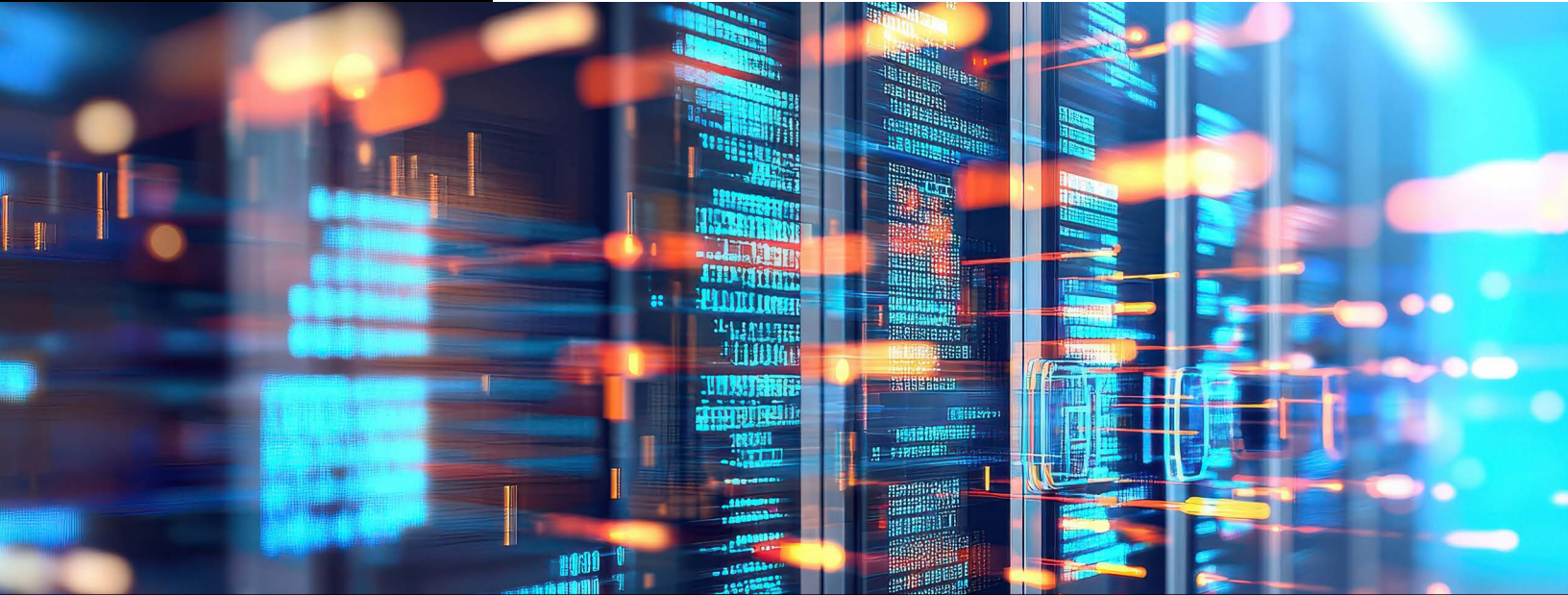




UNIPHORE BUSINESS AI CLOUD ON AMD ENTERPRISE AI REFERENCE STACK

SECURE, COMPOSABLE, SOVEREIGN AI AGENTS FOR THE ENTERPRISE.

EXECUTIVE SUMMARY

Every enterprise is running AI pilots. Very few are running AI in production at scale. The gap between experimentation and industrialization is where competitive advantage is won or lost.

Uniphore Business AI Cloud on AMD enterprise AI reference stack closes that gap. It is a complete, production-grade AI stack that takes enterprises from fragmented data to deployed AI agents, with sovereign governance, open infrastructure, and business outcomes measured in weeks, not years.

Uniphore delivers the enterprise AI platform: a composable system that connects enterprise data, governed knowledge, domain-tuned small language models, and agentic workflows across financial services, healthcare, telecommunications, and more. AMD delivers the infrastructure foundation: optimized inference containers, GPU-accelerated compute, and a Kubernetes-native deployment suite with zero licensing fees.

Together, they give enterprises a path from AI pilot to production-scale AI systems without the vendor lock-in, data migration, or governance gaps that stall most AI programs.

