

# **Edge Computing for Industrial AIoT Applications**

---

**Ethan Chen**  
*Product Manager*

**Alicia Wang**  
*Product Manager*

**Angie Lee**  
*Product Marketing Manager*

## Abstract

*IIoT applications are generating more data than ever before. In many industrial applications, especially highly distributed systems located in remote areas, constantly sending large amounts of raw data to a central server might not be possible. To reduce latency, reduce data communication and storage costs, and increase network availability, businesses are moving AI and machine learning to the edge for real-time decision-making and actions in the field.*

*These cutting-edge applications that deploy AI capabilities on IoT infrastructures are called the "AIoT." Although you still need to train your AI models in the cloud, data collection and inferencing can be performed in the field by deploying trained AI models on edge computers. This white paper discusses how to choose the right edge computer for your industrial AIoT application and provides several case studies to help you get started.*

## Bringing AI to the IIoT

The advent of the Industrial Internet of Things (IIoT) has allowed a wide range of businesses to collect massive amounts of data from previously untapped sources and explore new avenues for improving productivity. By obtaining performance and environmental data from field equipment and machinery, organizations now have even more information at their disposal to make informed business decisions. Unfortunately, there is far too much IIoT data for humans to process alone so most of this information goes unanalyzed and unused<sup>1</sup>. Consequently, it is no wonder that businesses and industry experts are turning to artificial intelligence (AI) and machine learning (ML) solutions for IIoT applications to gain a holistic view and make smarter decisions more quickly.

---

<sup>1</sup> IEEE (2020). "IoT Data Needs Artificial Intelligence" <https://innovationatwork.ieee.org/iot-data-artificial-intelligence/>

---

Released on July 15, 2020

© 2020 Moxa Inc. All rights reserved.

Moxa is a leading provider of edge connectivity, industrial computing, and network infrastructure solutions for enabling connectivity for the Industrial Internet of Things (IIoT). With over 30 years of industry experience, Moxa has connected more than 65 million devices worldwide and has a distribution and service network that reaches customers in more than 80 countries. Moxa delivers lasting business value by empowering industries with reliable networks and sincere service. Information about Moxa's solutions is available at [www.moxa.com](http://www.moxa.com).

### How to contact Moxa

Tel: 1-714-528-6777  
Fax: 1-714-528-6778



## Most IIoT Data Goes Unanalyzed

The staggering number of industrial devices being connected to the Internet continues to grow year after year and is expected to reach 41.6 billion endpoints in 2025<sup>2</sup>. What's even more mind-boggling is how much data each device produces. In fact, manually analyzing the information generated by all the sensors on a manufacturing assembly line could take a lifetime<sup>3</sup>. It's no wonder that "less than half of an organization's structured data is actively used in making decisions—and less than 1% of its unstructured data is analyzed or used at all"<sup>4</sup>. In the case of IP cameras, only 10% of the nearly 1.6 exabytes of video data generated each day gets analyzed<sup>5</sup>. These figures indicate a staggering oversight in data analysis despite our ability to collect more and more information. This inability for humans to analyze all of the data we produce is precisely why businesses are looking for ways to incorporate artificial intelligence and machine learning into their IIoT applications.

Imagine if we relied solely on human vision to manually inspect tiny defects on golf balls on a manufacturing assembly line for 8 hours each day, 5 days a week. Even if you could afford a whole army of inspectors, each person is still naturally susceptible to fatigue and human error. Similarly, manual visual inspection of railway track fasteners, which can only be performed in the middle of the night after trains have stopped running, is not only time-consuming but also difficult to do<sup>6</sup>. Likewise, manual inspection of high-voltage power lines and substation equipment also exposes human personnel to additional risks.

## Combining AI with IIoT

In each of the previously discussed industrial applications, the "AIoT" offers the ability to reduce labor costs, reduce human error, and optimize preventive maintenance. The "Artificial Intelligence of Things" (AIoT) refers to the adoption of AI technologies in "Internet of Things" (IoT) applications for the purposes of improving operational efficiency, human-machine interactions, and data analytics and management<sup>7</sup>. But what exactly do we mean by AI and how does it fit into the IIoT?

"Artificial intelligence" (AI) is the general field of science that studies how to construct intelligent programs and machines to solve problems that are traditionally performed through human intelligence. Artificial intelligence also includes "machine learning" (ML), which is a specific subset of AI that enables systems to automatically learn and improve through experience without being programmed to do so, such as through various algorithms and neural networks. Another related term is "deep learning" (DL), which is a subset of machine learning in which multilayered neural networks learn from vast amounts of data.

---

<sup>2</sup> IDC (2019). "The Growth in Connected IoT Devices Is Expected to Generate 79.4ZB of Data in 2025, According to a New IDC Forecast" <https://www.idc.com/getdoc.jsp?containerId=prUS45213219>

<sup>3</sup> Stack, Tim (2018). Cisco. *Internet of Things (IoT) Data Continues to Explode Exponentially. Who Is Using That Data and How?* <https://blogs.cisco.com/datacenter/internet-of-things-iot-data-continues-to-explode-exponentially-who-is-using-that-data-and-how>

<sup>4</sup> Dallemulle, L. & Davenport, T.H. (2017). "What's Your Data Strategy?" *Harvard Business Review*, May–June 2017, 112–121. <https://hbr.org/2017/05/whats-your-data-strategy>

<sup>5</sup> Intel Newsroom. (2018, May 16). *Intel Vision Products: See Your Data's Full Potential*. [Video] YouTube. [https://www.youtube.com/watch?v=4bK4rjatDBg&feature=emb\\_logo](https://www.youtube.com/watch?v=4bK4rjatDBg&feature=emb_logo)

<sup>6</sup> Ho, G. (2019) Global Railway Review. *Artificial intelligence in rail: Hype or reality?* <https://www.globalrailwayreview.com/article/78709/ai-hype-reality-transform-technology/>

<sup>7</sup> TechTarget (2019). "Artificial Intelligence of Things (AIoT)" <https://internetofthingsagenda.techtarget.com/definition/Artificial-Intelligence-of-Things-AIoT>

Since AI is such a broad discipline, the following discussion focuses on how **computer vision** or AI-powered video analytics, other subfields of AI often used in conjunction with ML, are used for classification and recognition in industrial applications. From data reading in remote monitoring and preventive maintenance, to identifying vehicles<sup>8</sup> for controlling traffic signals<sup>9</sup> in intelligent transportation systems, to agricultural drones and outdoor patrol robots, to **automatic optical inspection** (AOI) of tiny defects in golf balls<sup>10</sup> and other products, computer vision and video analytics are unleashing greater productivity and efficiency for industrial applications.

## Moving AI to the IIoT Edge

As previously mentioned, the proliferation of IIoT systems is generating massive amounts of data. For example, the multitude of sensors and devices in a large oil refinery generates 1 TB of raw data per day<sup>11</sup>. