

# The Ultimate Guide to Edge AI

The playbook for edge AI  
business transformation



EDGE IMPULSE

[edgeimpulse.com](https://edgeimpulse.com)

# AN INTRODUCTION FROM OUR CO-FOUNDER AND CEO

Artificial Intelligence (AI) is the latest disruptive technology that has captured the collective imagination of individuals and businesses alike. While AI offers many benefits and is already making its way into our personal lives, edge AI is emerging as the real game changer for businesses.

Edge AI — running AI algorithms directly on devices in the physical world rather than relying on the cloud — has rapidly evolved into a critical modern advancement utilized not only for commercial operations but also for applications that significantly impact quality of life. As we navigate a world increasingly driven by data, the ability to process information and make real-time decisions at the source has become not only a necessity but a distinct competitive advantage.

## How are companies using edge AI to solve business problems? What are they building, and how will their customers benefit from this technology?

Consider a manufacturing plant where edge AI-powered sensors can detect and predict equipment failures before they occur, a clinic offering more personalized and timely patient care with wearable devices capable of detecting anomalies and alerting healthcare providers instantly, or a car that can listen to the audio of its surroundings and detect events of interest, i.e., a collision or theft. These are all real on-device applications that are being built today, along with thousands of others, and we are just at the starting point. The use cases for edge AI are limitless.

Product managers and business executives are feeling the pressure to put AI into more products to meet consumer and industrial

demand for intelligent devices. In “**The Ultimate Guide to Edge AI**,” we shine the spotlight on how edge AI is transforming businesses, offering both practical and strategic insights for leveraging this groundbreaking technology.

Whether you’re a business leader looking to see dramatic improvements in operational efficiency, customer satisfaction, and a competitive advantage or a developer pushing the boundaries of what’s possible, edge AI has the potential to reshape your organization.

The future of business is intelligent, distributed, and happening at the edge. Discover how you can lead the way in the pages of this guide.



**Zach Shelby**  
**Co-Founder and CEO**  
**Edge Impulse**

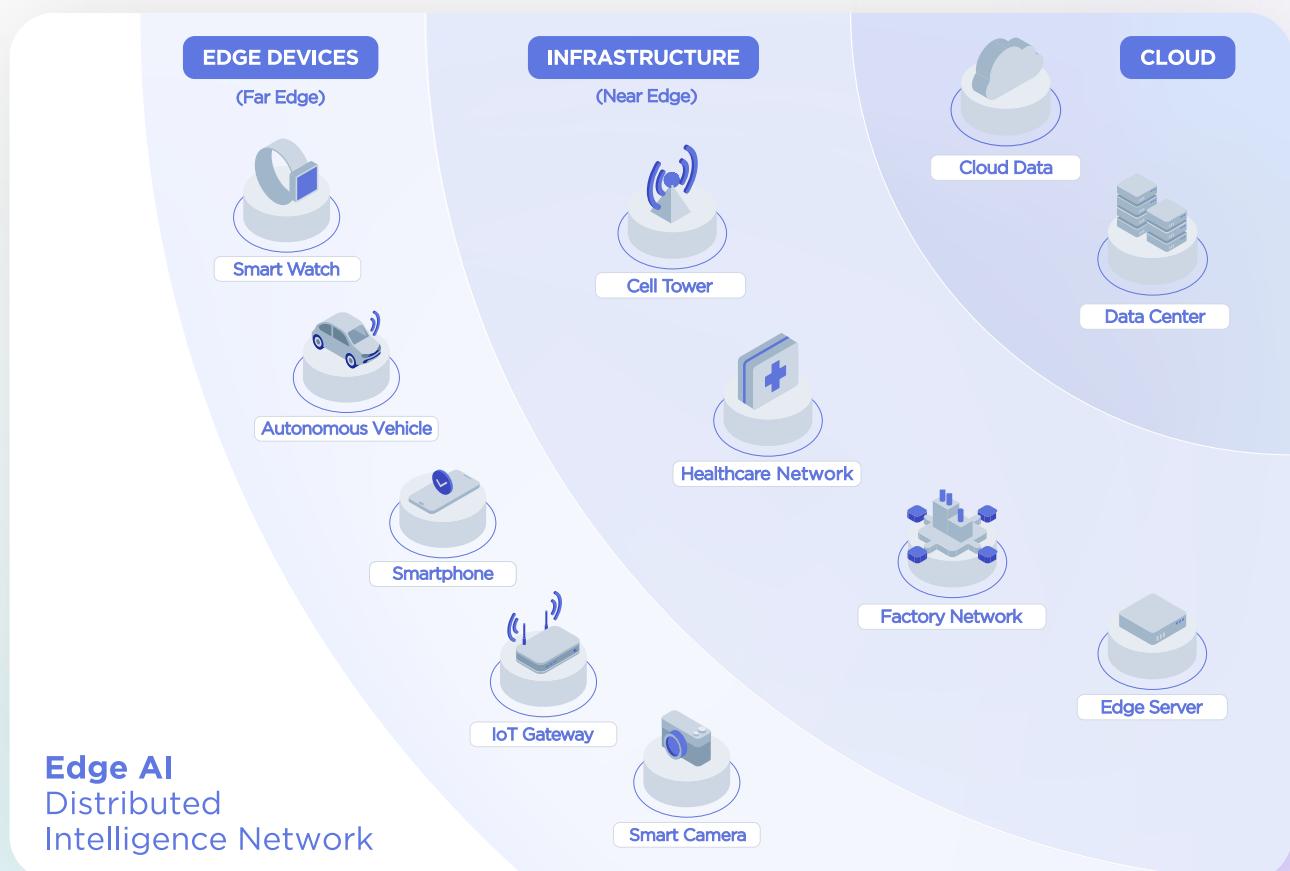
# Table of Contents

<b>Chapter 1:</b> How Edge AI is Reshaping the Business Landscape	<b>4</b>
<b>Chapter 2:</b> Edge AI Enablers in Intelligent Technologies	<b>7</b>
<b>Chapter 3:</b> The Business Case for Edge AI	<b>10</b>
<b>Chapter 4:</b> Real-world Case Studies and Success Stories	<b>12</b>
<b>Chapter 5:</b> Cutting Costs and Boosting Profits with Edge AI	<b>14</b>
<b>Chapter 6:</b> Accelerating AI Innovation with Edge AI + Cloud Computing	<b>15</b>
<b>Chapter 7:</b> The Rise of Gen AI and Foundational Models	<b>18</b>
<b>Chapter 8:</b> Powering the Future with Cutting-edge AI Hardware and Software	<b>19</b>
<b>Chapter 9:</b> Why Model Efficiency is Crucial to Successful Edge AI Deployment	<b>22</b>
<b>Chapter 10:</b> Getting Started with Edge AI	<b>23</b>
<b>Chapter 11:</b> In-House Development or an End-to-End Solution?	<b>29</b>
<b>Chapter 12:</b> Challenges and Barriers to Edge AI Adoption	<b>31</b>
<b>Chapter 13:</b> The Transformative Power of Edge AI	<b>33</b>
<b>Chapter 14:</b> The Future of Edge AI	<b>35</b>
<b>About Edge Impulse</b>	<b>37</b>

# HOW EDGE AI IS RESHAPING THE BUSINESS LANDSCAPE

Edge AI — the fusion of artificial intelligence with edge computing — is not just the new technology on the block; it's the key to unlocking unprecedented levels of efficiency, responsiveness, and innovation across every industry. Imagine a world where the boundaries between the physical and digital blend seamlessly, creating interconnected systems, smart devices, and immersive experiences. With advances in real-time computing and sensor technology, edge intelligence is reshaping how we live and work.

Businesses leverage these innovations to improve performance, while consumers benefit from personalized experiences like never before. From factories where machines predict and prevent their own failures, to healthcare devices that can make life-saving decisions, edge AI is making the seemingly impossible, possible. It's not only about faster computing or smarter algorithms; it's about bringing intelligence to the very edge of the digital world, where data is born and actions take place.



# Demystifying the edge: understanding the technology and its potential

Edge AI involves deploying intelligent algorithms to run *directly* on edge devices such as sensors, cameras, and industrial controllers rather than relying solely on centralized cloud servers. This approach enables real-time data processing, which is essential for applications requiring low latency, enhanced privacy, energy efficiency, and immediate decision-making.

Edge AI doesn't exist in a vacuum, however; some interconnected systems and technologies make it possible, including:

**Artificial Intelligence (AI)** — computer systems designed to make intelligent decisions based on data, often in a real-world context. These systems may be based on statistical models trained on large amounts of data or on rule-based programs created by software engineers.

**Machine Learning (ML)** — a subset of AI that focuses on developing algorithms that allow computing systems to learn from and make predictions based on data without being explicitly programmed for each task.

“By 2026, at least 50% of edge computing deployments will involve machine learning, compared to 5% in 2022.”



Source: Gartner Hype Cycle™ for Edge Computing, 2024, Thomas Bittman, 15 July 2024

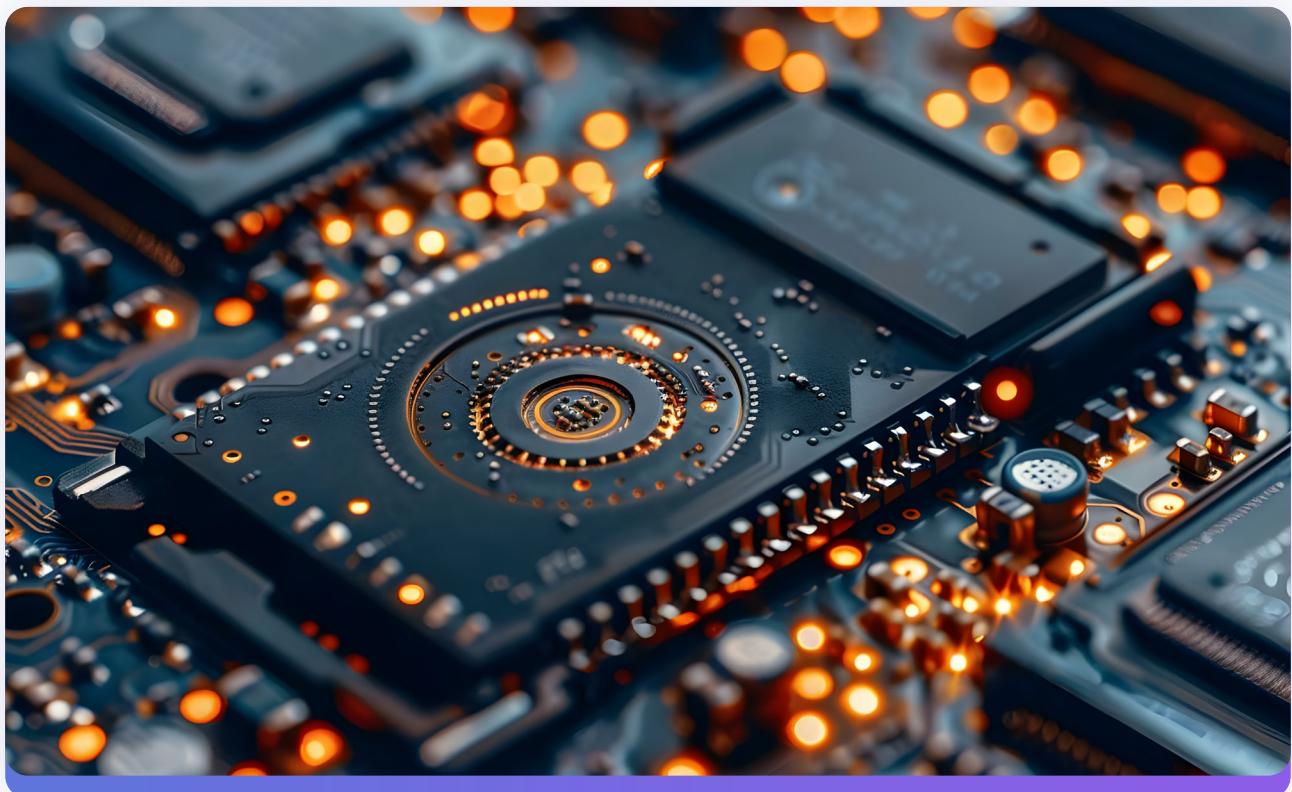
GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally, and is a registered trademark of Gartner, Inc. and/or its affiliates and is used herein with permission. All rights reserved.