THE ENTERPRISE AI CHALLENGE

The promise of enterprise AI has never been greater. The reality has never been more frustrating. Organizations across every industry have invested in AI pilots, proofs of concept, and innovation labs. But most of those experiments never reach production. They stall in the gap between what AI can do in a sandbox and what it takes to run AI as a core business operation.

The friction points are consistent across industries:

Data fragmentation is the first and biggest barrier. Enterprise data is fragmented across systems, clouds, and formats. Most organizations spend most of their AI project timelines on data preparation alone: wrangling, cleaning, migrating, and reconciling data before a single model is trained or agent is deployed. Without solving the data challenge, nothing else moves.

Governance and trust are unresolved. Most AI toolchains cannot meet the auditability, explainability, and compliance requirements that regulated enterprises demand. When AI decisions affect patients, policyholders, or account holders, the inability to explain and govern those decisions is a deployment blocker.

Scaling exposes hidden costs. Workloads that perform well in a pilot environment encounter latency, throughput, and infrastructure cost problems when pushed toward production volumes. What seemed affordable at ten queries per second becomes untenable at ten thousand.

Vendor lock-in limits agility. Closed AI stacks create dependencies at every layer, from models to infrastructure. As AI technology evolves rapidly, enterprises locked into a single vendor's ecosystem face rising costs and shrinking flexibility.

Regulatory and sovereignty constraints block deployment. In financial services, healthcare, telecommunications, and other regulated industries, data residency requirements, privacy regulations, and security mandates make it impossible to deploy AI solutions that do not offer sovereign control over data and models.

The enterprises that industrialize AI first will define the competitive landscape for the next decade. The rest will spend those years trying to catch up. The question facing every executive team is no longer whether to deploy AI. It is *whether* their infrastructure, their data, and their governance model can move AI from experiment to operating system.

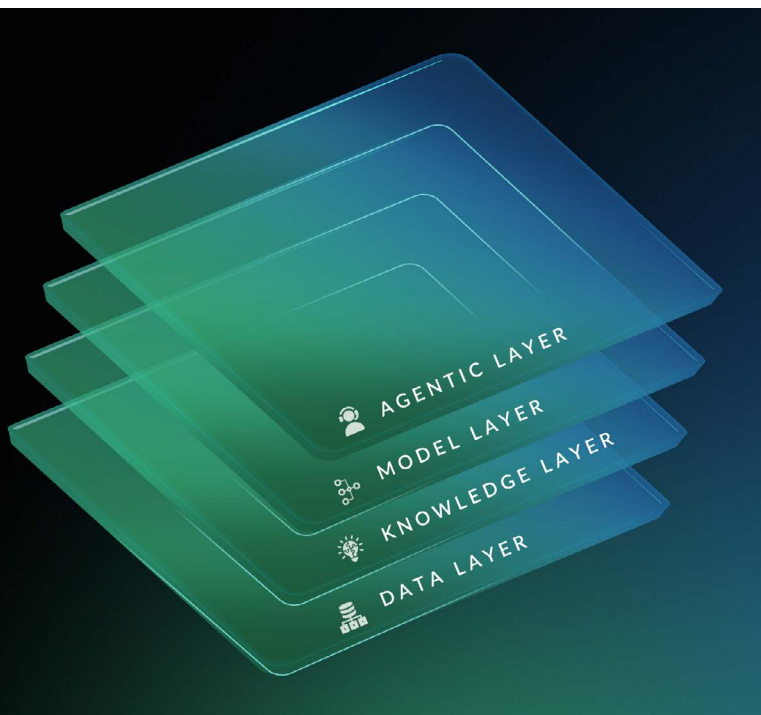


THE SOLUTION: UNIPHORE BUSINESS AI CLOUD ON AMD ENTERPRISE AI REFERENCE STACK

Uniphore and AMD have assembled a complete, production-grade AI stack that eliminates the barriers described above. From enterprise data to deployed AI agents, this joint solution is business-ready on day one, sovereign by design, and built on an open, cost-efficient infrastructure.

Uniphore Business AI Cloud: The Enterprise AI Platform

Uniphore Business AI Cloud is a composable, sovereign, secure enterprise AI platform that unifies enterprise data, knowledge, models, and agents into production-ready workflows. Its modular architecture spans four layers, each addressing a core enterprise AI challenge. Organizations can adopt each layer independently or deploy them as a unified platform.



