



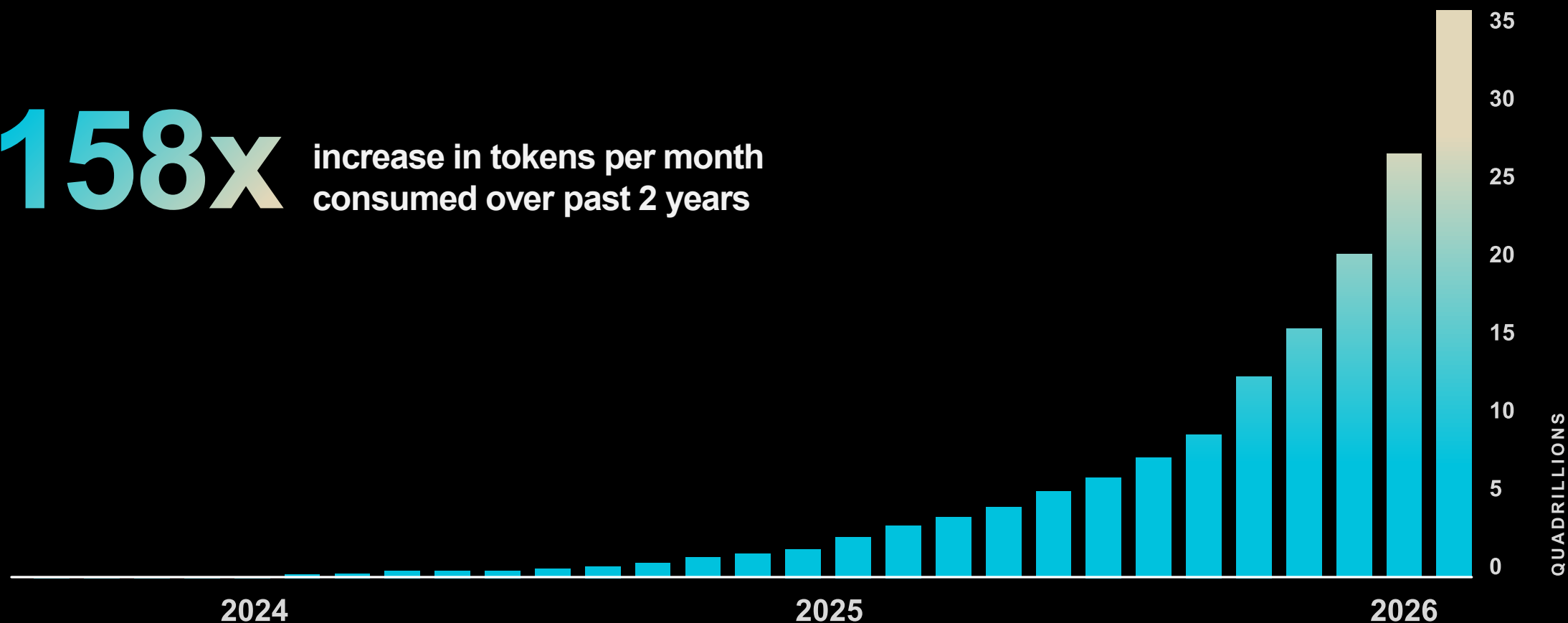
Building the Open Compute Ecosystem for the Agentic Era

