

# 5 GRÜNDE, WARUM AMD EPYC™ 9005 PROZESSOREN UNTERNEHMENS-KI-INFERENZ VORANBRINGEN

## AUF EINEN BLICK

Die unersättlich hohe Nachfrage nach KI ist zu einem entscheidenden Faktor für Unternehmen geworden, die vom Mehrwert dieser Technologie profitieren möchten. Im heutigen schnelllebigen Geschäftsumfeld besteht die Gefahr, dass Unternehmen, die KI-Innovationen nicht umsetzen, hinter der Konkurrenz zurückbleiben.

AMD EPYC 9005 Prozessoren beschleunigen und erleichtern Ihre KI-Transformation, indem sie eine Plattform für die Rechenzentrumskonsolidierung, einen Hostprozessor für GPU-beschleunigtes maschinelles Lernen und einen effizienten Prozessor für KI-Inferenz bereitstellen.

1

## TRADITIONELLE UND UNTERNEHMENS-KI-AUSLASTUNGEN KONSOLIDIEREN

AMD EPYC ist eine leistungsstarke Lösung für Organisationen, die das Potenzial der KI ausschöpfen und gleichzeitig ihre bestehenden Geschäftsanwendungen vorantreiben möchten. Dank ihrer robusten Architektur, ihrer erweiterten Funktionen und ihres wachsenden Softwareökosystems ermöglicht diese Lösung Innovationsschaffung und das Erreichen von Geschäftszielen in der Ära der KI.

2

## FÜHRENDES PORTFOLIO FÜR DIE BEWÄLTIGUNG VON KI-HERAUSFORDERUNGEN IN UNTERNEHMEN

AMD EPYC Prozessoren und AMD Instinct™ Beschleuniger bieten ein komplettes Lösungsportfolio für alle Ihre KI-Anforderungen. AMD EPYC Konsolidierungsvorteile und Inferenz machen Unternehmen fit für die KI und können für kleinere KI-Inferenzbereitstellungen verwendet werden. AMD Instinct bietet führende generative KI-Performance zur Unterstützung größerer KI-Modelle wie LLMs.

3

## FÜHRENDE PERFORMANCE FÜR KI-CPU-INFERENZ IN UNTERNEHMEN

AMD EPYC Prozessoren liefern unglaubliche Performance für Enterprise-KI-CPU-Inferenz. Zudem optimiert die AMD ZenDNN Bibliothek die neuronalen Netzwerkbausteine, die eine verbesserte Deep-Learning-Inferenz-Performance auf AMD CPUs ermöglichen. Diese APIs unterstützen gängige KI-Frameworks wie TensorFlow, PyTorch und ONNX Runtime für Bildklassifizierungs- und Empfehlung-Engines.

4

## ENERGIEEFFIZIENTE LÖSUNGEN ERSTELLEN

AMD EPYC Prozessoren sind für energieeffiziente Server geeignet, bieten eine außergewöhnliche Performance und helfen, die Energiekosten zu senken. Entdecken Sie neue Möglichkeiten, um die Kernnutzung zu optimieren, Ihre Gesamtbetriebskosten zu beeinflussen und Ihre Ziele in Bezug auf Nachhaltigkeit voranzubringen.

5

## NUTZEN SIE DAS VERTRAUEN IN ERWEITERTE SICHERHEITSFUNKTIONEN UND OFFENE STANDARDS

Verlassen Sie sich auf zuverlässiges Computing, da Sie wissen, dass Ihr Unternehmen die Sicherheitsherausforderungen von heute mit den erweiterten Sicherheitsfunktionen von AMD Infinity Guard<sup>1</sup> bewältigen kann. Darüber hinaus ist das anhaltende und konsequente Engagement von AMD für die Unterstützung offener Standards entscheidend für die Entwicklung eines robusten und wettbewerbsfähigen Computing-Ökosystems.

# TECHNISCHE EINBLICKE

## 1. TRADITIONELLE UND UNTERNEHMENS-KI-AUSLASTUNGEN KONSOLIDIEREN

- Die wichtigsten Funktionen ermöglichen es AMD EPYC, die Infrastruktur zu konsolidieren, Kosten zu optimieren und sich den wachsenden Anforderungen von traditionellen und KI-Auslastungen anzupassen.
  - **Hohe Anzahl von CPU-Kernen und Multithreading:** Ermöglicht effiziente Verarbeitung rechenintensiver Auslastungen und parallele Verarbeitung.
  - **Große Caches:** Ermöglicht hohe Performance durch schnellen Zugriff auf häufig verwendete Daten.
  - **Hardwarebeschleunigung:** Lagert rechenintensive KI-Aufgaben aus, um die Performance zu verbessern.
  - **Erweiterte Speicherverwaltung:** Liefert hohe Bandbreite und niedrige Latenz für anspruchsvolle Datensätze und KI-Kommunikation.
  - **Stabiles Softwareökosystem:** Unterstützt führende KI-Frameworks und Tools für optimierte Performance.