**The Internet of Things (IoT)** — a vast network of interconnected devices embedded with electronics, software, sensor, and network connectivity which enables these devices to collect and exchange data.

# Computing at the edge: from MCUs to MPUs to GPUs and NPUs

Edge AI spans a wide spectrum of computing capabilities, tailored to meet diverse application needs:

- **Microcontroller Units (MCUs)** — Ideal for ultra-low power, cost-sensitive applications. Examples include smart sensors, wearables, and simple IoT devices.
- **Microprocessors (MPUs)** — Sophisticated, integrated circuits that serve as a central processing unit (CPU) of a computer or advanced electronic system. Key characteristics include higher processing power, making them ideal for applications requiring significant computational power and versatility. Ideal cases for MPUs include gaming consoles, smartphones, tablets, and industrial control systems.
- **GPU-accelerated systems** — High-performance computing for intensive AI tasks. Use cases include autonomous vehicles, advanced robotics, and real-time video processing.



- **Neural Processing Units (NPUs)** — Specialized processors designed to accelerate neural network operations and AI tasks. Unlike general-purpose CPUs and GPUs, NPUs are optimized for parallel computing, making them highly efficient at handling AI-related tasks such as speech recognition, image processing, and more.
- **Hybrid Edge-cloud Systems** — Combining local processing with cloud resources. Examples include distributed AI for smart cities and large-scale IoT deployments.

As these devices become faster and more efficient, their ability to run AI models grows without the need for dedicated onboard AI-accelerated hardware. Nordic Semiconductor, Infineon, Espressif, Renesas, Silicon Labs, Texas Instruments, NXP, STMicroelectronics, and Microchip all have devices with demonstrated AI capabilities in different ways to support different markets.

We'll delve into edge AI-specific hardware in Chapter 8.

# EDGE AI ENABLERS IN INTELLIGENT TECHNOLOGIES

AI is a broad field encompassing multiple subfields and technologies, each with unique characteristics and applications. Edge AI and generative AI stand out as distinct approaches in the AI landscape. Let's take a closer look at the differences between the various subfields.

## Defining the intelligence spectrum

### AI:

Some of the key characteristics include:

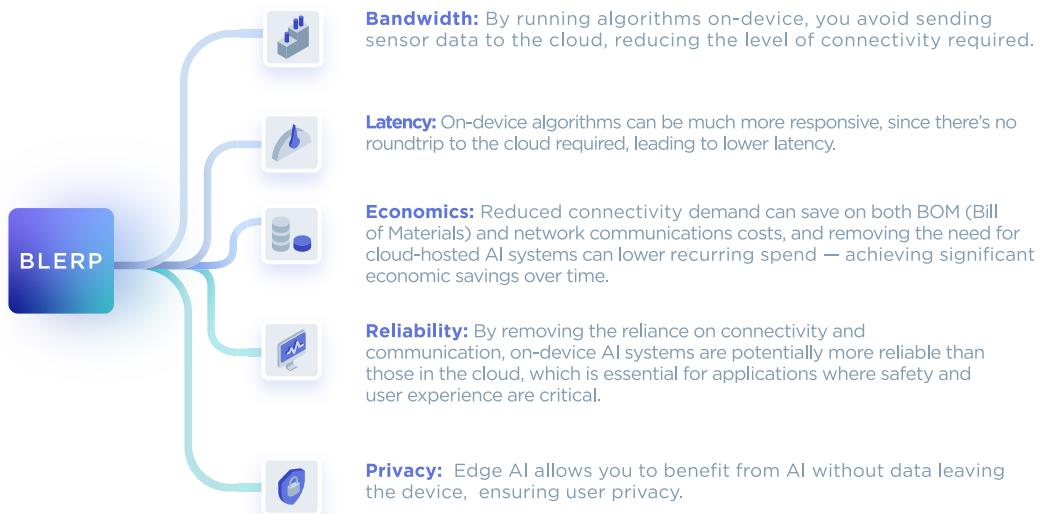
- **Data-driven** — AI systems learn from large amounts of data to make predictions, recognize patterns, and improve performance over time.
- **Versatile applications** — AI can be used in multiple fields, including manufacturing, healthcare, finance, transportation, and more.
- **Centralized processing** — Traditional AI models depend on resource-heavy, centralized servers and cloud computing for data processing and model training. Examples include predictive analytics in finance, natural language processing for chatbots, and image and speech recognition patterns.

### Edge AI:

Edge AI is a subset of AI where data processing and AI computation are performed locally on edge devices or hardware rather than in centralized data centers or the cloud. Edge devices include sensors, microcontrollers, and other IoT devices that generate and process data at the network's edge. This offers key advantages over cloud AI, including:

- **Local processing** — AI computations occur on the device where data is generated, reducing the need for transmission to any servers.
- **Low latency** — Immediate data processing and decision-making are possible and critical for real-time applications like security and safety.
- **Energy efficiency** — By processing data locally on-device, edge AI reduces the need for constant communication with cloud servers, conserving precious battery life and enabling more optimized and efficient use of energy.
- **Enhanced privacy** — Sensitive data remains on the device, reducing the myriad of security risks associated with transmitting data over networks, such as eavesdropping and man-in-the-middle attacks.
- **Efficient bandwidth usage** — Only relevant data or insights are sent to the cloud, saving bandwidth and reducing costs.

Examples include real-time video analytics on security cameras, health monitoring with wearable devices, and predictive maintenance for industrial machinery.



## Generative AI:

Gen AI focuses on creating new data instances that resemble existing data. It involves training models to generate text, images, audio, and other forms of content. These models learn the underlying patterns and structure of the input data and use this knowledge to produce new, similar data. Key characteristics include:

- **Content creation** — Gen AI excels at creating new content, including text, images, and code.
- **Deep learning models** — They often rely on advanced neural networks such as transformer architectures. Transformers are the backbone of many modern natural language processing (NLP) models<sup>1</sup>.
- **Creative applications** — Art installations, entertainment, marketing, and personalized content generation.

## Gen AI at the edge

At the edge, gen AI represents the convergence of two transformative technologies: edge AI, focused primarily on inference and decision-making based on pre-trained models, and gen AI, based on general knowledge of the internet and excels at tasks like labeling and generating samples of human speech or sounds.

Both work powerfully together when gen AI is used to train better edge models, for example, via synthetic data or knowledge distillation (where knowledge from a large model is “distilled” into a smaller one).

According to [Bloomberg Intelligence](#), gen AI is poised to become a \$1.3 trillion market by 2032<sup>2</sup>, and rising demand for gen AI products could add \$280 billion of new software revenue. As the market expands, we expect to see a significant portion of the growth driven by edge-based gen AI applications. Could the \$280 billion in new software revenue suggest a surge in demand for lightweight, efficient gen AI models optimized for the edge? We certainly think so.



— Size of gen AI market by 2032.

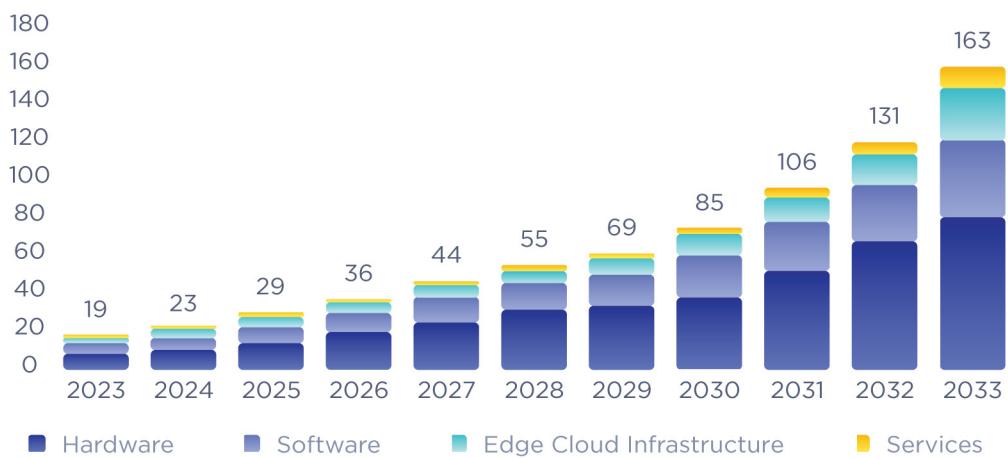
# THE BUSINESS CASE FOR EDGE AI

With its numerous compelling benefits, edge AI is an attractive technology in an increasingly data-driven world. Let's take a deeper dive into the business case for edge AI.

## Market momentum — global edge AI growth and projections

The global edge AI market size is expected to be worth around \$163 Billion USD by 2033, up from \$19 Billion USD in 2023<sup>3</sup>.