Mark Papermaster

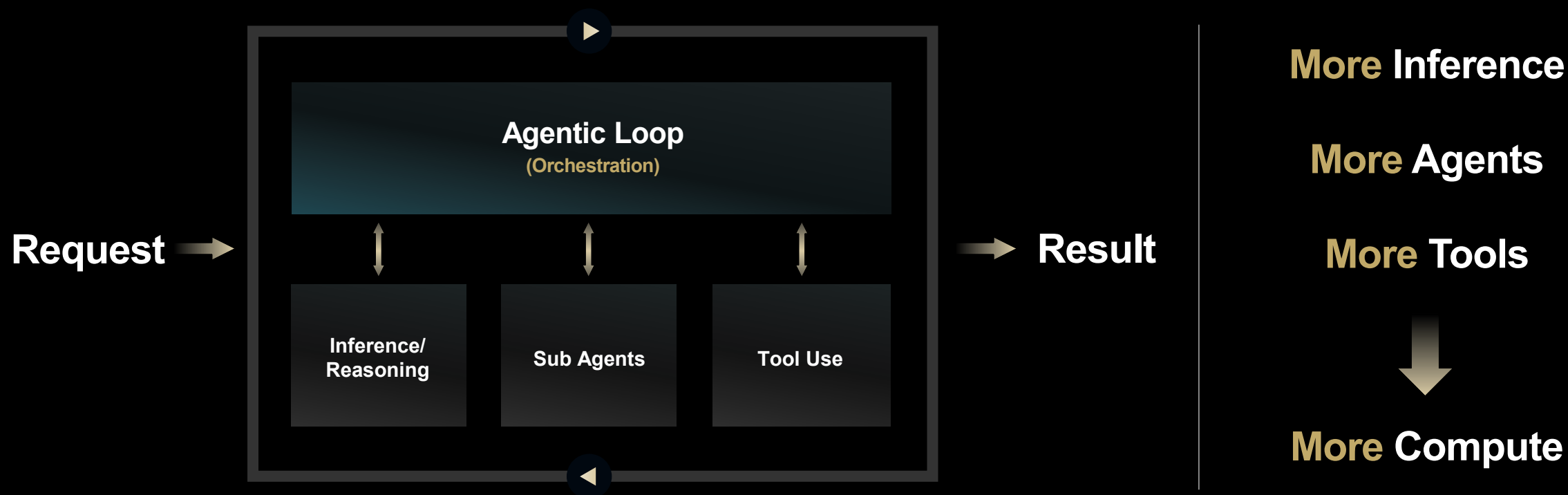
AMD EVP and CTO

Agentic AI Driving Demand Spike

158x increase in tokens per month consumed over past 2 years

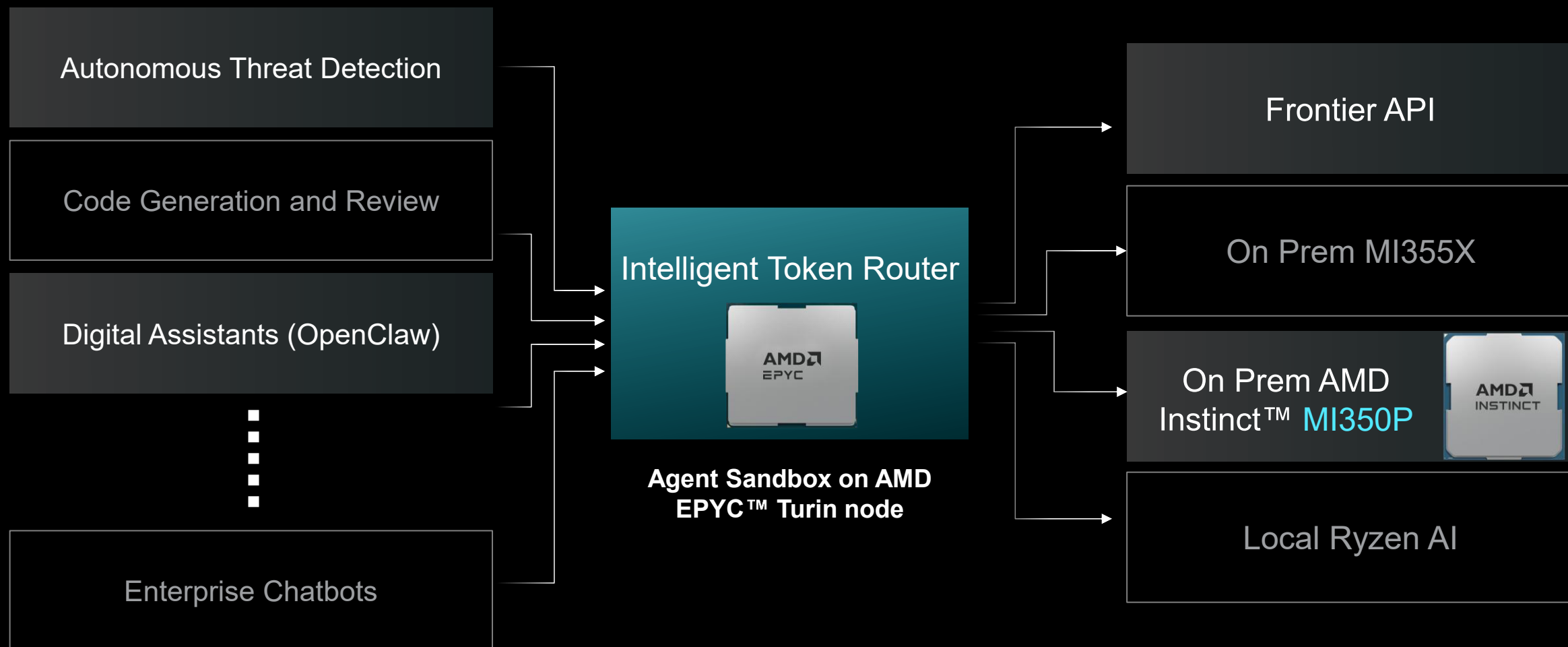


Agentic AI Further Accelerates Demand for Compute



AMD IT Use Case: Optimized Inference Costs

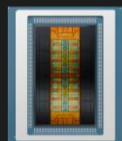
By intelligently routing to the right endpoint and model



Frontier Models Require Rack-Scale

Tightly integrated and fully optimized solution

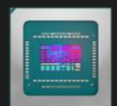
AMD  HELIOS



MI455X GPU
5th Generation Instinct™



"Venice" CPU
6th Generation EPYC™



Salina DPU
AMD Pensando™



Vulcano AI NIC
AMD Pensando™



Open Ecosystems Compound Innovation

The Pace of AI is Set By Shared Software



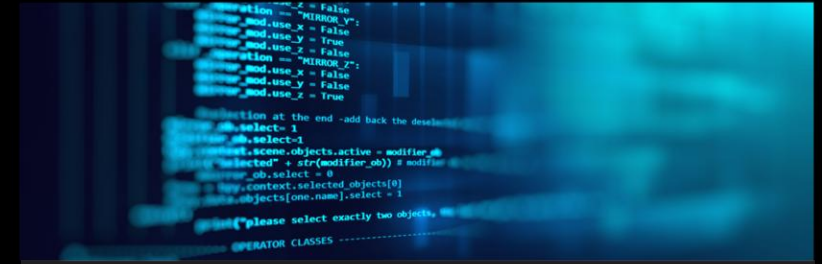
**Innovation lands in
the open first**

The community defines the stack
with silicon that meets it on day one



**Abstraction is the new
portability**

Developers want options suited to the
level of the problem without rewriting
for each generation of hardware



**AI is now writing the
accelerated code**

Agents are collapsing the cost of tuning and
porting, and they learn from what is public



