



EDGE IMPULSE



Building the Business Case for Edge AI

The growth and innovation advantage
you can't afford to ignore

BUILDING THE BUSINESS CASE FOR EDGE AI

The growth and innovation advantage you can't afford to ignore

Introduction

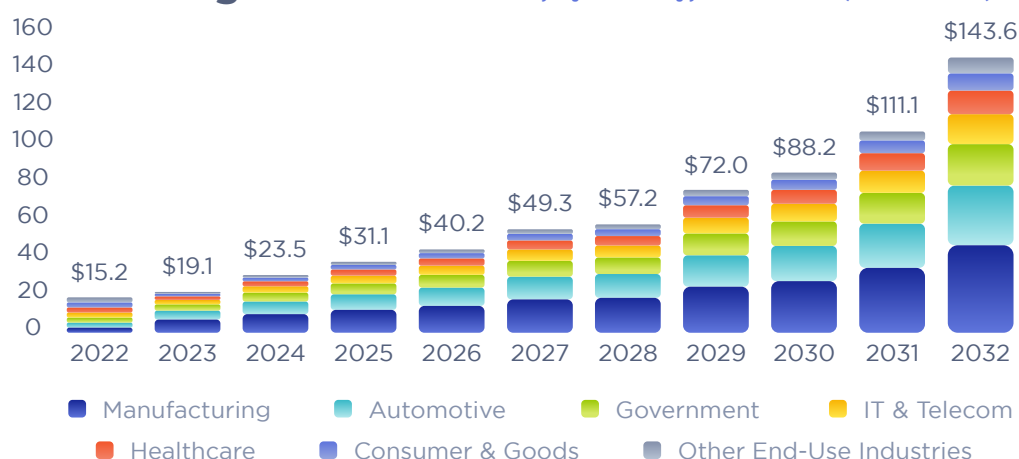
Organizations across almost every industry are leveraging the power of artificial intelligence (AI) to accelerate growth, innovation, and efficiency. When combined with the convergence of widespread IoT device adoption and edge computing, we arrive at edge AI, which may be an even bigger transformative force for businesses.

“Edge AI may be the most important development for the future of business.”

— FORBES MAGAZINE

What is edge AI and why is it such a game changer for businesses? Edge AI is the process of running artificial intelligence (AI) and machine learning (ML) algorithms on devices at the network's edge (periphery), rather than on large cloud servers. In other words, edge AI is the fusion of devices and artificial intelligence, enabling a wide range of possibilities to solve unique problems.

Global Edge AI Market Size, by industry, 2022-2023 (USD Billion)



The Market Will Grow at the CAGR of: **25.9%**

The Forecasted Market Size for 2032 in USD: **\$143.6Bn**



1. Forbes, The AI No One is Talking About: www.forbes.com/sites/forbestechcoucil/2023/10/11/edge-ai-the-ai-no-one-is-talking-about/?sh=6aa74189bfec&utm_content=287777028&utm_medium=social&utm_source=linkedin&hss_channel=lcpl-19239958

WHY EDGE AI IS TRANSFORMING BUSINESS

Edge AI adoption is accelerating across various industries, driven by the need for localized processing, improved data privacy, and enhanced operational efficiency. With its capacity to deliver flexibility, agility, and speed, edge AI is a disruptor unlike any other.

Moreover, edge AI presents distinct advantages over traditional centralized approaches. Unlike traditional AI architectures, edge AI brings intelligence to the source of data generation, typically on edge devices or computing nodes.

Imagine for instance, scenarios involving first responders or industrial workers facing life-threatening situations where they can't be reached. Under normal circumstances, cloud-based systems can monitor vital signs but if for any reason these workers “go dark” and are inaccessible via the cloud, that's where real-time decision making powered by edge AI has a huge impact.



Without the need for connectivity, algorithms running directly on existing hardware can provide instantaneous feedback on the risk of heat exhaustion in our example. SlateSafety, a company that helps leaders identify and prevent heat stress and lone worker injuries before they occur, pondered whether their award-winning BAND V2 could deliver real-time alerts through fully on-device processing. Through a collaboration with Edge Impulse, they honed in on key parameters and created an accurate, compact algorithm for **predicting heat exhaustion**.