**Global Edge AI Market** Size, by Component, 2023-2033 (USD Billion)



The Market Will Grow at the CAGR of: **24.1%**

The Forecasted Market Size for 2033 in USD: **\$163B**

market.us  
ONE STOP SHOP FOR THE REPORTS

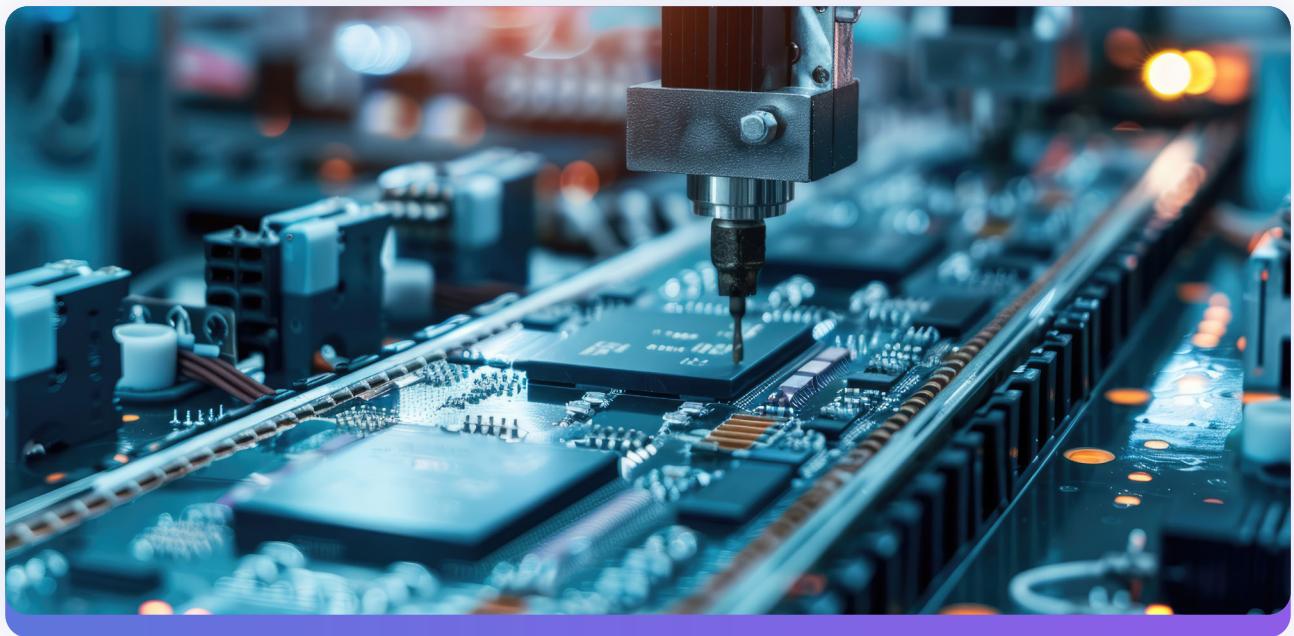
## Industry impact — top applications driving edge AI adoption

### Manufacturing

Edge AI is revolutionizing manufacturing through predictive maintenance, quality control, and process optimization. Fortune Business Insights predicts the global smart manufacturing market will grow to \$998.99 billion by 2032<sup>4</sup>. The market size is projected to increase from \$349.48 billion in 2024 to \$998.99 billion by 2032. Emerging technologies such as AI, cloud computing, big data, and machine learning are expected to drive this growth.

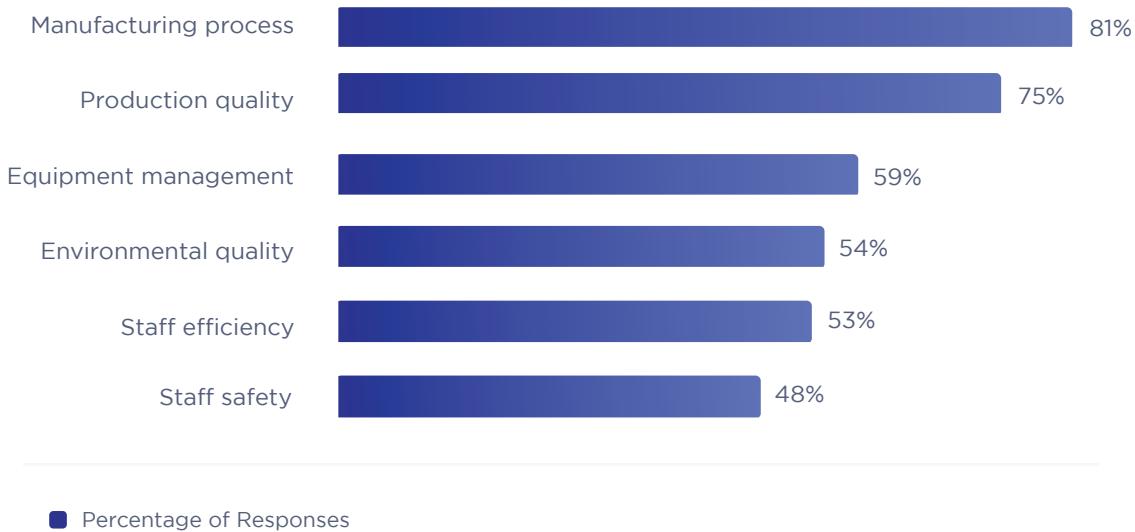
In a recent report by Edge Impulse and Manufacturing Dive, 100% of those surveyed responded **yes** to the question of whether their companies were working on initiatives to integrate AI and/or ML into their manufacturing environment.

Furthermore, 81% of respondents indicated that improvements in the manufacturing process are the area they're looking to leverage AI/ML the most, followed by production quality at 75%.



### Edge Impulse-Manufacturing Dive survey

These survey results reveal a considerable shift in the manufacturing sector's approach to AI and ML. The fact that 81% of respondents called out the manufacturing process itself as the area where AI and ML could have the most impact is particularly telling. This suggests a focus on core operations — possibly targeting efficiency, quality improvements, and cost reductions in the production line.



## Healthcare

In healthcare, edge AI is enabling faster diagnostics, personalized treatment plans, and remote patient monitoring. Companies like GE Healthcare use medical imaging devices to assist in quick and accurate diagnoses. Such devices allow healthcare providers to continuously monitor a person's health and gather large amounts of data.

“In the wearable monitor and healthcare market, we are now experiencing devices with built-in intelligence and machine learning capabilities powered by low-power yet powerful wireless computers like Nordic’s nRF54 Series Bluetooth LE and nRF91 Series Cellular IoT product lines. When critical situations happen, every second counts. Edge Impulse’s AI/ML capabilities and Nordic’s wireless solutions ensure that medical devices can make standalone and crucial decisions in real-time, backed by robust, global connectivity. Healthcare professionals can have confidence in the data and act swiftly to improve patient outcomes.”

---

— THOMAS SØDERHOLM, VP, BUSINESS DEVELOPMENT, NORDIC SEMICONDUCTOR



# REAL-WORLD CASE STUDIES AND SUCCESS STORIES

## Healthcare

*Disease insights and management: Hyfe*

Examples of edge AI in healthcare include Hyfe, which uses edge AI to extract **actionable insights from coughs**. Hyfe imported the world's largest cough dataset into Edge Impulse, generating a cough detection model capable of fitting on an Arm Cortex-M33 processor.

Hyfe initially utilized a traditional model approach where the model was trained on raw data. But to achieve the desired results, a pivotal step needed to be addressed: Digital Signal Processing (DSP).

DSP — the analysis, manipulation, and synthesis of digital signals from sources such as audio, images, or sensor data — is a can't-miss element of developing effective ML algorithms. DSP cleans up noisy signals, extracts the most relevant signals, and helps remove meaningful features from data.

By leveraging Edge Impulse's turnkey DSP-based extractors, Hyfe is successfully creating AI models that are not only effective in cough detection but also compatible and efficient on various devices, including high-performance hardware with low latency and small memory footprints. With this success, the company now supplies the Hyfe SDK to the Nordic nRF52-powered ActiGraph LEAP wearable, among other devices.

For people suffering from refractory chronic cough, along with conditions like asthma, chronic obstructive pulmonary disease, and allergies, the Hyfe application provided something previously unavailable — a way to qualify and quantify their coughs. The Hyfe app quickly became a resource for those sufferers, highlighting the importance of cough as a data point in managing and understanding various respiratory conditions.



### *Assisted living/elder care: CarePredict*

Anyone with an elderly relative who has suffered a fall knows how devastating such an event can be. CarePredict, a company specializing in home care and assisted living, is working on a fall detection device for the elderly, after the founder observed falls as a common problem that impacted his parents and many others.

However, CarePredict had a major problem: the occurrence of too many false positives in their current algorithm. They turned to Edge Impulse for model development.

With Edge Impulse's edge AI capabilities, CarePredict's fall detection algorithm has seen a 65% reduction in false alarms compared to their previous algorithm.



## Manufacturing

### *Predictive maintenance*

Manufacturers continuously monitor equipment performance by analyzing data from sensors and IoT devices in real time. Edge AI algorithms enable the **detection of anomalies, potential failures**, and recommend preemptive maintenance actions, helping to minimize downtime and maximize asset utilization.

## Consumer products and services

### *Electronics: Poly, an HP company*

In 2023, Poly, a subsidiary of HP, released two new Bluetooth headsets: the Voyager Surround 80 and 85 models, as well as the Voyager Free 60 (and 60+) earbuds. Poly Voyager Free 60 earbuds and Poly Voyager Surround 80 and 85 headsets can now be controlled using voice commands, allowing users to answer or ignore an incoming call by saying "Answer" or "Ignore," respectively.

This voice-controlled feature, using a keyword-detecting AI algorithm running directly on-device without any cloud interfacing at all, is the culmination of months of collaboration between Poly and Edge Impulse to train, build, and deploy en masse, not just a novel function that responds to the two keywords, but works with eleven different languages.

HP used Edge Impulse's platform to give HP engineers the tools to fast-track their model development and deployment phase. In fact, with Edge Impulse's help, HP was able to collect keyword data, train a production-grade ML model, and deploy it into their own custom workflow within just months.



## Smart cities

Edge AI is critical in developing smart cities, powering everything from traffic management to energy distribution. Singapore's [Smart Nation Initiative](#) extensively uses edge AI for urban planning and development<sup>5</sup>.

Smart camera systems designed to monitor and optimize traffic flow and enhance public safety are becoming more common components of modern urban infrastructure. These advanced systems leverage cutting-edge technologies, such as AI and computer vision, to analyze and respond to real-time traffic conditions. By deploying these [smart camera systems](#) at strategic locations throughout cities, municipalities can address a variety of challenges and significantly improve overall urban efficiency.

