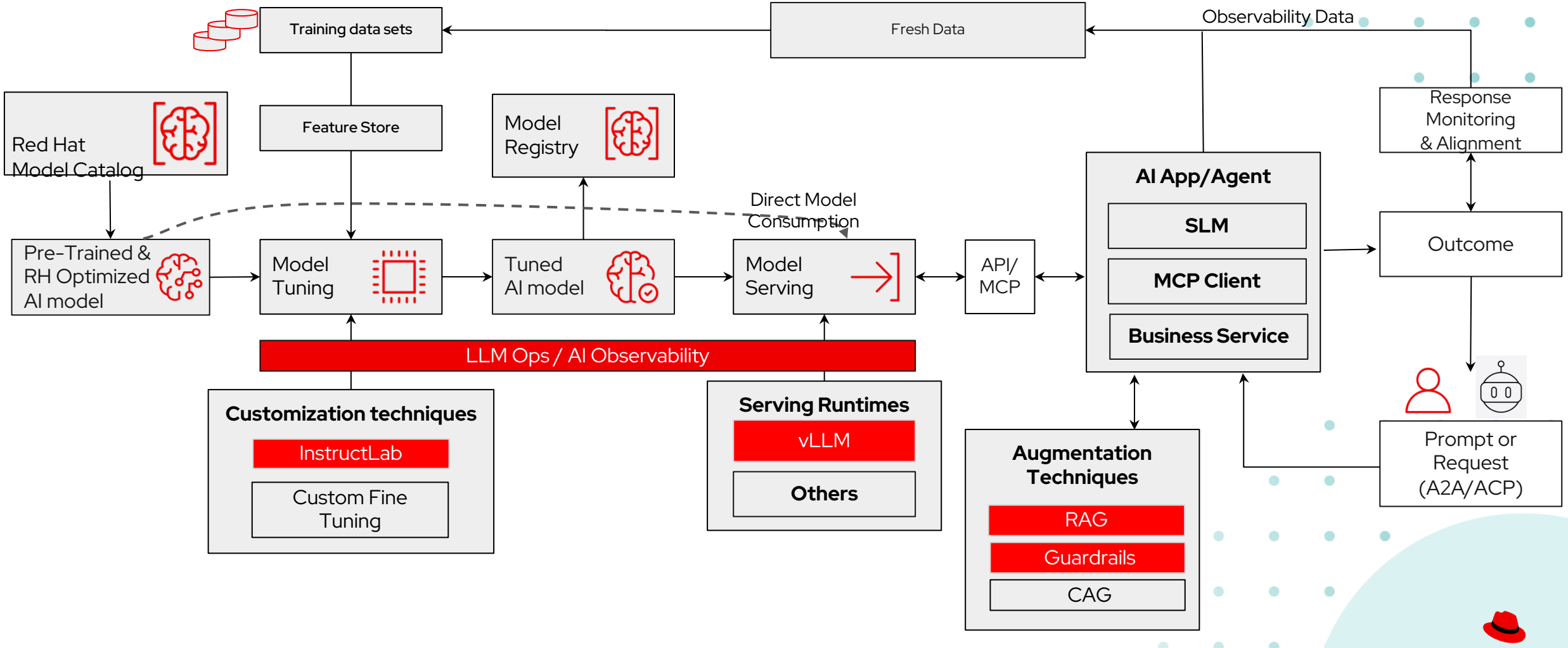
The Future of AI is Open

Red Hat's open source community engagement is a catalyst for powerful AI collaboration