Consider real-time decision making as one of the superpowers of edge AI—a critical capability for applications such as autonomous vehicles, industrial automation, and smart cities where split-second decisions have life and death implications.

Adaptability

The versatility and adaptability inherent in edge AI make it perfectly suited for a myriad of use cases and environments, allowing organizations to deploy and tailor to meet evolving business needs. By fine-tuning algorithms and models to address specific requirements and objectives, customization empowers businesses to extract maximum benefits from edge AI.

Seamless integration with existing infrastructure

Whether you're leveraging systems, IoT devices, or cloud platforms, edge AI complements and enhances these technologies without requiring significant modifications, thus minimizing disruptions to operations. This way, you can leverage your existing investments while tapping into the advantages edge AI offers.

Cost reduction

Edge AI has forever transformed the way data is processed and analyzed, presenting significant advantages in terms of costs and resource optimization. By leveraging computing power at the edge of the network, the expense associated with sending large volumes of data to the cloud for processing is drastically reduced. This results in minimizing bandwidth constraints and latency. An added benefit stems from edge AI's non-reliance on centralized cloud resources, which can be costly to maintain and scale.

Privacy and security

Whenever the topic of data is broached, concerns regarding privacy and security inevitably follow. Data breaches pose significant risks to individuals and organizations alike, jeopardizing sensitive information and, in the case of companies, undermining customers' trust. Unfortunately, anyone and any business can be vulnerable to a data breach or cyber attack.

Because traditional AI architectures often centralize the processing of data in cloud environments, transmitting potentially sensitive information over networks where it may be susceptible to hackers or others with unauthorized access is inescapable.

In one of the largest² data breaches in U.S. history, JP Morgan Chase disclosed that cyber attacks compromised over 76 million households and 7 million small businesses. Similarly, over 30,000 U.S. companies, (60,000 worldwide) were affected by an attack on the Microsoft Exchange email servers, one of the largest email servers in the world.

Edge AI offers a compelling alternative to traditional architectures. By performing AI inference and analytics at the edge, businesses can minimize the exposure of sensitive data to external threats, reducing the likelihood of data breaches.

The decentralized nature of edge AI empowers businesses to gain better control over their data and implement robust security measures tailored to specific use cases. Furthermore, edge AI solutions can capitalize on advanced encryption techniques and anomaly detection algorithms to safeguard data integrity and confidentiality.

“AI has shifted from simply being a cost-saving lever to a revenue driver, with 69% of respondents now using AI/ML to create new revenue streams.”

—— S&P 2023 GLOBAL TRENDS IN AI REPORT

² Biggest Data Breaches in U.S. History (Updated 2024) Upguard.com: www.upguard.com/blog/biggest-data-breaches-us

Less bandwidth required

By running algorithms on-device, you avoid sending sensor data to the cloud, reducing the level of connectivity required.

Faster response times

On-device algorithms can be much more responsive, since there's no round-trip to the cloud required, leading to lower latency.

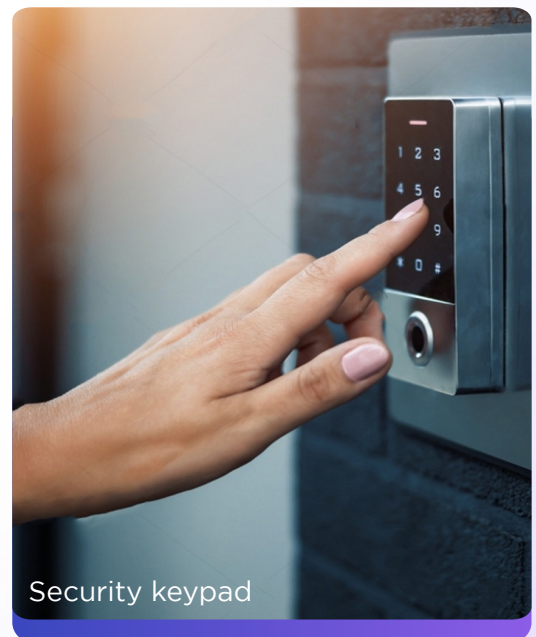
Increased reliability

By removing the reliance on connectivity and communication, on-device AI systems are potentially more reliable than those in the cloud, which is essential for applications where safety and user experience are critical.