# CUTTING COSTS AND BOOSTING PROFITS WITH EDGE AI

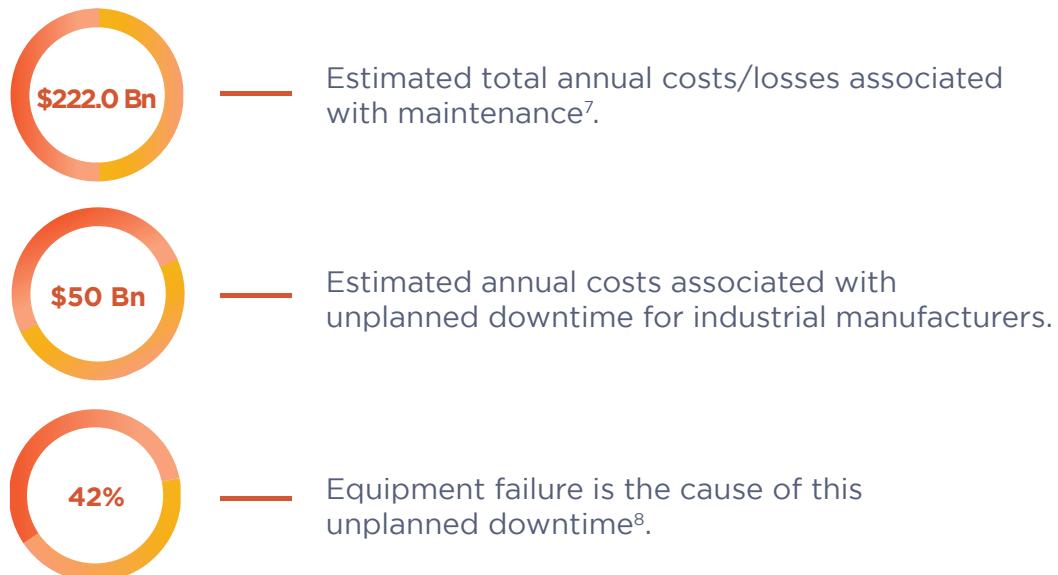
Edge AI adoption can reshape your bottom line if done right. It also has the potential to be an effective differentiator and competitive advantage. By embedding intelligence directly into devices and systems, businesses can offer unprecedented levels of personalization, efficiency, and functionality that set their products apart in crowded markets.

Edge AI's ability to operate in low-connectivity environments is opening up new use cases and customer experiences that were previously impossible. For instance, think of a smart home device with edge AI that can respond to user preferences without cloud latency or first responders who might "go dark" during a disaster and need to communicate.

By bringing intelligence directly to where data is generated, edge AI slashes operational costs, opening opportunities for increased profitability. Imagine reducing cloud computing expenses by as much as 35% year over year<sup>6</sup> as data processing shifts to the edge, or cutting network bandwidth costs.

In manufacturing, edge AI, via predictive maintenance, continuously monitors equipment in real time, detecting anomalies humans might miss. This practice extends equipment lifespan, offering up cost savings by addressing issues earlier, avoiding major breakdowns before they occur.

## The hidden costs of downtime — industrial machinery



By adopting edge AI, businesses across industries and other areas are not only improving operational efficiency by knowing when equipment will fail before it actually happens, but also **creating new revenue streams** and enhancing customer experiences. As the technology continues to grow, its transformative potential is only expected to increase, making it a critical consideration for forward-thinking organizations.

# ACCELERATING AI INNOVATION WITH EDGE AI + CLOUD COMPUTING

The debate between edge AI and cloud computing has gained significant traction as the demands for real-time data processing and low-latency responses grow. Cloud computing offers immense computational power, scalable resources, and centralized data storage. But as with any technology, there are both benefits and challenges. Organizations that leverage the strength of both cloud and edge AI will be well-positioned to accelerate their AI innovation ambitions.

## Edge AI strengths over cloud AI

Edge AI offers several distinct advantages over cloud computing, making it the preferred choice for many applications, especially those that require split-second decision-making.

*Reduced latency* — enables real-time processing and immediate responses by eliminating the need to send data to and from remote servers. This is significant in time critical applications where every millisecond counts.

*Enhanced privacy* — keeps sensitive data local, reducing the risk of breaches during data transmission. This is significant to prevent patient data from being leaked or preventing hacks of sensitive data.

*Improved reliability/offline functionality* — continues to function even with poor or no internet connectivity, ensuring uninterrupted operation in critical applications.

*Bandwidth efficiency* — reduces the amount of data sent over networks, significantly lowering cloud costs, and improving performance in bandwidth-constrained environments.

*Deployment scalability* — allows for easier deployment and management of AI capabilities across numerous devices without straining centralized resources.

*Energy efficiency* — consumes less power by minimizing data transmission, extending battery life of mobile and IoT devices, resulting in a much more marketable product.

*Personalization* — facilitates more personalized user experiences by processing user data locally, enhancing privacy and responsiveness.

## Limitations of edge AI

*Limited processing power* — edge devices typically have less computation capacity than cloud servers, constraining the complexity of AI models that can be run.

*Storage constraints* — edge devices often have limited memory and storage, restricting the size of AI models and the amount of data that can be stored locally.

*Update challenges* — deploying updates to AI models across numerous edge devices can be logistically complex and time-consuming.

*Limited data aggregation* — edge devices may not have access to the breadth of data availability in cloud environments, potentially limiting AI's learning and adaptation capabilities.

# Edge AI-powered wearable devices for health monitoring

*Offline or remote operations* — in scenarios where internet connectivity is intermittent or unavailable, such as during a [natural disaster](#)<sup>9</sup>.

*Cost efficiency for high-volume data* — for applications generating vast amounts of data, edge AI can significantly reduce cloud storage and bandwidth costs by processing data locally.

*When immediate responses are needed* — when real-time processing is required for immediate responses, such as autonomous vehicles or industrial robots, edge AI eliminates network latency.

*Data sensitivity* — when dealing with sensitive data, i.e., patient information, edge AI provides a way to process information locally without exposing the information to potential vulnerabilities in transit, such as during round trips to the cloud.



edge AI-powered wearable devices  
for health monitoring.

## Cloud AI advantages

**Powerful computing resources** — Another key advantage of cloud-based AI is access to near limitless computing resources. Leading cloud providers such as Amazon Web Services (AWS) offer vast amounts of processing power, storage, and advanced AI tools that are otherwise prohibitively expensive for most organizations to maintain in-house. This access allows businesses to perform complex computations, run sophisticated machine learning algorithms, and manage large datasets with ease.

**Computational resource flexibility** — Cloud-based AI offers unparalleled advantages in scalability and flexibility. Companies can start with minimal resources and expand their computational capacity as needed without investing in expensive hardware. The flexibility of cloud-based AI also means that companies can quickly adapt to changing market demands, deploy new AI models, and integrate emerging technologies with minimal disruption.

**Centralized data processing** — Centralized data and analysis further enhance the value of cloud-based AI. By aggregating data in a centralized cloud environment, businesses can perform comprehensive analyses and derive insights that drive strategic decision-making.

## Limitations of cloud-based AI

**Latency issues** — Despite its many advantages, cloud AI also presents several limitations that businesses must consider. One of the primary challenges is latency issues. Since data must travel from the user's device to the cloud server for processing and then back again, this roundtrip time can introduce delays, which can be problematic for applications requiring real-time responses. Latency remains a critical issue in cloud-based applications, particularly in scenarios where milliseconds can make a significant difference.

**Connectivity requirements** — Another considerable limitation of cloud AI is its dependency on stable and high-speed internet connectivity. For remote or rural areas with unreliable internet access, relying on cloud AI can lead to inconsistent performance and service disruptions. This connectivity requirement can also hinder the deployment of AI solutions in regions where infrastructure is not robust, limiting the potential reach and effectiveness of cloud-based AI systems.

As reported by the International Telecommunication Union (ITU), around 37% of the global population still lacks internet access, highlighting the [connectivity gap](#) that can affect cloud AI adoption<sup>10</sup>.

**Data privacy and security concerns** — Safeguarding sensitive information is also paramount when considering cloud AI. Storing such data in the cloud raises the risk of breaches and unauthorized access. Despite stringent security measures implemented by cloud providers, the centralized nature of cloud storage can make it an attractive target for cyberattacks.

**Ongoing operational costs** — While cloud AI offers scalability, the costs associated with continuous data transfer, storage, and computational power can add up over time. Cloud infrastructure continues to be one of the fastest-growing business expenditures.



## When to choose cloud over edge

*Complex, resource-intensive computations* — When your AI models require massive computational power or need to process extremely large datasets, cloud computing's scalable resources are often more efficient than edge devices.

*Non-time-sensitive applications* — For applications where real-time processing isn't critical and a little latency is acceptable, the robust processing power of cloud computing may be more beneficial.

*Initial model development and training* — Building and training of complex AI models often require significant computational resources, making cloud platforms ideal choices.

*Backup and disaster recovery* — Cloud computing typically offers more robust options for data backup, redundancy, and disaster recovery.

*Frequent model updates* — If your AI model needs constant retraining or updates based on large-scale data, cloud infrastructure offers easier management and deployment.

## The best of both worlds — the benefits of a hybrid approach

The convergence of edge and cloud computing can lead to a hybrid approach that leverages the strengths of both paradigms.

For instance, say a large retail chain wants to implement a hybrid edge-cloud approach to its Smart Stores; it might use edge AI in multiple scenarios: real-time inventory tracking using computer vision, personalized product recommendations via in-store kiosks, or for immediate theft detection and alert systems.

This same retailer could then utilize cloud AI for model training and updating based on data from all stores, or conduct cross-store trend analysis to help drive its marketing strategies.

In a healthcare scenario, a large hospital could use edge AI for patient care and operational efficiency. Their use of edge AI can involve wearable devices for real-time patient monitoring or bedside monitors for anomaly detection. In a cloud scenario, they would leverage the cloud for complex medical image processing, for example.

Here is an overview of the key advantages of a hybrid approach:

*Optimized performance* — Utilize edge for real-time processing and cloud for complex, resource-intensive tasks. This allows companies to achieve low latency for critical operations while maintaining access to cloud resources.

*Scalability and flexibility* — Scale cloud resources for big data analytics and machine learning, and deploy edge devices for localized processing, adding or removing as needed.

*Enhanced reliability* — Maintain basic functionality at the edge, even during network outages, and use the cloud for redundancy and backup, ensuring data integrity and system resilience.

*Comprehensive data management* — Process sensitive data at the edge for privacy, then utilize cloud for long-term storage.

*Energy efficiency* — Optimize power consumption by distributing processing between the edge and cloud, while reducing the energy footprint of data centers by offloading suitable tasks to the edge.

By adopting a hybrid edge-cloud approach, companies can create more responsive, efficient, and scalable AI systems. This strategy allows them to process data where it makes the most sense, balancing the need for real-time insights with the power of comprehensive data analysis.

# THE RISE OF GEN AI AND FOUNDATIONAL MODELS

Foundational models and gen AI are unlocking new capabilities in the field, from zero-shot learning to data generation, reasoning, natural language understanding, and creative content generation.