Data Layer: Solving the Data Preparation Bottleneck

Data preparation is the single largest time and cost sink in enterprise AI programs. The Uniphore zero-copy data fabric eliminates it. Rather than migrating data into a centralized repository, the Data Layer connects to enterprise data where it already resides, across systems, clouds, and formats, and prepares it in place.

Purpose-built data agents automate the pipeline from raw enterprise data to AI-ready inputs. The Data Discovery Agent identifies and maps relationships across data sources, generating metadata and enrichment. The Data Engineering Agent builds, automates, and orchestrates data pipelines for both structured and unstructured data. The Data Analyst Agent provides a self-service natural language interface for rapid access to insights across multi-modal data.

The result: **10x faster data preparation, zero data copy required, and 50-70% faster speed to insights.** For most enterprises, this alone transforms the economics and timeline of AI deployment.

Knowledge Layer: Domain-Specific, Enterprise-Tuned Intelligence

The Knowledge Layer transforms unstructured enterprise data into structured, governed knowledge optimized for AI consumption. It enables enterprise-grade retrieval, reasoning, explainability, and SLM fine-tuning, ensuring that AI agents operate on trusted, auditable information rather than opaque model outputs.

Model Layer: Open, Governed, and Secure Model Orchestration

The Model Layer supports fine-tuning and orchestration of both proprietary and open-source models, including domain-specific small language models (SLMs) built for production workloads. Enterprises can adapt AI behavior to their specific workflows, policies, and regulatory requirements, and swap models as technology evolves, eliminating vendor lock-in.

Agentic Layer: AI That Works Across Real Business Processes

The Agentic Layer delivers deterministic, governed AI agents that operate across real business workflows and enables enterprises to build, deploy, and manage agents that operate across real workflows: billing operations, customer onboarding, compliance processing, network management, and more. Agents are orchestrated via BPMN, integrating into existing business process frameworks rather than requiring enterprises to rebuild around AI.

AMD Enterprise AI Reference Stack: The Infrastructure Foundation

AMD enterprise AI reference stack is the engine that makes the Uniphore platform run at enterprise scale. It provides the optimized compute, simplified deployment, and operational governance that turn AI applications into production systems.

AMD Inference Microservices (AIMs) are pre-built, optimized containers that bundle the model, inference engine, optimized drivers and libraries, and all necessary configurations into a single deployable unit. AIMs automatically detect available GPU resources and apply hardware-aware tuning, selecting optimal precision and parallelism settings for AMD Instinct GPUs. Deployment is a single command. Supported inference engines include vLLM and SGLang, with OpenAI-compatible APIs for fast integration with existing applications.

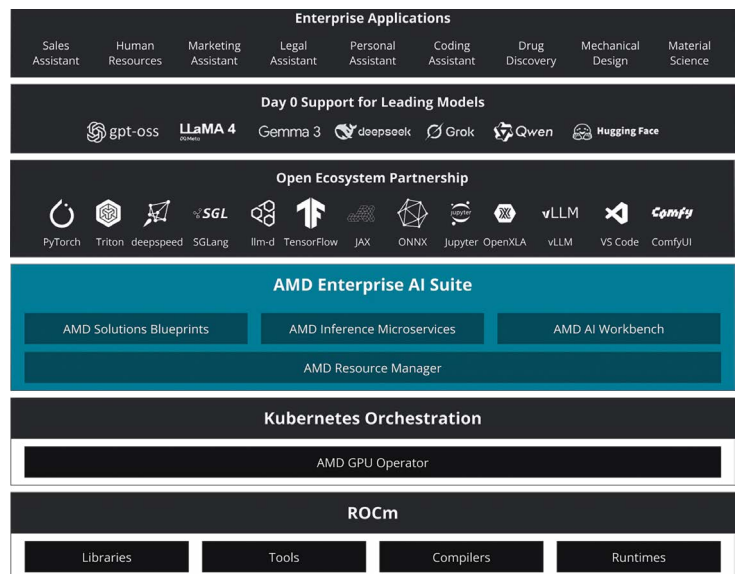
AMD AI Workbench provides a unified developer interface for deploying AIMs, accelerating model fine-tuning with expert-curated hyperparameter recipes, evaluating models side-by-side with Chat & Compare, and provisioning GPU-as-a-Service with built-in governance and quotas.

AMD Solution Blueprints are ready-to-deploy reference architectures that integrate AIMs with data pipelines, retrieval systems, and orchestration frameworks. They provide a fast path from prototype to production for generative and agentic

AI deployments.