## 2. FÜHRENDES PORTFOLIO FÜR DIE BEWÄLTIGUNG VON KI-HERAUSFORDERUNGEN IN UNTERNEHMEN

- AMD bietet beeindruckende Performance und Effizienz für CPU und GPUs. Zusammen mit AMD können Sie den besten Ansatz für 2P-Server mit AMD EPYC 9965 Prozessoren der 5. Generation liefern die folgenden Ergebnisse im Vergleich zu 2P-Servern mit AMD EPYC 9654 Prozessoren der 4. Generation:



- Gemischte Auslastungsinferenz
- Kleine bis mittlere Modelle
- Batch-/kleine Inferenz



- Dedizierte KI-Bereitstellungen, Training
- Mittlere bis große Modelle
- Große Inferenz

- Erzielen Sie herausragende durchgängige KI-Durchsatz-Performance bei einer großen Vielfalt von Anwendungsfällen. Unter Verwendung der TPCx-AI SF30 Benchmark liefern 2P-Server mit AMD EPYC 9965 Prozessoren mit 192 Kernen insgesamt bis zu 3,8-mal mehr KI-Testfälle pro Minute im Vergleich zu 2P-Servern mit Intel Xeon Platinum 8592+ mit 64 Kernen. [9xx5-012](#)

## 3. FÜHRENDE PERFORMANCE FÜR KI-CPU-INFERENZ IN UNTERNEHMEN

- AMD EPYC 9005 Prozessoren bieten im Generationenvergleich unglaubliche Verbesserungen bei Performance für KI-Inferenz. 2P-Server mit AMD EPYC 9965 Prozessoren der 5. Generation liefern die folgenden Ergebnisse im Vergleich zu 2P-Servern mit AMD EPYC 9654 Prozessoren der 4. Generation:

~39 % schnellerer LLM-Inferenzdurchsatz [9xx5-009](#)



Llama 3.1-8B BF16 (Token/Sek.)

~86 % schnellere Ähnlichkeitssuche [9xx5-011](#)



FAISS (Anfragen/Stunde)  
(Instanzen mit 16 Kernen bei FP32)

~101 % schnelleres maschinelles Lernen [9xx5-010](#)



XGBoost (Lauf/Stunde)

## 4. ENERGIEEFFIZIENTE LÖSUNGEN ERSTELLEN

- AMD bietet unternehmensgerechte IT- und KI-Lösungen an, die Energieeffizienz, technologische Innovation und niedrige Gesamtbetriebskosten bieten.
- AMD EPYC Prozessoren sind für energieeffiziente Server geeignet, bieten eine außergewöhnliche Performance und helfen, die Energiekosten zu senken.

## 5. NUTZEN SIE DAS VERTRAUEN IN ERWEITERTE SICHERHEITSFUNKTIONEN UND OFFENE STANDARDS

- Der AMD „Security by Design“-Ansatz umfasst moderne Sicherheitsfunktionen und ein Silizium-integriertes Sicherheitssystem. Ausgangspunkt ist eine Grundlage für die Verwaltung von Daten und Cyberbedrohungen über die Funktionen von AMD Infinity Guard<sup>1</sup>, die konsequent den AMD EPYC Prozessoren über die langjährige Sicherheits-Roadmap hinweg hinzugefügt wurden.
- Mit den AMD EPYC Prozessoren der 5. Generation wird Trusted-IO in den Funktionsumfang von Infinity Guard integriert. Damit wird die Integrität der Datensicherheit auf externe vertrauenswürdige Geräte erweitert. Sie werden derzeit nach FIPS 140-3 zertifiziert, was die neueste Benchmark für die Validierung kryptografischer Hardware ermöglicht.
- Da wir uns seit Langem konsequent für offene Standards einsetzen, können Unternehmen mit AMD EPYC Prozessoren sich auf beispiellose Flexibilität und Auswahl, mühelose Skalierbarkeit und herausragende Kosteneffizienz verlassen.

## AMD EPYC™ PROZESSOREN DER 5. GENERATION BRINGEN DIE RECHENZENTRUMS-KI VORAN

TOGETHER WE ADVANCE AI



MEHR ERFAHREN UNTER [AMD.COM/AI](https://www.amd.com/ai)

<sup>1</sup> Die Funktionen von AMD Infinity Guard variieren je nach EPYC™ Prozessorgeneration und/oder Serie. Sicherheitsfunktionen von Infinity Guard müssen von Server-OEMs und/oder Cloud-Diensteanbietern vor Betrieb aktiviert werden. Wenden Sie sich an Ihren OEM oder Anbieter, um die Unterstützung dieser Funktionen zu bestätigen. Mehr erfahren über Infinity Guard unter <https://www.amd.com/de/technologies/infinity-guard>. GD-183A