

Willkommen in einer neuen Ära:

## Edge-KI ist der echte Game Changer



Mit dem vermehrten Einsatz intelligenter Systeme vor Ort in Fabriken, im Einzelhandel und an anderen entfernten Standorten verlagert sich der Fokus von KI-Anwendungen: Es ist nicht mehr die schiere Rechenleistung entscheidend, sondern vielmehr die Frage, wo Daten generiert und verarbeitet werden müssen. Dazu gehört auch die Fähigkeit, schnelle Entscheidungen treffen zu können und KI-gestützte Lösungen in großem Umfang effizient zu managen. Schon heute streben 86 % der Unternehmen mit aktiven Edge AI-Implementierungen nach agentenbasierten Edge-Fähigkeiten. Die Bedeutung von Edge AI für die Business-Strategie von Unternehmen wurde erst 2026 in einer vom Branchenexperten Zededa in Auftrag gegebenen Umfrage unter CLevel und IT-Führungskräften in den USA und Deutschland untersucht.

### Herausforderungen im laufenden Betrieb treiben Edge-KI an

Unternehmen erkennen zunehmend, dass ihre wertvollsten Daten außerhalb zentralisierter Systeme entstehen. Z. B. in den Produktionshallen selbst, in Filialen oder an geografisch verteilten Standorten. Diese Entwicklung führt zu neuen Prioritäten für KI-Anwendungen: KI-Inferenz vor Ort in Echtzeit, geringerer Stromverbrauch und Kosteneffizienz im großen Maßstab sowie sichere Bereitstellung auf beispielweise Zehntausenden von End-Geräten sind heute konkrete Anforderungen. Gleichzeitig verändern die Möglichkeiten von Edge-KI Geschäftsmodelle, Prozesse und IT-Infrastrukturen grundlegend. Mitarbeitern vor Ort stehen ganz neue Optionen zur Verfügung: Etwa durch Computer Vision, spezielle LLMs oder mit anderen lokal bereit-

gestellten KI-Applikationen erhalten sie neue Fähigkeiten. Und so können sie Fehler schneller beheben, qualitativ hochwertigere Ergebnisse erzielen und autonom Entscheidungen treffen. Diese neuen Möglichkeiten zwingen Unternehmen dazu, nicht nur ihren Tech-Stack zu überdenken. Es geht vor allem auch darum, wie sie den Betrieb, die Daten und Infrastruktur in ihrer dezentralen Struktur in der Edge managen.

### Die Edge-Inferenz ist da

Schon bis Ende 2026 wird sich der eigentliche Wettbewerb im Bereich KI hin zur Edge-Inferenz verlagern. Rechenzentren bleiben zwar wichtig, doch immer mehr Unternehmen realisieren, dass die relevantesten Daten in Fabriken, in ihren Einzelhandelsfilialen oder an entfernt verteilten Standorten generiert werden - d. h. vor allem dort, wo Echtzeitentscheidungen entscheidend sind. Und bei großflächigen Edge-Implementierungen ergeben sich zusätzlich andere Einschränkungen. Wenn Unternehmen beispielsweise 20.000 Geräte irgendwo dezentral verteilt einsetzen, werden die Kosten pro Einheit und der Stromverbrauch zu entscheidenden Faktoren. Diese Gegebenheiten wiederum verbreitern das Ökosystem von Hardware-Anbietern und deren Portfolios. NVIDIA, Qualcomm, AMD und andere spezialisierte Chiphersteller sind bestens positioniert, um im Bereich Edge-Inferenz aggressiv zu konkurrieren und einen Markt zu gestalten, der weniger von Spitzenleistung als vielmehr von Effizienz, Flexibilität und Skalierbarkeit geprägt ist. Unternehmen müssen demnach in zunehmend heterogenen Hardwareumgebungen agieren. Orchestrierung aller Komponenten und Abstraktionsmöglichkeiten auf Plattform-Ebene werden zum zentralen Erfolgsfaktor.

### KI-Superkräfte für das Team vor Ort

Edge- und Physical AI verlagern Entscheidungsfindung stärker an die Basis. Mitarbeitende, die bisher keinen Zugang zu solch fortschrittlicher Intelligenz hatten, werden jetzt von KI-Systemen unterstützt, die direkt am Arbeitsplatz eingesetzt werden. Kollaborative Roboter werden immer häufiger repetitive und körperlich anstrengende Aufgaben übernehmen, z. B. Materialtransport, Montage oder einfache Arbeitsschritte. Gleichzeitig werden sich Menschen auf Tätigkeiten konzentrieren, die Urteilsvermögen erfordern. Dazu gehören z. B. Qualitätskontrolle, Fehlerbehandlung und Finalisierungsarbeiten – unterstützt durch Computer Vision und kleine, lokal laufende LLMs.

Die Herausforderung besteht darin, diese Funktionen sicher und konsistent an bis zu Tausenden von Standorten bereitzustellen, an denen häufig Mitarbeiter ohne IT-Expertise tätig sind.



