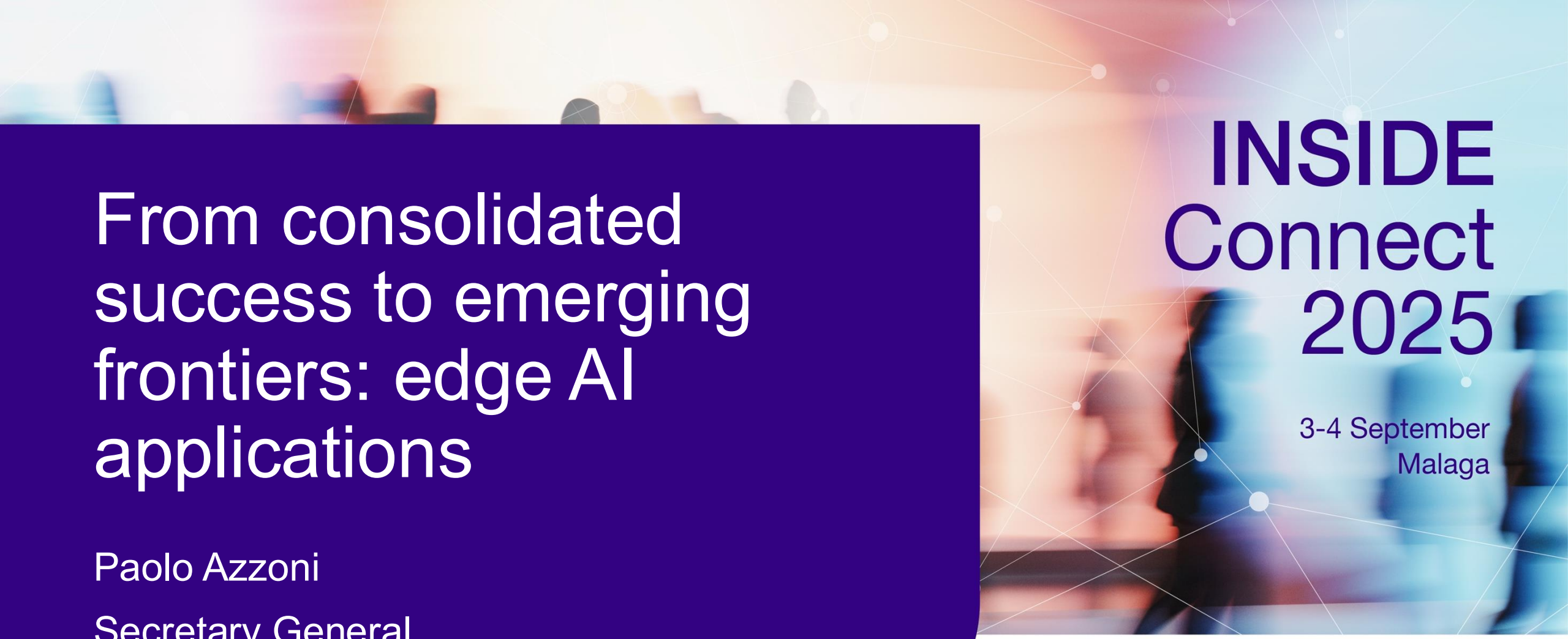


From consolidated success to emerging frontiers: edge AI applications

Paolo Azzoni

Secretary General

INSIDE Connect, Málaga, 04/09/2025

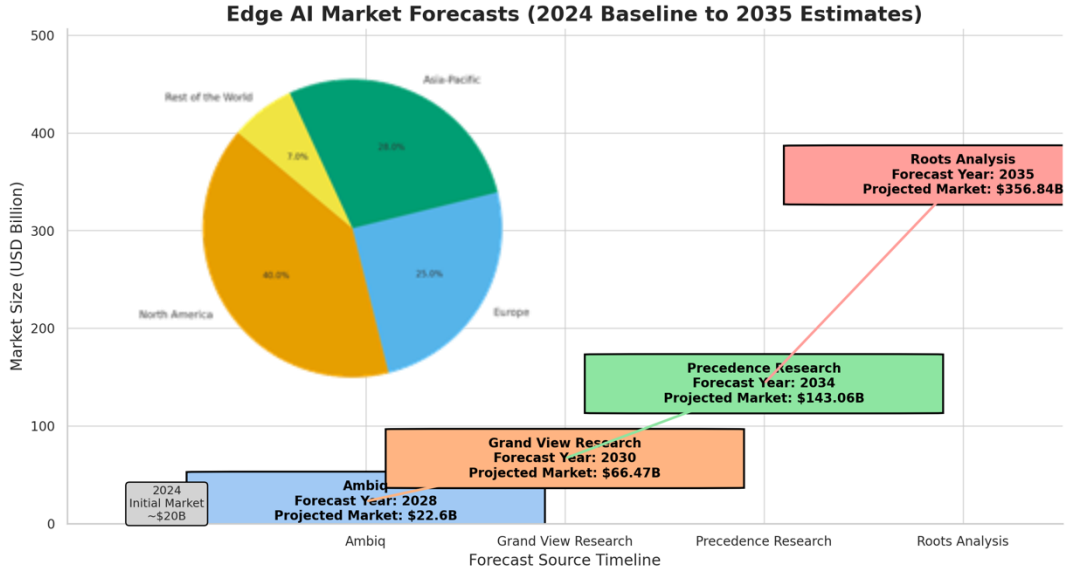
INSIDE Industry Association

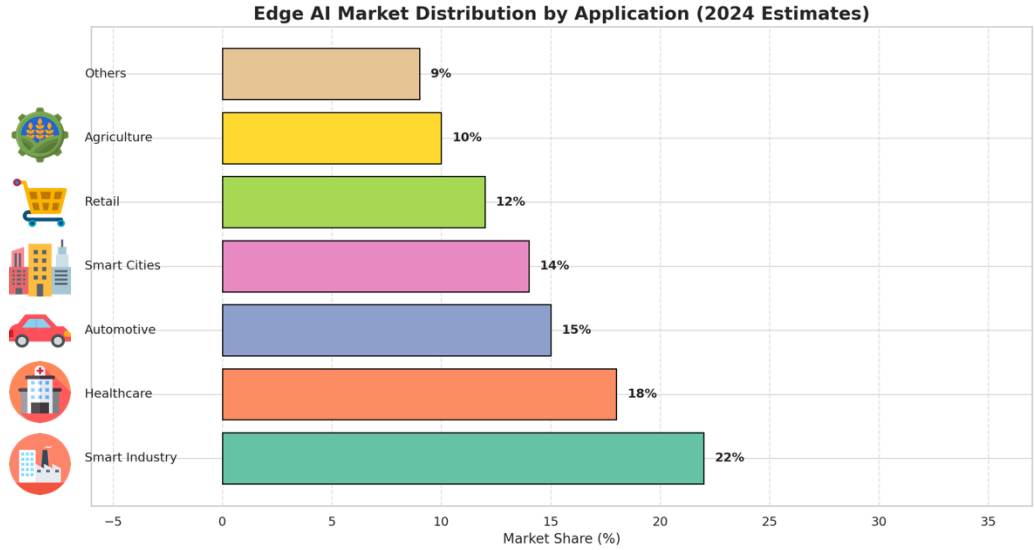
Proud member of Chips JU

The edge is the new centre

Intelligence is quickly moving from the cloud to the device.

- **Edge AI is shaping how machines see, sense, and act in real time**: decisions now happen in milliseconds, not seconds.
- **Edge AI is not a niche anymore**, a default enabler in systems that demand speed, autonomy, and privacy.
  - Global market growth: USD 356.84 B by 2035 (CAGR 27.8%)
  - North America leads with 40% share
  - Asia-Pacific is poised to be the fastest-growing region
- **Edge vs. Cloud shift**
  - Gartner predicts that by 2025, 75% of enterprise-generated data will be created and processed at the edge, compared to just 10% in earlier years.
  - Edge computing now accounts for 75% of all data computation, with the global edge computing market valued at over USD 200 B annually.





What exactly is Edge AI?