Rapid growth and opportunities

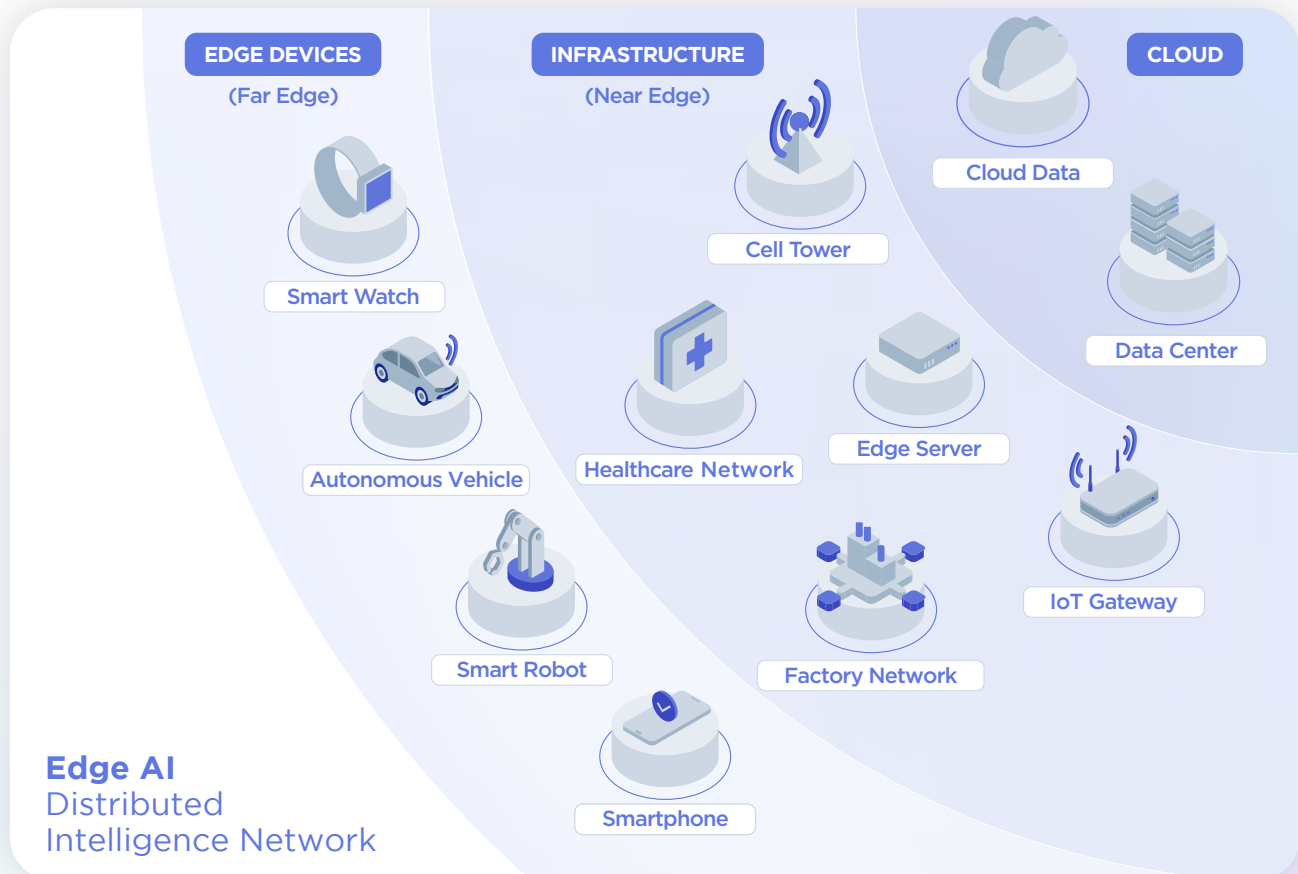
The proliferation of IoT devices is fueling edge AI growth.

Examples of edge AI systems



HOW DOES EDGE AI WORK?

The edge isn't a single destination; it's more like a broad region. Devices at the edge of the network can communicate with each other, and also communicate with remote servers. There are even servers that live at the edge of the network.