Real World Problems > Real World Solutions

Solution BluePrint

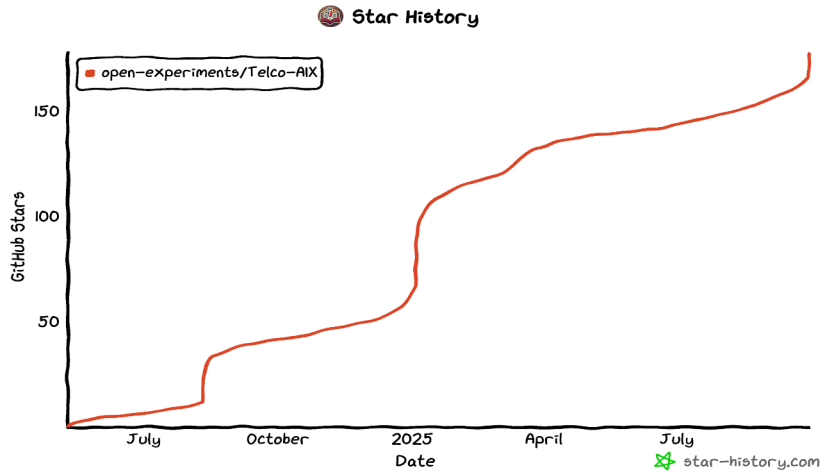


Telco-AIX

https://github.com/open-experiments/Telco-AIX

Team

Role	Name	LinkedIn	Region
Maintainer	Alessandro Arrichiello	Profile	EMEA
Maintainer	Ali Bokhari	Profile	North America
Maintainer	Atul Deshpande	Profile	APAC
Program Manager	Arun Thomas	Profile	Texas
Business Development	Paul Lancaster	Profile	North America
Business Development	Sankar Panneerselvam	Profile	Texas
Founder	Fatih E. NAR	Profile	Texas



Experiments

Domain	Project	Focus Area
💰 Revenue Management	RAFM	Revenue Assurance & Fraud Detection
📊 Service Quality	Service Assurance Churn Prediction	Latency & NPS Predictions & Churn Predictions
🌐 Network Operations	5G Network Ops	Fault Predictions
🌿 Sustainability	Energy Efficiency	Green Telecom Initiatives
🔒 Security	SecOps-AI	Networking Security
⚡ Smart Infrastructure	AI Powered SmartGrid	Grid Optimization
🛡️ IoT Security	IoT Perimeter Security	Perimeter Security
🚀 Advanced AI	5G CNF RCA with LLM	Root Cause Analysis
💬 Customer Experience	CRM Voice App Intent Classification	Intelligent Customer Interactions
🔍 Anomaly Detection	RootCause Analysis	Model Chaining & RAG
🛰️ Connectivity	Starlink QoE	Satellite ISP Experience
🖥️ Network Operations	NoC AI Augmentation	OSS Optimization
🎩 IT Management	ITSM Automation	Intelligent Service Management
🤖 Agentic Telco-AI	Agentic Framework Autonomous 5G Network	Agentic Telco
🥇 Telco Expert Portal	One Stop KnowHow Shop	Prompt Engineering & Embeddings Search

SCAN ME

Ent-ObX

(https://github.com/open-experiments/Ent-ObX)

1. Python Instrumentation

Python Instrumentation **Focus**: Jupyter Notebook-based AI Application Stack Observability

- Detailed metrics and insights
- Seamless integration with data science workflows

2. Java Instrumentation

Java Instrumentation **Focus**: Enterprise Java Application Monitoring

- Minimal code overhead
- OpenTelemetry framework integration
- Optimized for OpenShift deployments

3. C++ Instrumentation

C++ Instrumentation **Focus**: High-Performance Application Monitoring

- Native OpenTelemetry integration
- Low-overhead metrics collection
- UBI 8 based OpenShift deployment

4. eBPF BlackBox Harvesting

eBPF Instrumentation **Focus**: Black-Box Data Harvesting in Kernel [Work in Progress]