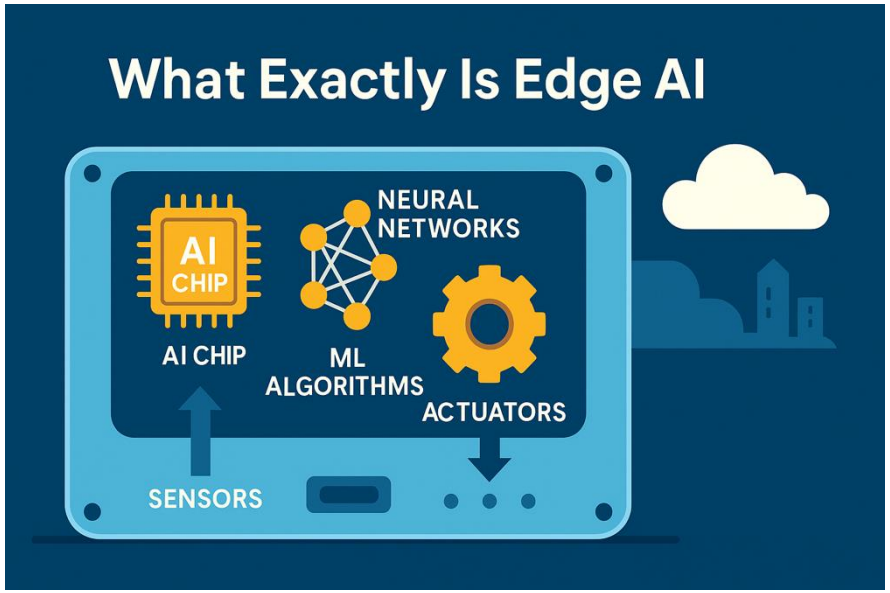
Edge AI is often confused with IoT, missing what makes it special:

- Edge AI refers to artificial intelligence that runs directly on local devices, without relying on the cloud. It brings compute closer to where data is generated.
- Edge AI combines AI and edge computing: processing data and making decisions on-site, where latency, bandwidth, privacy, security and autonomy matter most.
- **An analogy:** *if the cloud is the brain of a SoS, Edge AI is its nervous system, providing sensing and immediate reactions.*

Enablers examples:

Powered by light ML, model compression (e.g. quantization, pruning), neuromorphic technologies, and AI accelerators (e.g., Jetson Nano, Coral TPU, STM32 MCU), etc., that bring inference to the device level.

Edge AI is the fusion of data locality, computational efficiency, and embedded autonomy: it's the technical foundation of responsive, intelligent systems in the physical world



From the cloud to the edge: why now?

The transition from cloud-based AI to edge AI it's a necessity, it's a reality! And It is driven by a combination of technical constraints and domain-specific requirements:

Latency-sensitive applications

Applications like autonomous driving, robotic surgery, and predictive maintenance in smart factories **require sub-millisecond decision-making**. Waiting for round trips to the cloud introduces unacceptable delays.

Bandwidth & cost constraints

Constantly transmitting high-resolution images, video, or sensor data to the cloud is both costly and bandwidth-intensive. **Edge AI reduces this load by filtering, compressing, or fully analysing data locally.**

Privacy & data protection

Regulations like GDPR and HIPAA demand **tighter control over personal and sensitive data**. Edge AI allows processing to occur on-device or on-site, minimizing exposure and satisfying compliance.

Energy efficiency

Specialized hardware (e.g., NPU-enabled MCUs, neuromorphic chips) **enables AI workloads to run on ultra-low-power devices**, which is crucial for wearables, drones, and remote sensors.

Scalability in IoT ecosystems

In large IoT deployments, edge AI reduces the dependence on centralized infrastructure. This results in **more resilient, autonomous systems that can operate offline or in degraded networks**.

From the proven to the possible

To understand where Edge AI applications are going, it helps to look at where they already are and where they are heading next.

Quickly maturing emerging domains	Solid, fast-growing domains
Robotics (drones, humanoids)	Healthcare
Retail	Smart Home / IoT
Smart Cities	Smart Agriculture
Environmental Monitoring	Smart Industry / Manufacturing
Defense, Aerospace	Automotive
Logistics, Mobility	Energy

- **Edge AI in healthcare,**
  Juan Montiel Nelson — Universidad de Las Palmas de Gran Canaria
- **Edge AI in smart home/IoT**
  Eike Schultz – Universität zu Lübeck
- **Edge AI in smart agriculture**
  Markus Tauber – Research Studios Austria
- **Edge AI in smart industry**
  Gianvito Urgese – Politecnico di Torino
- **Edge AI in automotive**
  Mohammed Abuteir – TTTech Computertechnik AG
- **Edge AI in energy**
  Atmojo Udayanto – Aalto University

Emerging domains, quickly maturing

Robotics

- Consumer and industrial humanoid robots
- Drones & UAVs
  - Energy & Environmental Monitoring
  - Construction & Industrial Drones
  - Public Safety & Emergency Response

Retail & consumer environments

- Real-time inventory management
- In-store analytics
- Cashier-less & smart checkout
- On-Device AI for
  - Fraud detection
  - Personalization
  - Offline resilience