AMD Resource Manager is a centralized control plane for multi-tenant GPU infrastructure. It provides cluster management, workload scheduling, quotas, and monitoring, giving IT administrators the governance and visibility they need to operate AI infrastructure at scale.

The full stack is Kubernetes-native, open-source, and carries zero licensing fees.



WHY UNIPHORE + AMD

Uniphore delivers the business outcomes. AMD makes those outcomes faster, simpler to deploy, and scalable across the enterprise. Together, they create something neither could deliver alone.

Uniphore brings production-ready AI workflows that drive measurable outcomes across regulated industries. Its composable, sovereign platform provides governance, guardrails, and auditability at every layer. Its zero-copy data fabric and purpose-built data agents eliminate the data preparation bottleneck that stalls most enterprise AI programs. And its domain-tuned SLMs deliver the performance and accuracy that production workloads require.

AMD brings an open infrastructure that removes cost and complexity barriers. AMD Instinct GPUs with high-bandwidth memory enable larger models per GPU and better economics. AIMs containers bundle optimized drivers and libraries for single-command deployment. The open-source ecosystem, including vLLM and SGLang, eliminates infrastructure-layer lock-in. And Kubernetes-native orchestration with centralized resource management makes enterprise GPU operations manageable.

The combined value

Ease of deployment. Pre-built AI agents on pre-optimized inference containers, deployed on Kubernetes-native infrastructure. Production in weeks, not months.

Better economics at scale. Zero-license-free AMD enterprise AI reference stack plus high-performance GPUs enable lower total cost as workloads scale.

Security and sovereignty. Data, models, and operations remain under enterprise governance across cloud, hybrid, and on-premises deployment models. No data leaves the enterprise's control.

Business outcomes from day one. Pre-built agents delivering measurable impact across financial services, healthcare, and telecommunications. Not in future roadmaps, but in current production deployments.

No vendor lock-in at any layer. Open models, open infrastructure, open APIs. Enterprises retain the flexibility to evolve their AI strategy as the technology landscape changes.

BUSINESS OUTCOMES: PRODUCTION RESULTS ACROSS INDUSTRIES*

Data readiness is the foundation for every outcome that follows. With 10x faster data prep and zero data copy required, the Uniphore platform delivers AI-ready data as the starting point, not the finish line.



Financial Services: Customer Onboarding and Compliance

THE CHALLENGE

Customer onboarding in financial services is a fragmented, high-friction process. Identity verification, KYC checks, and regulatory disclosures span multiple disconnected systems. The result is high application abandonment, slow time-to-revenue, and compliance risk from inconsistent monitoring. Manual processes cannot keep pace with the volume of regulatory requirements or the expectations of digital-first customers.

THE SOLUTION

Uniphore's agentic AI automates the end-to-end onboarding workflow. The zero-copy data fabric unifies semantic and multi-modal data into AI-ready knowledge without migration. Domain-tuned models handle document classification, identity verification, and disclosure processing. Integrated governance strengthens compliance at every step. AMD-accelerated inference ensures real-time processing at the volume financial institutions require.

30-40%
reduction in onboarding time

80-90%
compliance adherence improvement

20-30%
conversion lift

50-70%
faster processing cycle times

20-30%
AHT reduction

5-15%
first-call resolution improvement



HEALTHCARE PAYER: PRIOR AUTHORIZATION AND CLAIMS AUTOMATION

THE CHALLENGE

Healthcare payers face an escalating operational burden. Prior authorization decisions require extensive manual review, creating slow cycle times and rising administrative costs. Fraud, waste, and abuse case volumes are growing, with high false positive rates that consume investigator time without proportional results. On the member-facing side, inconsistent guidance and high call volumes strain service operations.

THE SOLUTION

AI agents automate document processing and decision workflows for prior authorization, reducing manual touchpoints while maintaining compliance. A multi-agent framework handles claims adjudication, fraud detection, and case prioritization, surfacing the highest-value cases for human review. Agentic assistance guides frontline staff with compliant, real-time information retrieval. The zero-copy data layer connects to existing claims and member systems without data migration, preserving sovereignty and accelerating deployment.



TELECOMMUNICATIONS: BILLING, NETWORK OPERATIONS, AND FIELD SERVICE

THE CHALLENGE

At telecom scale, small inefficiencies become major cost drivers. Billing complexity overwhelms contact centers and slows resolution, while network downtime and congestion increase maintenance and degrade quality of service. Field teams face scheduling inefficiencies and uneven access to region-specific training, driving repeat visits, high travel costs, and missed SLAs.