## Foundation Models

- Pre-trained models
- Trained on broad datasets
- Applied to tasks outside their training
- Tend to be large! Hundreds of megabytes to terabytes

**Text • Audio • Image • Code  
Genomics • Time Series**

## Generative AI

- Create data in addition to consuming
- Can be implemented using foundation models
- Size can vary greatly depending on task

**Writing • Speech • Denoising  
Images • Code • Music**

These models, pre-trained on vast datasets, can be adapted for a variety of tasks without extensive retraining. While typically deployed in cloud environments due to their size and complexity, Edge Impulse is finding ways to [bring these models to the edge](#)<sup>11</sup>.

In the current landscape, it's common to use gen AI models earlier in the machine learning (ML) workflow, for example, for labeling data or to help train smaller models via model distillation. As for edge AI, bringing these powerful models to the edge is rare, with the exception of very specific use cases. However, the future is wide open as the evolution of more efficient hardware and optimized algorithms continue to evolve rapidly.

“We don’t need to wait for models like GPT to run on edge devices. There are already ways to harness the power of these foundational models without needing to deploy the full-scale versions at the edge.”

— DANIEL SITUNAYAKE, DIRECTOR OF MACHINE LEARNING, EDGE IMPULSE



Techniques such as knowledge distillation are central to this approach. This process transfers the knowledge of large models into smaller, more efficient ones, making edge deployment more feasible. Edge Impulse is leveraging gen AI to create synthetic data, which significantly lowers the cost and time of data collection for training AI models. This is especially important for applications in industrial settings where collecting real-world data might be difficult or expensive.

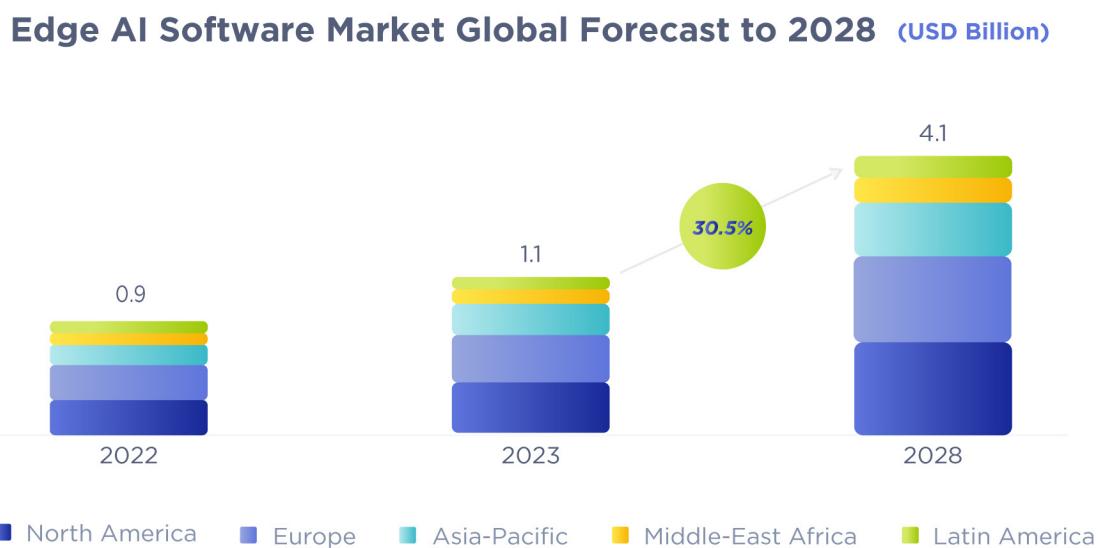
In the future, we may see more sophisticated, adaptive, and context-aware intelligence directly on edge devices.

# POWERING THE FUTURE WITH CUTTING-EDGE AI HARDWARE AND SOFTWARE

## Software — global market size and growth projections

Edge AI software solutions are becoming an integral part of modern digital infrastructures. This growing market encompasses a wide range of tools and platforms designed to enable AI inference and training on edge devices, from smartphones and IoT sensors to industrial equipment.

The global market for edge AI software is projected to grow from \$1.1 billion USD in 2023 to \$4.1 billion USD by 2028<sup>12</sup>. The growth is fueled by the increasing adoption of IoT devices, the need for real-time data processing, and the emergence of 5G networks, says MarketsandMarkets in their [Edge AI Software Market Report](#).



## Software platforms painting the future

**NVIDIA** — The Jetson platform developed for AI at the edge includes small, power-efficient production modules and developer kits that offer the most comprehensive AI software stack for high-performance acceleration to power generative AI at the edge.

**Edge Impulse** — The leading platform for building, deploying, and scaling edge machine learning models. The company empowers ML teams to run AI at peak performance on any edge hardware with unmatched ease and speed.

*Amazon Web Services* — AWS IoT Greengrass is an open-source edge runtime and cloud service for building, deploying, and managing device software. AWS IoT Greengrass makes it easy to bring intelligence to edge devices, including anomaly detection in precision agriculture or powering autonomous devices.

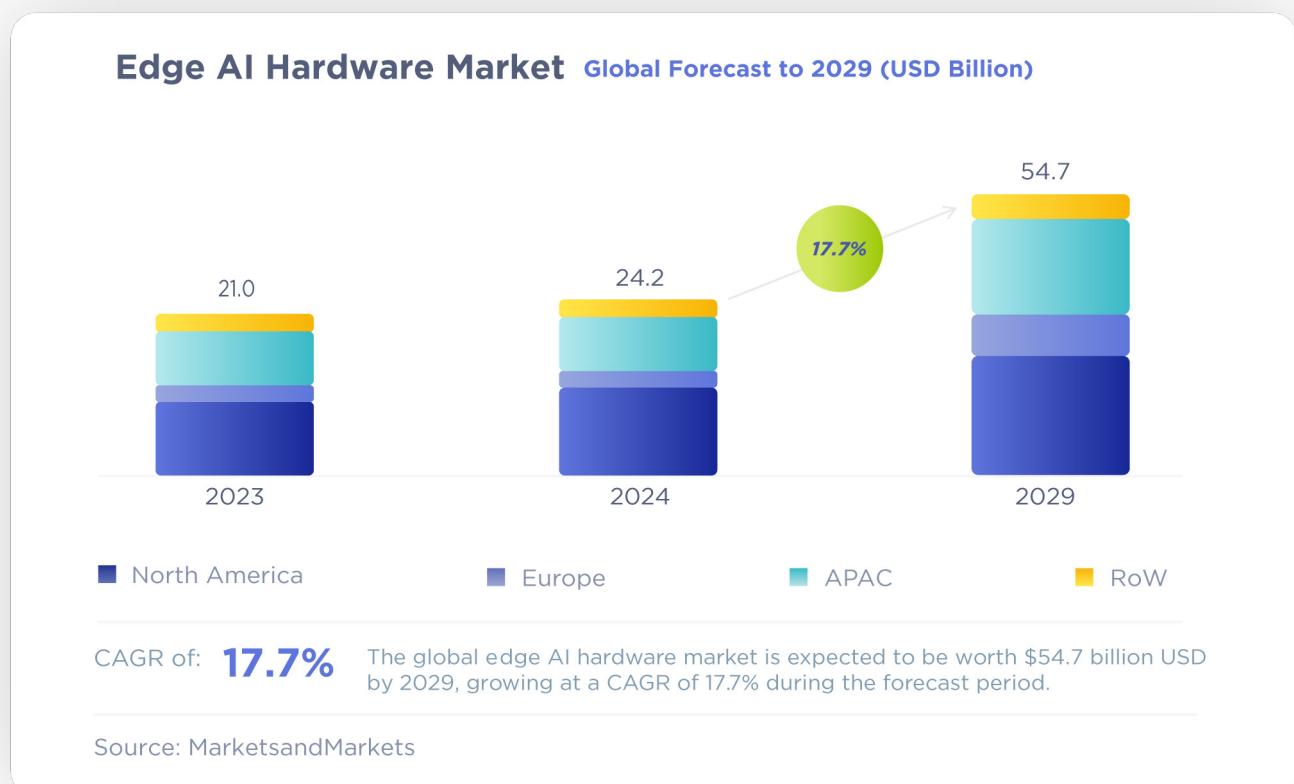
*Google Coral* — Google's edge AI offering, Coral, provides tools for building edge AI applications, including hardware, software, and a suite of tools designed to run machine learning models on edge devices.

Take a look at the [\*\*Top Ten Platforms for Developing Edge AI Applications\*\*](#)<sup>13</sup>.

## Hardware — market size and projections

Edge devices and processors form the backbone of edge AI implementations, enabling local processing and decision-making across a wide range of applications. These devices range from specialized AI accelerators and neural processing units (NPUs) to more general-purpose systems on chip (SoCs) optimized for edge computing. Major players like NVIDIA and Arm continue to innovate in this space, developing increasingly unique chipsets that are both task-optimized and energy-efficient edge AI chips and hardware accelerators.

This shift is driving rapid growth in the edge AI hardware market. According to [\*\*MarketsandMarkets\*\*](#), the global edge AI hardware market is expected to be worth \$54.7 billion USD by 2029<sup>14</sup>.



## Hardware solutions empowering AI at the edge

**Edge GPUs** — NVIDIA has developed AI-specific chips like the Jetson Orin series to run optimized models for edge devices. Production-ready devices from Advantech, Adlink, AAEON, Lanner, Vecow, Lexmark, Seeed, and many others use these NVIDIA solutions in designs that have the right specifications to operate in multiple spaces.

**SoCs with AI Capabilities** — Arm has its own Cortex-M and Cortex-A cores that feature capabilities to drive AI functions. Companies like STMicroelectronics, NXP, Infineon, Microchip, and many others continue to invest in this technology. Meanwhile, others like Silabs have added special cores, such as the MVP (Matrix Vector Processor), to run DSP capabilities. With the latest Arm announcement of the Cortex-M85, Renesas' RA8 family uses built-in Helium technology to deliver the highest scalar, DSP, and ML performance that Arm offers in its MCU cores.

**Edge AI Accelerators** — Devices from MemryX, Hailo, Deepx, Blaize, and Axelera are AI coprocessors that run models in an accelerated manner, thereby freeing the host processor from computationally heavy tasks. Many of these can run the latest vision models.

**Neural Processing Units (NPUs)** — These are accelerated coprocessors integrated on the same silicon as the host processor. Arm's Ethos IP can be easily added alongside Cortex-based IP devices for accelerated inferencing. The latest U85 is designed for gen AI at the edge with native support for transformer networks, along with support for Tensor Operator Set Architecture (TOSA) as a standard.