In simple terms, edge AI is a way of making everyday devices smarter by giving them the ability to think and make decisions on their own, without needing to connect to the internet. For example, picture a security camera that can spot unusual activity and alert you, all by itself, without sending video to a distant server. The “edge” means the thinking happens right there on the device. This makes it faster and more private since your information doesn't need to travel across the internet.

BEYOND THE CLOUD: HOW EDGE COMPUTING ENABLES EDGE AI

Edge computing—a computer networking strategy where data is processed and stored at the periphery of the network—plays a pivotal role in facilitating the deployment of AI at the edge. IDC predicts³ worldwide spending on edge computing is expected to be \$232 billion in 2024, an increase of 15.4% over 2023. Combined enterprise and service provider spending across hardware, software, professional services, and provisioned services for edge solutions will sustain strong growth through 2027 when spending will reach nearly \$350 billion.

Anything that runs locally on your computer or smartphone is considered edge computing. That includes word processing, spreadsheets, most programming development environments, and many video games. The following highlights from Accenture’s research report, “Leading with Edge: How to Reinvent with Data and AI,”⁴ make the case for why edge computing is crucial to edge AI and the future of business:

- Edge is set to become a ubiquitous lever of scale and reinvention—especially as AI becomes more pervasive in the enterprise.
- 83% of research respondents believe that edge computing will be essential to remaining competitive in the future.
- Of the four adopter types, the most advanced edge adopters are 4x more innovative, 9x more efficient, and nearly 7x more cost-effective.



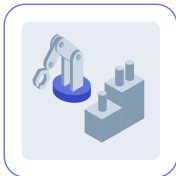
³ New IDC Spending Guide Forecasts Edge Computing Investments Will Reach \$232 Billion in 2024: www.idc.com/getdoc.jsp?containerId=prUS51960324

⁴ Leading with edge: How to reinvent with data and AI: www.accenture.com/us-en/insights/cloud/edge-computing

EDGE AI IN ACTION - PRACTICAL APPLICATIONS

Edge AI has emerged as a transformative force and many industries are experiencing its profound impact. In the manufacturing sector, predictive maintenance has quickly become one of the top use cases, where systems powered by edge AI monitor the health of industrial machinery and predict potential failures before they occur.

Top applications and use cases for Edge AI include:



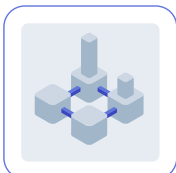
Industrial & Manufacturing

Enhances predictive maintenance, quality control, and process optimization in manufacturing plants



Healthcare

Facilitates remote patient monitoring, medical imaging analysis, and personalized treatment recommendations



Smart cities

Enables intelligent traffic management, public safety monitoring, and safer roads



Retail

Customizes behavior analysis, inventory, management, and personalized shopping experiences

Anomaly detection use case

In applications such as industrial inspection, medical imaging, and logistics, anomaly detection is critical. Training **visual anomaly detection models** involves developing algorithms to identify unusual patterns or anomalies in image data that do not conform to the expected behavior. With edge AI, organizations now have access to visual anomaly detection for production-ready edge AI pipelines.

Health and wellness use case

Edge AI is also making a tangible difference in the healthcare industry where technology is becoming increasingly vital to enhancing patient care and driving better clinical outcomes. AI systems can analyze vast amounts of medical data, including images, genetic information, and patient records, to assist healthcare professionals in diagnosing diseases and predicting outcomes with unprecedented accuracy.

In many instances, healthcare interactions speak to symptoms, not the cause. The founders of Hyfe, a respiratory wellness company that tracks and monitors coughs, needed a solution for streamlining their ML models for full on-device processing. The founders felt that there could be valuable, actionable insights stored in the sound of our coughs, so they set out to find a system that could unlock those learnings to help **address illness more directly**.

Energy use case

Electric power grids consist of complex interconnections that sustain our modern way of life. Even a minor infrastructure failure can lead to catastrophic consequences, evidenced by events like the California wildfires, which have caused numerous human casualties and extensive property damage. With an increase in energy demand, it's more important than ever to preemptively predict potential outages and failures. With edge AI, you can monitor what's happening with the power grid in real time. Discover how edge AI built one of the world's **most advanced grid monitoring systems**.



CHALLENGES AND LIMITATIONS OF EDGE AI

Implementing edge AI comes with challenges that organizations must navigate to unlock its full potential.