THE SOLUTION

Uniphore on AMD addresses telecommunications challenges across three interconnected domains:

Billing: A billing-specific SLM and dedicated AI agent handle customer billing inquiries through a self-service chat portal. The solution deploys entirely within the customer's own cloud environment, preserving data sovereignty. The SLM is purpose-trained on billing data, call transcripts, and standard operating procedures.

Network operations: Agentic AI analyzes telemetry from hybrid infrastructure spanning cloud and on-premises systems to forecast equipment failures, coordinate preventive maintenance, and dynamically reallocate network resources to match demand peaks across markets.

Field operations: AI agents assign field jobs based on engineer skills, proximity, and real-time traffic conditions. Field engineers access SLMs trained on region-specific technical manuals, historical repair data, and compliance standards, available offline for remote locations.

Additional Uniphore solutions span Insurance, Healthcare Provider, and Retail with documented impact metrics. Contact your Uniphore or AMD representative for details on these and other industry applications.

REDUCED

billing-related contact center volume

LOWER

unplanned downtime and maintenance costs

INCREASED

first-time fix rate and improved SLA compliance

CONTRACT LEAKAGE DETECTION & RECOVERY

Challenge

Contract leakage and overbilling often stem from fragmented contracts, clauses, invoices, and operational data, making it difficult to verify spend and detect non-compliance early.

Solution

An agentic workflow works by

- Connecting and enriching contract documents, clauses, and invoices
- Establishing a contract intelligence knowledge base and fine-tuning a domain-specific SLM
- Deploying AI agents to calculate leakage, detect non-compliance, and flag invoice discrepancies

Outcomes

- Reduced leakage and overpayment
- Higher fund recovery rates
- Earlier detection of non-compliance and discrepancies

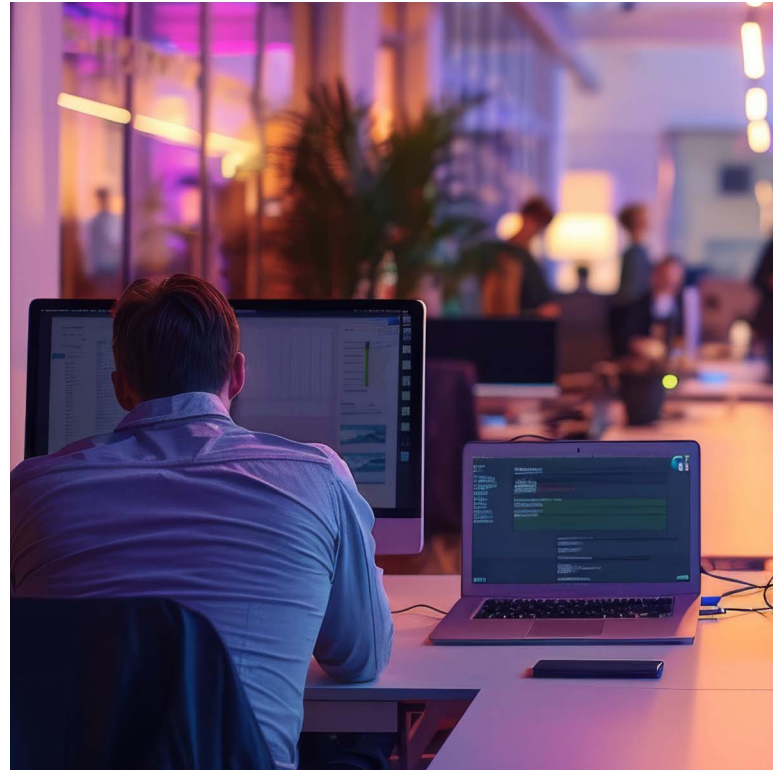
DEPLOYMENT MODEL

The Uniphore + AMD solution supports flexible deployment aligned with enterprise IT strategies and regulatory requirements. Organizations can deploy in the cloud, in hybrid environments, or on-premises without sacrificing performance or governance.

Hybrid: Kubernetes-native architecture enables workload splitting across cloud and on-premises environments, with unified governance and monitoring through AMD Resource Manager.