*Autor*  
*Pdraig Stapleton*  
*Senior Vice President*  
*Chief Product Office*  
*ZEDEDA*  
*<https://zededa.com/>*

Unternehmen, denen das Management dezentraler Edge-KI-Umgebungen gelingt, werden sich in Branchen wie der Fertigungsindustrie, der Öl- und Gasindustrie sowie der Automobilindustrie, einen entscheidenden Wettbewerbsvorteil verschaffen. Das gilt umso mehr für Bereiche, in denen Fachkräftemangel herrscht.

## Small Language Modelle und Computer-Vision überzeugen

Kleine und visuelle Sprachmodelle schaffen den Sprung aus Test- und Forschungsumgebungen in die Praxis und bieten schon jetzt einsatzfähige Edge-Systeme. Anstatt auf ständige Cloud-Verbindungen angewiesen zu sein, bieten solche Ansätze unmittelbaren, lokalen Mehrwert. Ein konkretes Beispiel ist die direkte Einbettung von Gerätehandbüchern und Betriebsdokumentationen in kleine Sprachmodelle. So können Mitarbeiter im Außendienst Fehlercodes oder technische Störungen sofort beheben. Ebenso können visuelle Modelle in Edge-Umgebungen kamerabasierte Inspektion und Überwachung leisten, indem sie visuelle Daten lokal verarbeiten und mit sprachbasierten Anweisungen kombinieren – ganz ohne ständige Cloud-Verbindung. Dieser Ansatz stellt Fachwissen genau dort und dann bereit, wo es benötigt wird. Und zwar ohne Latenz oder die Abhängigkeit von einer permanenten Verbindung. Dieser Wandel bedeutet eine Abkehr von cloudabhängiger generativer KI hin zu praktischer, vor Ort verfügbarer Intelligenz, die den täglichen Betrieb unterstützt und den Mitarbeitern sofortigen Zugriff auf relevantes Wissen ermöglicht.

## Weg von zentralisierter Kontrolle hin zur Edge-Autonomie

2026 werden Organisationen ihre Betriebsmodelle weiter umstrukturieren und von zentralisierter Entscheidungsfindung am Hauptsitz hin zu dezentraler Autonomie an den Außenstellen übergehen. Mit Edge-Computing-Plattformen können Unternehmen Daten lokal in Produktionsstätten, Filialen und Betriebsstätten speichern. Lokale Teams können Daten dadurch besser managen, Anwendungen bereitstellen und aktualisieren sowie die Leistung in Echtzeit optimieren – und das alles innerhalb der



unternehmensweiten Governance- und Sicherheitsrichtlinien. Mit diesem Ansatz können Unternehmen ihre betrieblichen Herausforderungen bezüglich Skalierung Größe und dezentraler Verteilung meistern. Mit der Ausbreitung von Edge-Umgebungen wird Autonomie auf lokaler Ebene von einer Option zur Pflicht.

## Die Konvergenz von IT und OT wird zum Standardverfahren

Die seit langem bestehende Kluft zwischen IT und operativer Technologie wird sich im Jahr 2026 weiter auflösen. Hersteller setzen vermehrt IT-Workflows, Containerisierung und moderne Management-Tools direkt in der Produktionsumgebung ein. Diese Konvergenz realisiert einheitlichere Infrastrukturstrategien und integriert Cybersicherheit, Automatisierung und Lebenszyklusmanagement auf ganzheitliche Weise in industrielle Abläufe. Dadurch können Unternehmen schneller agieren und zeitgleich die für den Betrieb erforderliche Zuverlässigkeit und Sicherheit gewährleisten. Anstatt als getrennte Bereiche zu agieren, werden IT und OT Teil eines einzigen, koordinierten Betriebsmodells.

## Ab in die Zukunft

Edge-KI wird sich fest als operative Infrastruktur etablieren. Der Fokus verlagert sich vom Experimentalstadium zur Umsetzung: Künstliche Intelligenz wird in großem Umfang bereitgestellt, Mitarbeiter im direkten Kundenkontakt werden unterstützt und komplexe Edge-Umgebungen werden konsistent und kontrolliert verwaltet. Für Organisationen, die diesen Übergang bewältigen, hängt der Erfolg von der Fähigkeit ab, diverse Hardware, Anwendungen und KI-Workloads in verteilten Umgebungen sicher zu orchestrieren und so die Komplexität in der Edge in einen strategischen Vorteil zu verwandeln.

## Wer schreibt:

Als erfahrene Führungskraft mit einer Erfolgsbilanz bei der Entwicklung innovativer Lösungen unter Einsatz von KI/ML, Big Data und Cloud-Technologien hat Padraig Stapleton nachweislich globale Entwicklerteams sowohl verschiedenen Unternehmen aufgebaut und geleitet. Bevor er zu Zededa kam, war er Chief of Engineering bei Mobileum, baute und führte die Big-Data-Engineering- und DevOps-Teams bei AT&T. ◀

## FACHKRÄFTE GESUCHT?

Finden Sie Mitarbeiter, die zu Ihnen passen – mit einer Stellenanzeige in unseren Fachzeitschriften!

Fordern Sie gern ein Angebot bei mir an!



[www.beam-verlag.de](http://www.beam-verlag.de)