Alif's Ensemble and Himax's WiseEye2 both have Ethos-U55 integrated, while in the high-end market, NXP is using the latest Ethos-U65 in the I.MX93 to achieve better performance with a Linux-based host processor. Others, like Renesas's DRP-AI and TI's specialized DSPs, have their own dedicated AI accelerators integrated to offload AI tasks from the host processor.



# WHY MODEL EFFICIENCY IS CRUCIAL TO SUCCESSFUL EDGE AI DEPLOYMENT

At the heart of the edge AI revolution lies the critical challenge of model optimization — the art and science of refining AI models to run efficiently on resource-constrained devices without sacrificing accuracy. It serves as the linchpin between ambitious AI capabilities and the practical realities of the constraints of edge devices.

## Benefits of model optimization

By reducing model size and computational requirements, compression enables sophisticated AI models to run more efficiently on resource-limited hardware, dramatically expanding the potential applications of edge AI.

The benefits of model optimization are multifold:

- Reduced energy consumption, leading to longer battery life
- Decreased memory usage, allowing for more complex functionality on simpler devices
- Lower latency, enabling real-time responsiveness crucial for many edge AI applications

## How to achieve model optimization

Edge Impulse's market-leading [Edge Optimized Neural \(EON\) Compiler](#)<sup>15</sup>, produces models that run more efficiently on-device while having reduced RAM and flash usage. It supports a diverse range of neural networks from TensorFlow to PyTorch, as well as classical ML models from scikit-learn, LightGBM, and XGBoost. This ensures rapid execution of neural network models for real-time or near-real-time scenarios, all the while maintaining accuracy and:

- Up to 70% less RAM usage
- Up to 40% less flash usage
- Faster inference times
- Reduced overall power consumption

Companies that fail to prioritize model optimization may find themselves unable to deploy their AI solution on targeted devices, struggle with poor performance, or face prohibitively high costs for hardware upgrades. Moreover, unoptimized models can lead to slower time-to-market and missed opportunities.

# GETTING STARTED WITH EDGE AI

## Cultivating edge-centric thinking

Embarking on the edge AI journey requires a fundamental shift in mindset. As companies take their first steps into this transformative realm, adopting an “edge first” approach is crucial. This means prioritizing the value and experience delivered directly at the edge rather than treating edge capabilities as an afterthought.

By centering the edge in their strategic thinking, businesses can unlock the full potential of AI at the point of action. This edge-centric mindset is an important precursor to leveraging edge AI as a true business driver.

As we delve into the practical aspects of getting started with edge AI, remember that the most successful implementations begin with the pivotal shift in perspective — seeing the edge not as a peripheral consideration but as the new frontier of innovation and competitive advantage.

“Edge AI will be a highly transformative trend for the next decade, with proliferation of IoT endpoints and more pervasive network connectivity via 5G.”



GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.

# Assessing edge AI readiness

This edge AI maturity assessment checklist is designed as a starting point to help you evaluate your organization's preparedness for implementing edge AI solutions. It's not an exhaustive roadmap, but rather a guiding tool to identify key areas of focus and potential challenges. Use it to spark discussions and inform your edge AI strategy.

## Edge AI Maturity Assessment Checklist

### 1. Strategic Alignment

- Clear business objectives for implementing edge AI identified
- Edge AI initiatives aligned with overall company strategy
- Key stakeholders across departments are engaged and supportive

### 2. Development Capabilities

- Internal AI/ML expertise assessed
- Skills gap analysis completed
- Training or hiring plan to address skill gaps in place
- Familiarity with edge AI development tools and frameworks

### 3. Use Case Definition

- Specific edge AI use cases prioritized and documented
- Return on Investment (ROI) for each use case estimated
- Success metrics and Key Performance Indicators (KPIs) defined for each use case

### 4. Technical Infrastructure

- Edge devices/hardware requirements specified
- Network capabilities assessed (e.g., connectivity, latency, bandwidth)
- Cloud-edge hybrid architecture designed (if applicable)
- Scalability needs identified and plans created for addressing them

### 5. Ethical AI Considerations

- Ethical implications of edge AI applications assessed
- Bias detection and mitigation strategies implemented
- Ensured AI decision-making processes are transparent and understandable
- Governance structure for ethical AI use established

## 6. Data Readiness

- Relevant data sources identified and accessible. Create a data collection program if starting from scratch
- Data quality and quantity metrics defined
- Data privacy and security requirements understood and addressed
- Plan for ongoing data collection and management in place

## 7. Model Development

- Experimental set of models and DSP algorithms selected or custom development plan in place
- Model optimization for edge deployment strategy defined
- Testing and validation procedures established
- Plan for model updates and maintenance created

## 8. Performance Optimization

- Benchmarking criteria for on-device performance established
- Power consumption and efficiency targets set
- Latency requirements for real-time applications defined
- Plan for continuous performance monitoring and improvement in place

## 9. Integration and Deployment

- Integration points with existing systems identified
- Deployment strategy (e.g., gradual rollout, pilot testing) defined
- Monitoring and logging systems for edge AI performance in place
- Model monitoring and update strategy defined
- Rollback and failover procedures established

## 10. Security and Compliance

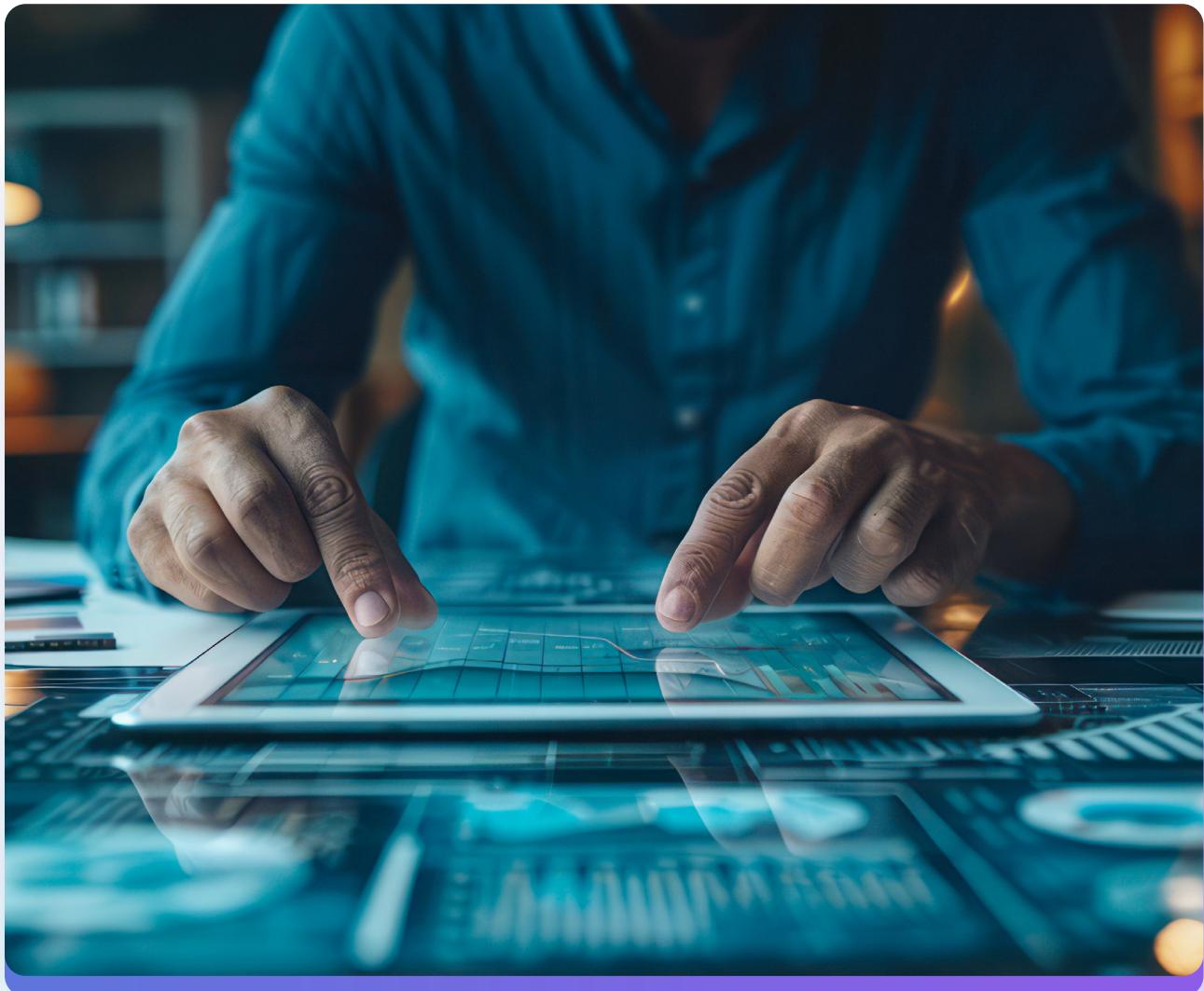
- Edge AI-specific security risks assessed
- Data protection measures for edge devices implemented
- Compliance with relevant regulations (e.g., GDPR, CCPA) ensured
- Regular security audit process established

## 11. Commercial Readiness

- Go-to-market strategy for edge AI-enabled products/services defined
- Pricing model for edge AI solutions determined
- Customer support processes for edge AI offerings established
- Partner ecosystem for deployment and scaling identified

## 12. Future-Proofing

- Roadmap for future edge AI capabilities and use cases created
- Flexibility for hardware/software upgrades built into the strategy
- Continuous learning and adaptation mechanisms for AI models planned
- Strategy for staying current with edge AI trends and advancements



## ROI Calculator – build vs. buy

Ready to quantify the potential impact of edge AI on your business? After completing the maturity assessment, take the next step with the [ROI calculator](#). This calculator is designed to compare the costs and benefits of developing machine learning (ML) projects using a do-it-yourself (DIY) approach versus utilizing Edge Impulse's platform.

It takes into account various factors, including team composition, project complexity, and the nature of the data and sensors involved in your project.

Product • Solutions • Developers • Pricing • Company • Blog

## ROI Calculator

Easily calculate the ROI of using Edge Impulse versus traditional in-house development for your edge machine learning platform. See how Edge Impulse can save you time and money, helping you make an informed decision for your next project.