Smart cities & public infrastructure

- Traffic & mobility management
- Public safety & surveillance
- Smart lighting & waste detection
- Energy Systems & mobility

Environmental monitoring & ecosystem protection

- Air/water quality & pollution alerts
- Wildlife & habitat surveillance
- Edge AI for rural areas service continuity

Defence & aerospace

- Autonomous defence systems
- Tactical edge networks
- Satellite & spacecraft operations
- Surveillance & reconnaissance
- Cyber defence at the edge

Logistics & mobility

- Fleet optimization
- Autonomous delivery systems
- Real-time cargo monitoring
- Warehouse robotics & vision
- Multimodal transport hubs

Technologies and challenges uniting use cases

Key enablers

- **TinyML & Lightweight Models**

  Efficient deep learning models (like quantized CNNs or pruned transformers) that can run on microcontrollers or low-power SoCs with <1MB RAM.

- **Edge AI Accelerators**

  Hardware platforms like NVIDIA Jetson, Intel Movidius, Google Coral TPU, STM32, and Syntiant NDPs enable fast, local inference at the device level with low energy cost.

- **Federated Learning & Local Training**

  Privacy-preserving training across distributed devices, allowing models to learn collaboratively without centralizing data — critical for regulated sectors like healthcare and finance.

- **Low-Latency & Private Connectivity**

  Emerging 5G, private LTE, and TSN (Time-Sensitive Networking) architectures ensure fast communication between edge devices and coordination with cloud services when needed.

Cross-cutting challenges

- **Hardware Heterogeneity**

  Deploying AI across a fragmented ecosystem of devices (MCUs, NPUs, ASICs, FPGAs) requires custom model optimization, compilers, and orchestration tools.

- **Model Optimization & Deployment**

  Running modern AI on the edge demands significant adaptation: quantization, pruning, distillation, and even architecture redesign. No “one-size-fits-all” model.

- **Scaling & Managing Edge AI Systems**

  Coordinating thousands of deployed AI endpoints means solving DevOps for the edge: version control, A/B testing, observability, rollback strategies, and dynamic model updates.

- **Privacy, Trust & Governance**

  Keeping data local doesn’t eliminate risk — it changes the threat surface. Organizations must address explainability, compliance (e.g. GDPR), and AI ethics at the device level.

Where do we go from here?

What if every sensor could think?

- Imagine a world where trillions of sensors and devices don't just collect data, **they interpret, act, and collaborate autonomously**.

From devices to environments: embedded intelligence as infrastructure

- Edge AI will be woven into roads, bridges, buildings, fields, and factories, turning passive systems into active, adaptive ones.

Edge AI is redefining infrastructure

- Not just enabling use cases, it's reshaping the **architecture** of systems and SoS in healthcare, agriculture, mobility, energy, and the built environment.

A call for ethical design, resilience, and societal alignment

- The more power we distribute to the edge, the more responsibility we carry to ensure it's **safe**, **transparent**, **inclusive**, and **resilient**.

References

- **Grand View Research (2024) -** *“Edge AI Market Size, Share & Trends Analysis Report By Component, By Data Source, By End-use, By Region, And Segment Forecasts, 2024 - 2030”*, https://www.grandviewresearch.com/industry-analysis/edge-artificial-intelligence-ai-market
- **Precedence Research -** *“Edge AI Market (By Component, Data Source, End-Use, Region): Global Opportunity Analysis and Industry Forecast, 2024–2034”*, https://www.precedenceresearch.com/edge-ai-market
- **Roots Analysis (2024) -** *“Edge AI Market, 2024–2035”*, https://www.rootsanalysis.com/reports/edge-ai-market.html
- **Ambiq (2024) -** *Ambiq: Edge AI Ecosystem Stats (from corporate publications & press releases)*, https://ambiq.com/news
- **MarketsandMarkets (2024) -** *Edge AI Market by Region Report*, https://www.marketsandmarkets.com/Market-Reports/edge-ai-hardware-market-256103867.html
- **Statista (2024) -** *IoT Connected Devices Worldwide 2019–2030,* https://www.statista.com/statistics/1101442/iot-number-of-connected-devices-worldwide/
- **Gartner -** *“Gartner Says 75% Of Data Will Be Processed At The Edge By 2025”*, https://www.forbes.com/sites/forbestechcouncil/2021/12/17/gartner-says-75-of-data-will-be-processed-at-the-edge-by-2025
- **IDC & Intel (2024) -** *“Edge Computing: 75% of All Data Computation Now Happens Outside the Cloud”*, https://www.intel.com/content/www/us/en/edge-computing/overview.html
- **IEEE Spectrum (2023–2024) -** *Articles on edge AI hardware (neuromorphic chips, NPUs)*, https://spectrum.ieee.org/tag/edge-ai