- Advanced system-level observability
- Kernel-level insights and telemetry

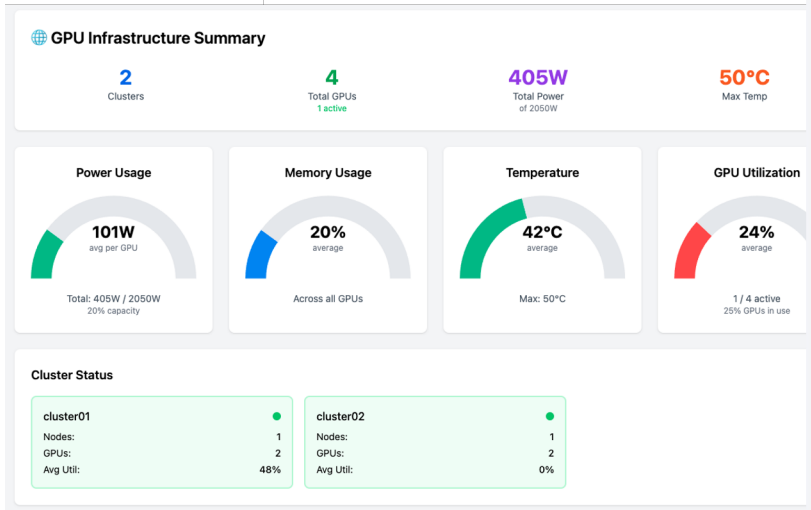
5. GPU-Console

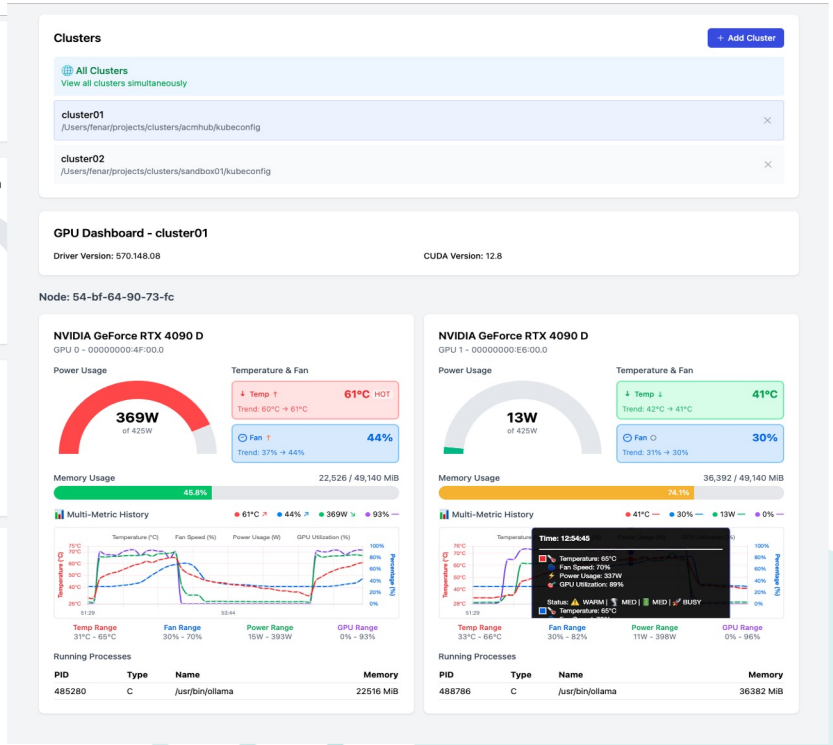
GPU Dashboard Single Pane of Glass **Focus**: Multi-Cluster GPU Resource Observability [Work in Progress]

- Single Pane of Glass for GPU Resources & Their Consumption Levels

Our mission is to provide ready to use observability examples through:

- Custom Data (Metrics, Events, Logs and Traces -MELT) harvesting & data-pipelines with OTel
- Advanced Data Harvesting Techniques with eBPF





SCAN ME

Open-Experiments: Get Involved

- **Explore the repository and contribute**
  - Dive into the code, data, and documentation, and become an active contributor
- **Collaborate with the community**
  - Engage with other users and developers to share knowledge and build better solutions
- **Build and share your own AI solutions**
  - Develop and contribute your own AI models and use cases to expand the repository's capabilities

Red Hat Summit

Connect

Thank you

 linkedin.com/company/red-hat

 facebook.com/redhatinc

 youtube.com/user/RedHatVideos

 twitter.com/RedHat