**How many engineers are on your team?**

1      2-5      5-10      10

**What is the mix of your team?**

Mostly ML  Mostly Infra  Mostly product  A mix of both  Just me

**How new is your team to Edge ML?**

New      Somewhat experienced      Experienced      Mature

**How complex is your dataset?**

Small & Structured      Medium & Semi-structured      Large & Unstructured      Unknown

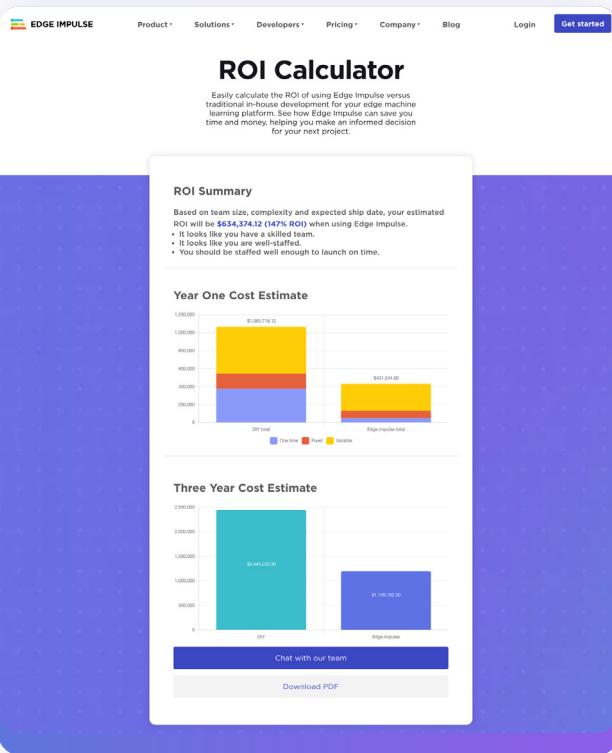
**How complex are your sensors?**

Simple      Moderate      Advanced      Complex

**When would you like to ship your product?**

Yesterday  Within 12 months  Within 2 years  No rush

**Submit**



Put the [ROI calculator](#) to the test.

## Developing an edge AI roadmap

Embarking on the edge AI journey also requires a well-crafted roadmap that aligns with your organization's strategic goals and operational realities. Developing the roadmap is a critical step in navigating the complex landscape of distributed intelligence, ensuring that your investment in the technology yields tangible results. Below, we'll explore the key components of an effective edge AI roadmap.

### Define Vision and Objectives

- Articulate long-term edge AI vision
- Align with overall business strategy
- Set clear, measurable objectives

### Assess Current State

- Evaluate existing infrastructure and capabilities

### Identify skill gaps and resource needs

- Analyze current data landscape and quality

### Identify Key Milestones

- Define short-term (6-12 months) goals
- Outline medium-term (1-3 years) objectives
- Envision long-term (3-5 years) aspirations

## Prioritize Use Cases

- Select initial pilot projects
- Plan for scaling successful pilots
- Outline the timeline for implementing various use cases

## Technology Stack Planning

- Choose appropriate edge devices and sensors
- Select AI frameworks and development tools
- Plan for integration with existing systems

## Data Strategy

- Develop data collection and management plan
- Address data privacy and security concerns
- Plan for data governance and quality assurance



## Skill Development and Team Building

- Outline hiring and training strategies
- Plan for continuous learning and skill upgrades
- Consider partnerships and external expertise

## Budget and Resource Allocation

- Estimate costs for each phase of implementation
- Allocate resources for different initiatives
- Plan for ongoing operational expenses

## Risk Assessment and Mitigation

- Identify potential technical and operational risks
- Develop contingency plans
- Consider regulatory and compliance challenges

## Performance Metrics and KPIs

- Define success criteria for each phase
- Establish monitoring and evaluation processes
- Plan for regular reviews and adjustments

## Stakeholder Communication Plan

- Develop strategies for keeping stakeholders informed
- Plan for regular progress updates and demonstrations
- Prepare for managing change across the organization

## Scalability and Future-Proofing

- Plan for scaling successful implementations
- Consider potential pivots based on technological advancements
- Align roadmap with emerging edge AI trends

# IN-HOUSE DEVELOPMENT OR AN END-TO-END SOLUTION?

When it comes to implementing edge AI solutions, organizations often face a critical decision: embark on a DIY journey or adopt an end-to-end platform. In the early days of edge AI, the DIY approach was often the only option, requiring companies to assemble a complex puzzle of components, tools, and resources. This meant building a team of specialized talent:

## Core Team Composition

- AI researchers
- Data scientists/ML engineers for model development and optimization
- Embedded engineers to handle hardware and deployment challenges
- IoT engineers for sensor integration and data collection
- Software developers for application development and integration
- DevOps engineers for continuous integration and deployment
- Data engineers for data pipeline and infrastructure management



## Supporting Roles

- Project managers to oversee edge AI initiatives
- UX/UI designers for creating intuitive interfaces for edge AI applications
- Security specialists to ensure data protection and device security
- Legal/compliance experts to navigate regulatory landscapes
- Domain experts to provide industry-specific insights

Companies also needed to navigate the intricacies of hardware selection, model optimization, and system integration, often leading to lengthy development cycles and unforeseen challenges. While DIY can offer maximum customization, the landscape has evolved.

DIY is a challenging approach that demands extensive domain expertise to be feasible. Developing and maintaining customized solutions often require significant time and resources, which is impractical for most organizations. Additionally, lacking specialized knowledge can lead to suboptimal outcomes and potential setbacks.

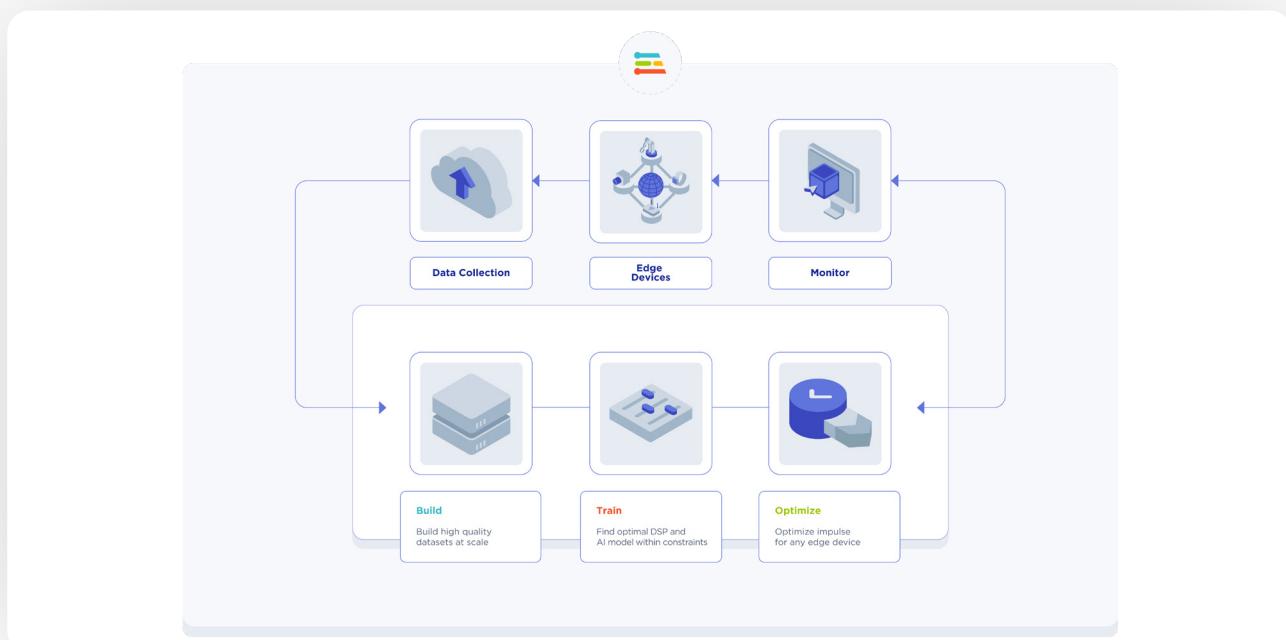
## End-to-end solutions — the power of leveraging MLOps

On the other hand, end-to-end edge AI platforms offer a more streamlined path to implementation. These offerings are designed to assist developers with the entire process of developing an application:

- Collecting, managing, and exploring datasets
- Performing feature engineering and digital signal processing
- Training machine learning models, optimizing algorithms for embedded hardware, deploying to embedded systems, and evaluating systems' performance on real-world data.

### Here's a snapshot of the Edge Impulse platform for comparison to DIY:

- **Cost Efficiency and Predictability** — Reduce total cost of ownership by eliminating the need for in-house development and maintenance teams, leading to more predictable budgeting.
- **Faster Time to Market** — Deploy solutions quickly without the delays inherent in developing and maintaining a system from scratch.
- **Scalability** — Easily scale up or down based on your needs without worrying about infrastructure limitations or additional development.
- **Reliability and Uptime** — Benefit from a solution that has been thoroughly tested and proven in various environments to ensure maximum reliability.
- **Integration Capabilities** — Seamlessly integrate with existing systems and platforms, reducing the complexity and cost of adoption.



**Edge Impulse end-to-end platform**

- **New Features, Regular Updates, and Maintenance** — Receive ongoing software updates, patches, and new features without additional effort on your part.
- **Focus on Core Competencies** — Allow your team to concentrate on developing your primary business applications, rather than diverting precious resources to develop and maintain AI tooling.
- **Enhanced Security Measures** — Benefit from advanced security features, such as SOC2 compliance and customer data oversight.
- **Intellectual Property Assurance** — Avoid potential licensing pitfalls — including GPL compliance — and other IP issues that can arise from developing new technology internally.
- **Optimization Features** — Powerful tools like EON Tuner and EON Compiler allow for hardware-aware optimization of algorithms and one-model deployment to any device. Model metrics available within the platform provide detailed insights on model accuracy, resource usage, and inference speed.

Use the [\*\*ROI calculator\*\*](#) to help you estimate cost savings and efficiency gains.

For organizations looking to quickly capitalize on edge AI benefits without the overhead of building everything from scratch, an end-to-end solution is an attractive option. Such solutions often come with robust support and regular updates, ensuring that the systems remain current with the latest technological advancements.

# CHALLENGES AND BARRIERS TO EDGE AI ADOPTION

While the potential of edge AI is immense, the path to widespread adoption is not without hurdles. Organizations may face a complex landscape of challenges that span technical, operational, and strategic areas. Here are a few challenges to consider.

## Resource constraints

- [Budget limitations](#) — balancing cost and innovation
- [Talent scarcity](#) — addressing the edge AI skills gap
- [Hardware constraints](#) — optimizing for limited computing power

Ultimately, striking the right balance between model complexity, accuracy, and resource utilization is critical to pushing the limits of edge devices.

However, there are approaches available to developers looking to minimize hardware costs and enhance the feasibility of [\*\*deploying advanced AI in resource-constrained environments\*\*](#).

## Data privacy and security concerns

As intelligent systems move closer to the edge, the point of data collection, they handle vast amounts of personal, operational, and proprietary information. We live in an era where data [\*\*breaches and privacy violations\*\*](#)<sup>16</sup> dominate the headlines. Although edge AI offers inherent advantages in data protection by processing information locally, it also introduces new and complex security challenges that organizations must carefully navigate.