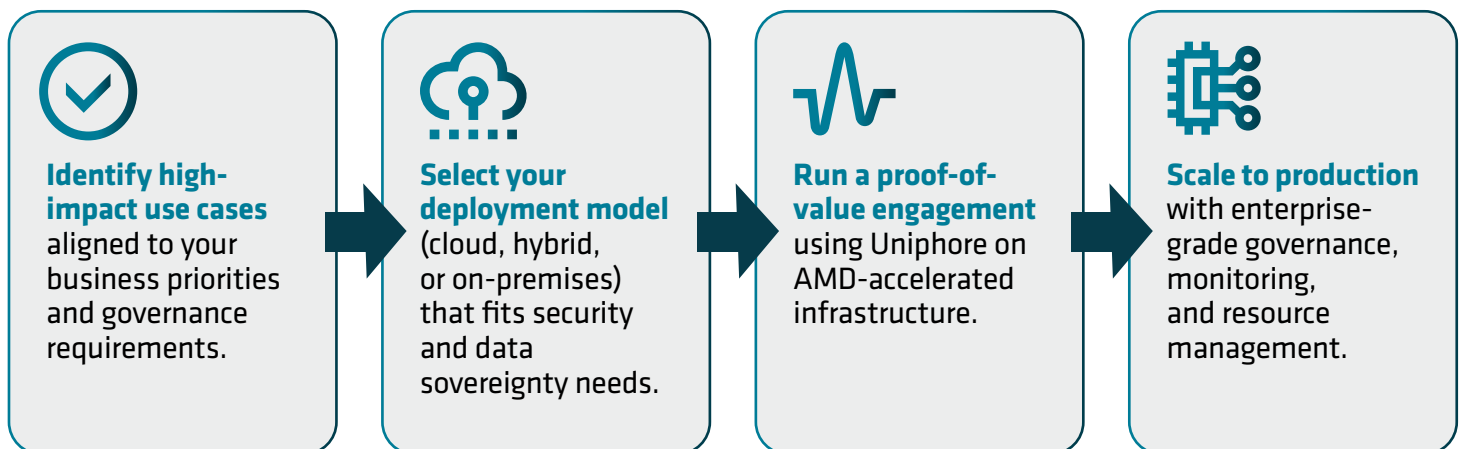
On-premises: On-premises supported.

AMD enterprise AI reference stack is designed for all three deployment models from the ground up. AIMs containers deploy identically across environments, ensuring consistent performance and simplified operations regardless of where the workload runs.



NEXT STEPS

Ready to move from AI experimentation to production outcomes? Engage with Uniphore and AMD to:





CONTACT YOUR AMD OR UNIPHORE REPRESENTATIVE FOR A STRATEGY SESSION.

ABOUT AMD

More than 50 years driving innovation in high-performance computing. AMD Instinct GPUs power AI workloads for leading cloud providers, enterprises, and sovereign AI initiatives worldwide. AMD's commitment to open standards and open ecosystem ensures enterprises can deploy AI with control, flexibility, and long-term vendor independence.

ABOUT UNIPHORE

Uniphore is the Business AI Company that unlocks the agentic enterprise with a complete, composable AI platform spanning agents, models, knowledge, and data. Its platform, the Business AI Cloud, bridges the AI divide between consumer AI and enterprise AI—combining the simplicity of consumer AI with the rigor, security and scalability required for the enterprise. Uniphore allows business users to effortlessly harness AI and deliver results immediately, while providing CIOs the foundation to deliver powerful AI applications that are embedded into workflows, trained on enterprise data.

Trusted by more than 2,000 businesses globally, recognized by Gartner, Forrester, and listed on the Deloitte Fast 500, Uniphore delivers on the promise of AI as a transformative force for business. Learn more at www.uniphore.com.

*All performance and/or cost savings claims are provided by the 3rd party organization featured herein and have not been independently verified by AMD. Performance and cost benefits are impacted by a variety of variables. Results herein are specific to such 3rd party organization and may not be typical. GD-181a.

©2026 Advanced Micro Devices, Inc. all rights reserved. AMD, the AMD arrow, and combinations thereof, are trademarks of Advanced Micro Devices, Inc. Other product names used in this publication are for identification purposes only and may be trademarks of their respective companies. Certain AMD technologies may require third-party enablement or activation. Supported features may vary by operating system. Please confirm with the system manufacturer for specific features. No technology or product can be completely secure.

PID: 2645550