AMD ROCm.AI

Easy, Accessible Development

ROCm.AI in Action

1. Establish Performance Baseline



2. Profile & Pinpoint Bottlenecks



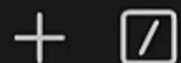
3. Tune Serving Config



4. Generate & Rewrite GPU Kernels



5. Validate Gains



</> Edit automatically

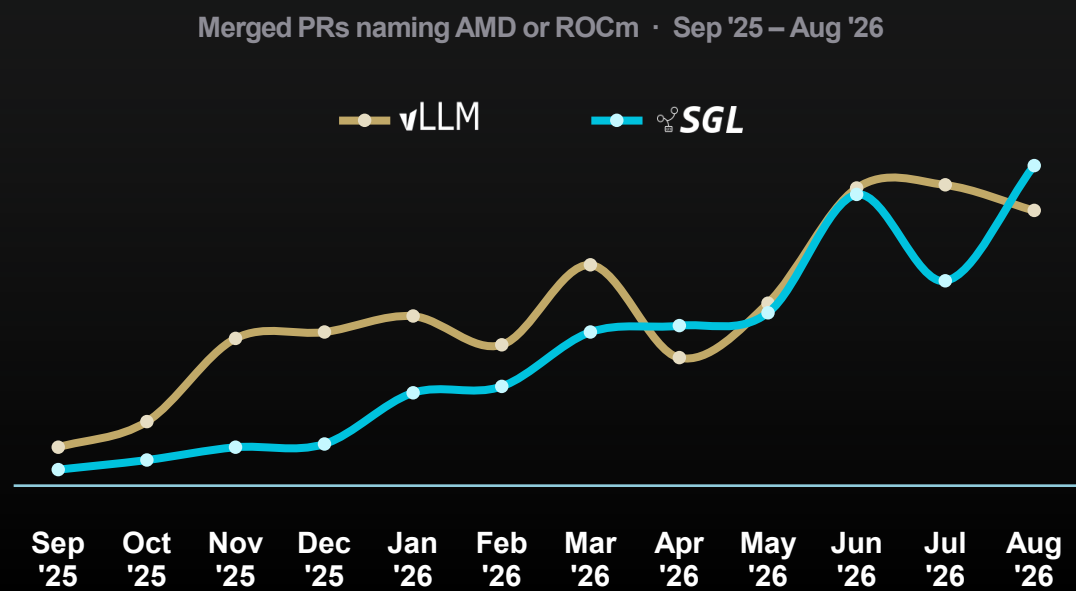


AMD Accelerates Ecosystem Innovation

Contributing to The Open Ecosystem



Growing Upstream Contribution



Enormous Opportunity In Front of Us

Design Holistically

Optimize each
component together

Move Faster

Development models
accelerate on pace with AI

Stay Open

Provides the best innovation
and sovereignty



Disclaimers

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions and typographical errors.

The information contained herein is subject to change and may be rendered inaccurate for many reasons, including but not limited to product and roadmap changes, component and motherboard version changes, new model and/or product releases, product differences between differing manufacturers, software changes, BIOS flashes, firmware upgrades, or the like. AMD assumes no obligation to update or otherwise correct or revise this information. However, AMD reserves the right to revise this information and to make changes from time to time to the content hereof without obligation of AMD to notify any person of such revisions or changes.

AMD MAKES NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE CONTENTS HEREOF AND ASSUMES NO RESPONSIBILITY FOR ANY INACCURACIES, ERRORS OR OMISSIONS THAT MAY APPEAR IN THIS INFORMATION.

AMD SPECIFICALLY DISCLAIMS ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE. IN NO EVENT WILL AMD BE LIABLE TO ANY PERSON FOR ANY DIRECT, INDIRECT, SPECIAL OR OTHER CONSEQUENTIAL DAMAGES ARISING FROM THE USE OF ANY INFORMATION CONTAINED HEREIN, EVEN IF AMD IS EXPRESSLY ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

© 2026 Advanced Micro Devices, Inc. All rights reserved. AMD, EPYC™, Instinct™, Infinity Fabric™, Pensando™ DPU accelerators, AMD ROCm™, Ryzen™, the AMD Arrow logo, and combinations thereof are trademarks of Advanced Micro Devices, Inc.

Ultra Accelerator Link™ and its abbreviation UALink™ are trademarks of the UALink Consortium.

ESUN and Open Compute Project® is a registered trademark of the Open Compute Project Foundation.

The Mistral Large language model is a trademark of Mistral AI.

PyTorch, the PyTorch logo and any related marks are trademarks of The Linux Foundation.

Ultra Accelerator Link™ and its abbreviation UALink™ are trademarks of the UALink Consortium.

ComfyUI, Comfy and the Comfy logo are trademarks of Comfy Org.

Hugging Face, the Hugging Face logo and any related marks are trademarks of Hugging Face, Inc.

JAX is a trademark of Google LLC.

Disclaimers Continued

llama.cpp is an open-source project maintained by ggml.ai; Llama is a trademark of Meta and its affiliate companies.

llm-d is a trademark of LF Projects, LLC.

Ollama is a trademark of Ollama Inc.

ONNX and the ONNX logo are trademarks of LF Projects, LLC.

TensorFlow, the TensorFlow logo and any related marks are trademarks of Google Inc.

Unsloth and the Unsloth logo are trademarks of Unsloth AI.

verl is a trademark of ByteDance Ltd.

vLLM is a registered trademark of LF Projects, LLC.

Other product names used in this publication are for identification purposes only and may be trademarks of their respective companies.