For product innovation teams, these privacy and security requirements can impact feature development and user experience. Balancing the need for data protection with the desire for seamless, user-friendly interactions often requires careful trade-offs.



## Navigating the complex regulatory compliance landscape

Organizations must ensure their edge AI implementations adhere to a complex web of data protection laws. Then, there are industry-specific mandates, such as finance, healthcare, and more, that are becoming increasingly stringent. It's also worth noting that the compliance landscape can vary dramatically across different regions and sectors, presenting a daunting challenge for those that operate globally.

To help navigate the edge AI compliance, companies can:

**Implement privacy by design principles** — Implement privacy considerations from the outset of product development and conduct privacy assessments for new edge AI implementations.

**Develop a comprehensive data governance framework** — Create clear policies for data collection, processing, storage, and deletion.

**Utilize advanced encryption and security measures** — Implement end-to-end encryption on all data in transit and at rest, and use secure enclaves or trusted execution environments for sensitive processing.

There are pros and cons, risks, and challenges to implementing edge AI. Here are five top risks to watch out for.

## Top 5 risks to consider when implementing an edge AI project:

### 1. Suitability

- The task you want to perform may simply be impossible, so it's crucial to disprove this as early as possible in the R&D process. Is AI, or edge AI, the right solution for this problem, or would another technology be more effective?
- Can your application perform adequately in real-life scenarios?
- Is it possible to validate your application sufficiently to gain the confidence needed for deployment? Do the safety margins of your use case allow for the use of AI, knowing that it will always involve some degree of error?
- Is the idea commercially viable, for example, given the hardware requirements?

### 2. Dataset

- It can be too difficult or expensive to obtain an adequate dataset to achieve the required real-world performance. Data also expires, and models need to be updated over time.

### 3. Constraints

- Is the hardware you're required to use capable of running the algorithms you need?
- Are there communication constraints that cause a problem (bandwidth, latency, etc.)?
- Device heterogeneity — can you keep up your product scales as the market evolves?

## 4. Organization and infrastructure

- Is your organization set up to support an edge AI project throughout a potentially long lifecycle?
- Is it possible to safely and effectively integrate with existing systems and infrastructure?

## 5. Expertise

- Does your team have the required skills to execute the project from end to end?

“AI has revolutionized industrial workflows, particularly in machine vision, with increasing accuracy in AI inferencing and machine learning. But one of the biggest hurdles with implementing AI/ML on the edge is the complexity of the infrastructure. Edge Impulse solves this problem by integrating their software with edge devices (like the ICAM-540 Advantech Industrial AI Camera) to create an ecosystem that bridges the gap between the cloud and edge.”

— MELKY REBOLLEDO, ASSOCIATE PRODUCT MANAGER, ADVANTECH



# THE TRANSFORMATIVE POWER OF EDGE AI

When considering the business case for edge AI, every organization, based on business objectives and available resources, including AI/ML expertise, must decide whether the benefits far outweigh the challenges. Industry data and insights support the thesis that edge AI is a game changer and one of the key technologies that will continue to shape the future of innovation.

## The economics of edge AI

- Cost savings beyond the initial investment
- New revenue streams
- Operational efficiency
- Reduces cloud computing and data transfer costs, especially for applications processing large volumes of data

## Energy efficiency

- Consumes less power by minimizing data transmission, extending the battery life of mobile and IoT devices
- Scalable energy savings — as edge AI is deployed across millions of devices, the cumulative energy savings compared to cloud-centric approaches can be substantial
- Sustainability — green IoT applications such as smart building systems can optimize energy use in real time

## Reliability in action

- Unwavering performance — continues to function even with poor or no internet connectivity, ensuring uninterrupted operation in critical applications
- Always-on intelligence
- Predictive maintenance

## Smarter, more responsive systems

- Real-time decision-making (insight to action in milliseconds)
- Contextual awareness
- Adaptive learning

## Privacy and security

- Safeguards data at the source
- Data minimization by processing data where it matters
- Encryption at the edge

## Sector-specific impact

- Manufacturing — intelligent factories
- Healthcare — personalized medicine at the point of care
- Retail — reimagining the shopping experience
- Smart cities — urban environments that adapt and respond



# THE FUTURE OF EDGE AI

As we look forward, it's crucial to understand emerging trends, anticipate industries' impact, and strategize on how to leverage edge AI for a competitive advantage and long-term growth.

## Emerging technology trends in edge AI

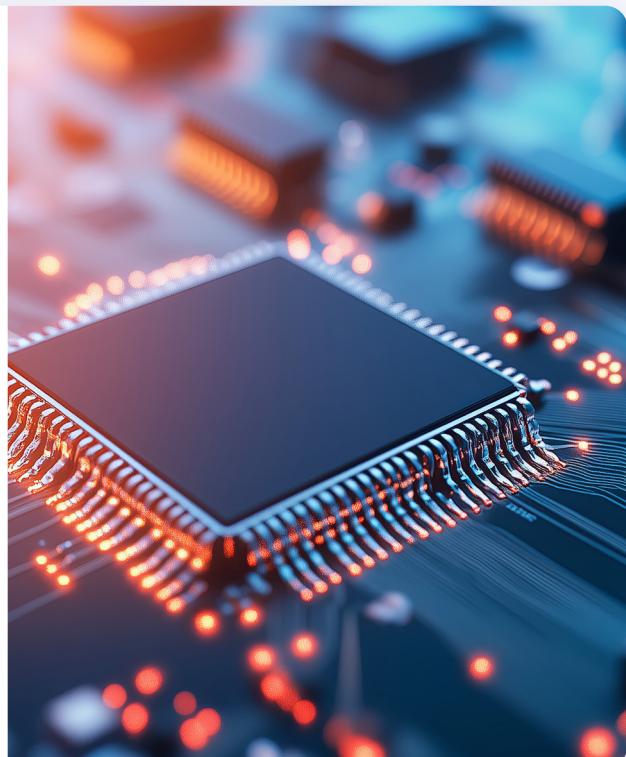
**AI-optimized Hardware** — development of specialized chips (NPUs, TPUs) designed for efficient AI processing at the edge.

**Federated Learning** — Enables collaborative model training across distributed devices without centralizing data and addresses privacy concerns while enabling more personalized AI.

**Tiny ML** — Development of ultra-lightweight AI models capable of running on severely constrained devices, thereby bringing AI capabilities to a wide range of small and low-power devices.

“Neuromorphic processing for edge AI is inspired by the brain, focusing on computing relevant events — a leap forward in efficiency and adaptability. When integrated with platforms like Edge Impulse, it puts the power of event-based computing into the hands of innovators everywhere.”

— STEVEN BRIGHTFIELD, CHIEF MARKETING OFFICER, BRAINCHIP



**Neuromorphic Computing** — Development of brain-inspired computing architectures for more efficient AI processing and the potential to dramatically reduce power consumption in edge AI systems. BrainChip's first-to-market, digital neuromorphic processor IP, Akida, keeps AI/ML local to the chip and independent of the cloud, dramatically reducing latency while improving privacy and data security.

## Maximize the edge AI Advantage:

1. Develop an edge AI strategy — Create a comprehensive roadmap that aligns edge AI initiatives with your overall business objectives and digital transformation efforts.
2. Invest in skills and talent — Build internal expertise in edge AI technologies through training programs and strategic hiring to drive innovation from within.
3. Prioritize data strategy — Implement robust data collection, management, and governance practices to ensure high-quality data for edge AI models.
4. Focus on user-centric designs — Ensure that edge AI implementations prioritize user experience and deliver tangible value to end-users.
5. Security — Implement robust security measures specifically designed for edge AI systems to protect against emerging threats.
6. Measure and communicate value — Develop clear metrics to measure the impact of edge AI initiatives and effectively communicate their value to stakeholders.
7. Continuous learning and adaptation — Stay agile and be prepared to pivot strategies as the edge AI landscape evolves, continuously evaluating new technologies and use cases.

By embracing the right strategy and staying ahead of emerging trends, you can ready your organization to not only ride the wave of edge AI innovation but to shape its future. Organizations that successfully integrate edge AI into their core operations and product offerings will be well-positioned to drive growth, enhance customer experiences, and maintain a competitive edge in an increasingly AI-driven world.

## Footnotes:

- <sup>1</sup> [Top 10 Deep Learning Algorithms You Should Know in 2024.](#)
- <sup>2</sup> [Generative AI to Become a \\$1.3 Trillion Market by 2032, Research Finds.](#)
- <sup>3</sup> [Global Edge AI Market, Market.us.](#)
- <sup>4</sup> [Smart Manufacturing Market Size, Share & Industry Analysis, Fortune Business Insights.](#)
- <sup>5</sup> [Smart Urban Living, Smart Nation, Singapore.](#)
- <sup>6</sup> [What Companies Can Do About Cloud Spend Wastage, Forbes.](#)
- <sup>7</sup> [Maintenance Costs and Advanced Maintenance Techniques in Manufacturing Machinery: Survey and Analysis, National Library of Medicine.](#)
- <sup>8</sup> [Unlocking Performance: How Manufacturers Can Achieve Top Quartile Performance, Wall Street Journal Partners.](#)
- <sup>9</sup> [Prevent Heat Exhaustion Case Study, Edge Impulse.](#)
- <sup>10</sup> [2.9 Billion People Still Offline, International Telecommunication Union \(ITU\).](#)
- <sup>11</sup> [Roundup: Edge Impulse's Implementations of LLM Tools and Techniques for Edge AI.](#)
- <sup>12</sup> [Global Edge AI Hardware Market Report, Markets and Markets.](#)
- <sup>13</sup> [Top 10 Platforms for Developing Edge AI Applications, Analytics Insight.](#)
- <sup>14</sup> [Global Edge AI Hardware Market Report, Markets and Markets.](#)
- <sup>15</sup> [Edge Optimized Neural \(EON\) Compiler.](#)
- <sup>16</sup> [Biggest Data Breaches in U.S. History, Upguard Blog.](#)

# The future of edge AI is here. Get started and see what's possible.

[Learn more](#)

## About Edge Impulse

Edge Impulse streamlines the creation of AI and machine learning models for edge hardware, allowing devices to make decisions and offer insight where data is gathered. Edge Impulse's technology empowers developers to bring more AI products to market, and helps enterprise teams rapidly develop production-ready solutions in weeks instead of years. Powerful automations make it easier to build valuable datasets and develop advanced AI for edge devices from MCUs to CPUs to GPUs. Used by health and wearable organizations like Ultrahuman, industrial organizations like Halma, as well as top silicon vendors and over 160,000 developers, Edge Impulse has become the trusted ML platform for enterprises and developers alike.



**EDGE IMPULSE**

[edgeimpulse.com](http://edgeimpulse.com)