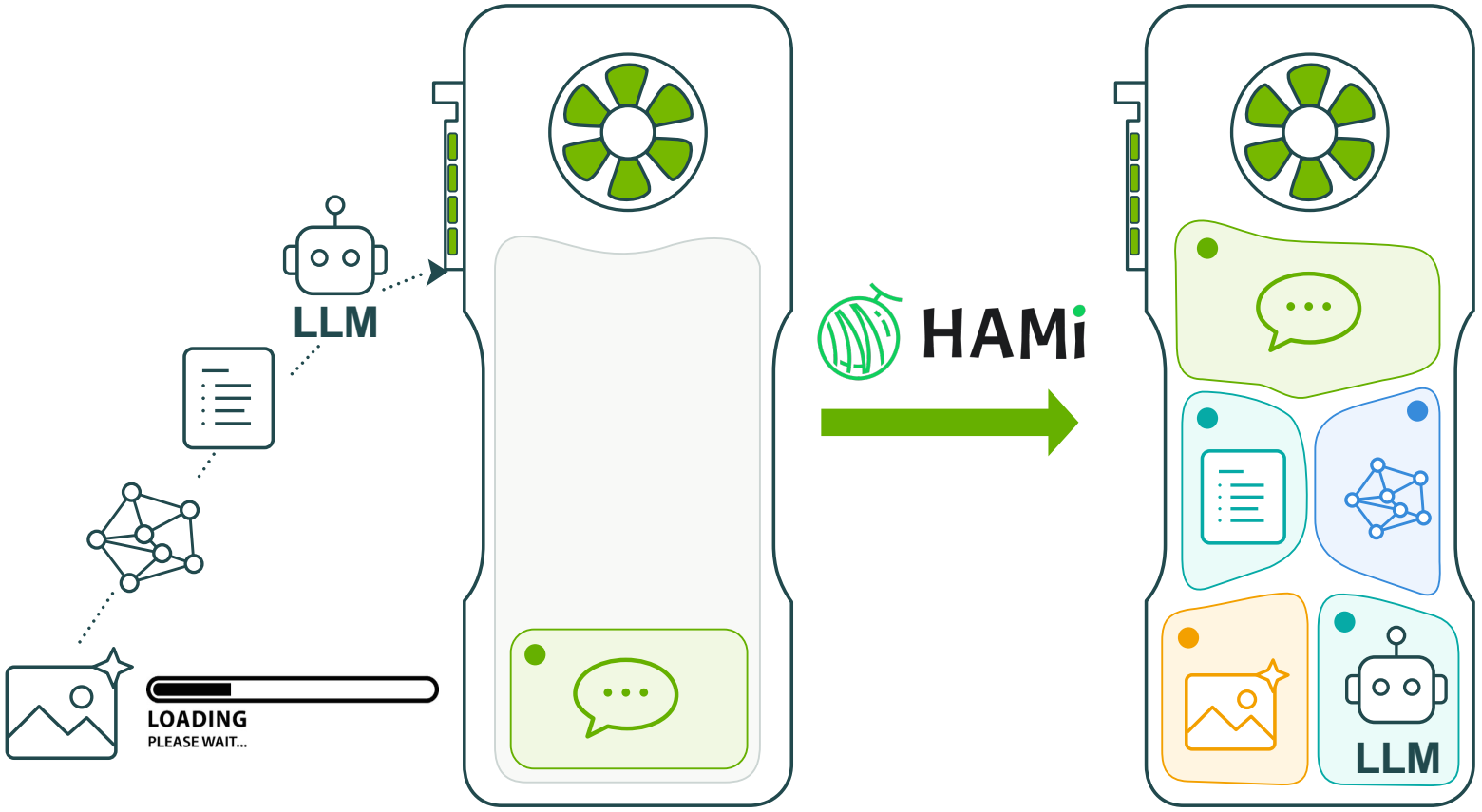
GPU sharing with HAMi



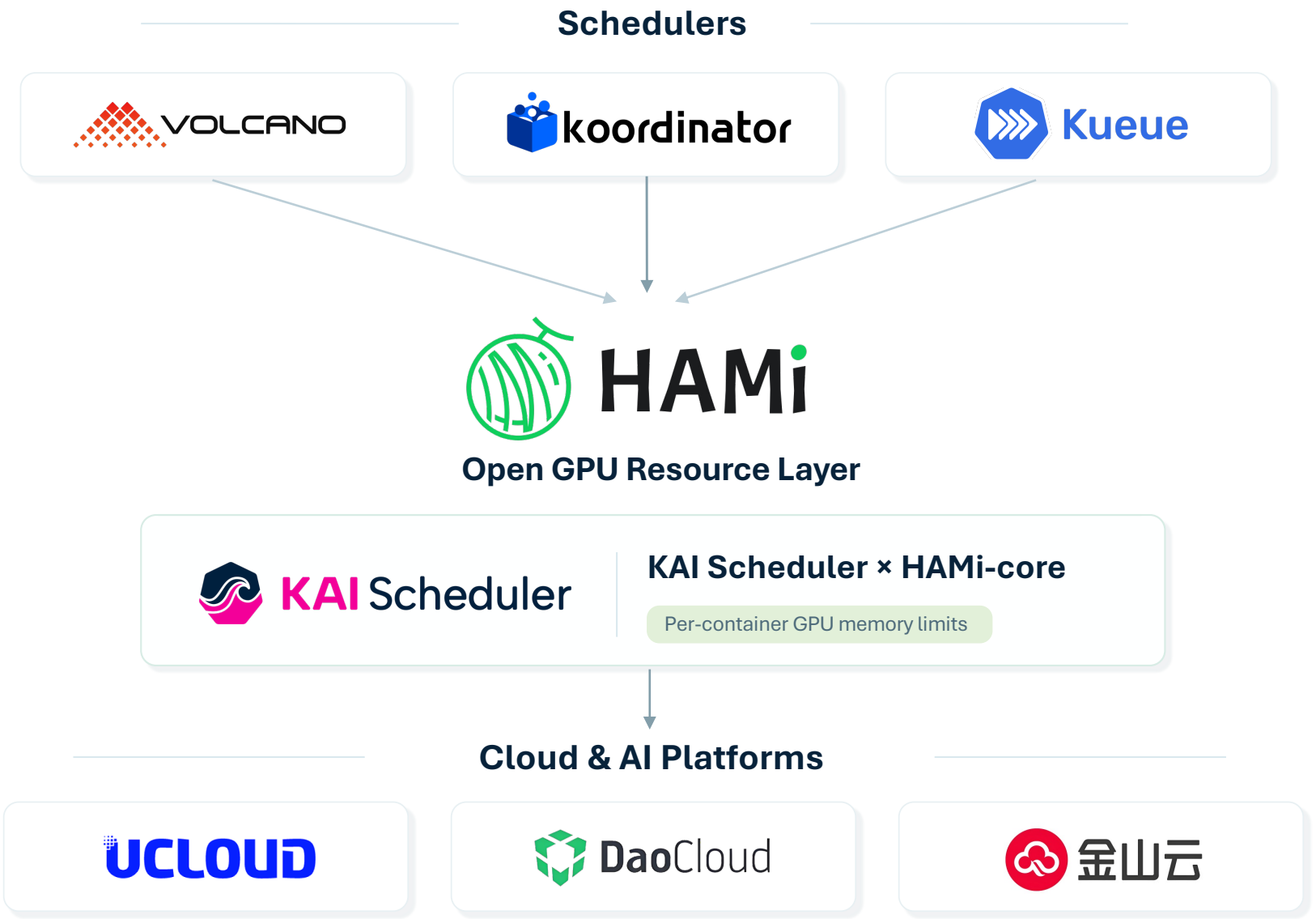
GPUs & accelerators

AMD
Ascend
aws
壁仞科技 BIREN TECHNOLOGY
Cambricon 寒武纪
Enflame 燧原科技
HYGON 中科海光
天数智芯 Iluvatar CoreX
昆仑芯 KUNLUNXIN
METAX 沐曦
摩尔线程 MOORE THREADS
NVIDIA
瀚博半导体 POWERED BY VASTSTREAM

Capabilities vary by device.

Xiao Zhang · CEO, Dynamia · HAMi Maintainer

HAMi in the cloud native ecosystem



723

Contributors

1075

Organizations

44

Countries & Regions

CLOUD NATIVE COMPUTING FOUNDATION

Incubating since July 2026

HAMi in production

招商銀行 CHINA MERCHANTS BANK

+46.7%

Single-card inference throughput

ResNet50 · batch 32

NIO

30%

fewer GPU hours

Driving simulation

SNOW

50%

fewer GPUs

Train-to-inference pipelines

PREP

90%

of GPU infrastructure

Managed by HAMi

Dynamia's engineering contributions

llm-d

MetaX integration

Prefill / decode compatibility

Under review

Lupine

vLLM compatibility

Per-client GPU metrics

Merged upstream

密瓜智能 DYNAMIA.AI

Founded by HAMi initiators
Maintains & drives
HAMi adoption

HAMi

CNCF Incubating Project

Later today · 8 Sep (UTC+8)

09:59–10:04	PD Disaggregation vLLM Deployment on Alternative AI Accelerators Using llm-d
11:14–11:19	Project Lightning Talk: From Static Slices to Elastic GPUs: Dynamic MIG with HAMI
14:30–15:00	How Intsig Serves Billions of Document Scans: GPU Virtualization at Scale with HAMi

Booth area: Grand Ballroom I, Project Pavilion, **T-1**

HAMi community

WeChat assistant