Thank you for watching

INSIDE Industry Association

Proud member of Chips JU

INSIDE Connect 2025

The edge is the new centre

- **Global Edge AI Market Growth**
- Estimated at **USD 20.78 billion in 2024**, projected to reach **USD 66.47 billion by 2030**, at a **CAGR of 21.7%** .
- Other forecasts estimate an even more aggressive rise—from **USD 24.05 billion in 2024** to **USD 356.84 billion by 2035** (CAGR 27.8%) .
- Another projection: **USD 21.19 billion in 2024** to **USD 143.06 billion by 2034**, CAGR ~21% .
- **Regional Insights**
- **North America** accounted for **40% of the global edge AI market in 2024**, with the **U.S. alone at USD 5.93 billion**. By 2034, the U.S. market is expected to reach **~USD 45.85 billion** .
- **Asia-Pacific** is poised to be the fastest-growing region, driven by urbanization, smart city projects, and 5G expansion .
- **Ultra-Low Power Edge AI Device Adoption**
- Ambiq estimates the **edge-AI market** to be around **USD 13.9 billion in 2025**, increasing to **USD 22.6 billion by 2028**, particularly in **smart-home, medical, and wearable devices**. Ambiq chips are already used in **270 million devices**, with **42 million units shipped last year**—>40% of which are AI-capable .
- **Edge Processing vs. Cloud Shift**
- Gartner predicts that by **2025, 75% of enterprise-generated data** will be **created and processed at the edge**, compared to just 10% in earlier years .
- Edge computing now accounts for **75% of all data computation**, with the global edge computing market valued at **over USD 200 billion annually** .
- **Speaker Notes**

- “Let’s start with the numbers—Edge AI isn’t niche. The global market is already in the tens of billions: around **USD 20–24 billion in 2024**, with strong projections reaching as high as **USD 66 billion by 2030** or even **USD 143 billion by 2034**—and some forecasts stretch to **USD 357 billion by 2035** .
- Regionally, North America dominates with roughly **40% share**, and the U.S. market alone is set to skyrocket toward **USD 46 billion** by 2034. Meanwhile, Asia-Pacific is surging fastest, led by smart city and 5G investments .
- On the device front, Ambiq—a leader in ultra-low-power edge AI chips—notes that the edge AI sector is already worth **USD 13.9 billion today**, with rapid growth expected by 2028. Their chips are in over **270 million devices**, including wearables and medical hardware

- And critically: by 2025, Gartner expects roughly **75% of enterprise data to be processed at the edge**, up from just 10% previously—marking a seismic shift toward localized intelligence . In fact, edge computing already accounts for **3/4 of data computation worldwide**, with a market valuation exceeding **USD 200 billion** .
- These metrics underscore that Edge AI is not just emerging—it’s rapidly becoming the default for speed, autonomy, and scalability.”

From the cloud to the edge: why now?

The transition from cloud-based AI to edge AI is no longer optional, it's a necessity!

It is driven by a combination of **technical constraints** and **domain-specific requirements**

- **Latency-sensitive applications**
  - Applications like autonomous driving, robotic surgery, and predictive maintenance in smart factories require **sub-millisecond decision-making**. Waiting for round trips to the cloud introduces unacceptable delays.
- **Bandwidth & cost constraints**
  - Constantly transmitting high-resolution images, video, or sensor data to the cloud is both **costly** and **bandwidth-intensive**. Edge processing reduces this load by filtering, compressing, or fully analyzing data locally.
- **Privacy & data protection**
  - Regulations like **GDPR** and **HIPAA** demand tighter control over personal and sensitive data. Edge AI allows processing to occur **on-device or on-site**, minimizing exposure and satisfying compliance.
- **Energy efficiency**
  - Specialized hardware (e.g., **NPU-enabled MCUs**, neuromorphic chips) enables AI workloads to run on **ultra-low-power devices**, which is crucial for wearables, drones, and remote sensors.
- **Scalability in IoT ecosystems**
  - In large IoT deployments, pushing intelligence to edge nodes reduces dependence on centralized infrastructure. This results in **more resilient, autonomous systems** that can operate offline or in degraded networks.