Lack of industry standards — The absence of standardization, especially in hardware, exacerbates compatibility issues according to [Wevolver's State of Edge AI 2024](#) report. Varying computing capabilities across different edge devices makes it difficult for developers striving to create universally compatible edge AI applications. The report suggests establishing industry-wide hardware standards is essential to facilitate seamless integration and scalability of edge AI solutions across diverse environments and devices. However, there is some good news with certain [software standards](#).

Data management — Edge AI involves managing and processing large amounts of data at edge devices rather than the cloud, which presents challenges around data and storage. According to the S & P 2023 State of AI report, data management is the top technical inhibitor to AI adoption.

Resource constraints — In general, edge devices offer fewer computational resources than their cloud-based counterparts. For example, edge devices like sensors, cameras, and IoT devices are typically designed for compact and mobile or remote deployment, which imposes strict limitations on their processing power, memory, capacity, and energy efficiency.

On the flip side, cloud servers can offer powerful processors and large amounts of memory. If your machine learning (ML) model cannot be optimized or constrained to run on an edge device, you should consider a cloud-based solution.

Unlike powerful cloud servers, these devices must operate within stringent hardware constraints, often relying on low-power processors, minimal RAM and battery or energy harvesting power sources. Ultimately, striking the right balance between model complexity, accuracy, and [resource utilization](#) is critical to pushing the limits of edge devices.

Limited remote access — Prediction serving from the cloud offers easy access from any device with internet access. Remotely accessed edge devices often require special network configuration, such as running a [VPN service](#).

Scaling — Scaling prediction services of cloud models typically requires cloning your servers and increasing computing power spend with your service providers. With edge computing, you need to purchase and configure additional hardware.

Data security and privacy concerns — As intelligent systems move closer to the edge, the point of data collection, they handle vast amounts of personal, operational and proprietary information, increasing the risk of exposure. Stringent security measures must be implemented.

Embedded.com suggests the following steps to [mitigate these risks](#): encryption of data both at rest and in transit, regular security audits, and regular software updates to address vulnerabilities. Additionally, techniques, such as federated learning and differential privacy (a mathematical technique for publicly sharing information about a dataset without including information about individuals in the dataset), can be used to ensure that sensitive data is protected while still enabling the benefits of AI.

The right skill sets and talent — Building edge AI systems requires a unique blend of interdisciplinary skills, which can make it challenging for some businesses. Proficiency in AI, machine learning, edge computing, and data analytics is critical. However, expertise in embedded systems, hardware, and software design is also needed to tackle the stringent constraints of edge devices.

Fortunately, there are systems designed to [solve well-known edge AI challenges](#) so engineers and developers can save time and focus on their domain expertise.

Part VI

GETTING STARTED WITH EDGE AI

Start with clear objectives — Begin with a strategic approach by identifying high-value use cases that align with your business objectives and can leverage the strength of edge computing.

Level up data management strategies — Robust data management is essential, beginning with techniques for efficient data collection, processing, and storage at the edge, while maintaining high security and privacy standards.

Prioritize the development of edge architecture — this is needed to account for diverse hardware requirements, connectivity constraints, and scalability for edge deployments. Your architecture should incorporate mechanisms for secure over-the-air updates, device management, and real-time monitoring for seamless operation and maintenance.

Part VII

CONCLUSION

Edge AI is already revolutionizing business, particularly in how companies deliver value to their customers. As we enter an era of [unprecedented connectivity](#), the ability to extract, process, and act upon data generated from billions of devices and machines generating data is paramount to successfully harnessing the power of edge AI.

The future is here, and we'll continue to see businesses unlocking new revenue streams, drive innovation, get to market faster, and stay competitive in an increasingly data-driven world.

Embracing edge AI is no longer an option but a necessity for businesses seeking to thrive in the era of intelligent, connected systems.



EDGE IMPULSE

Build. Train. Optimize. AI for the Edge

Sign up for your **free enterprise trial** and learn how Edge Impulse can help you run AI at peak performance on any edge device with unmatched speed and ease.

[Sign Up](#)

[Contact Us](#)

Check out these additional resources for more tips and insights:

Solutions:

[**Transform Your Product Strategy with Edge AI**](#)

Blog:

[**Revolutionizing Smart Manufacturing with Edge AI**](#)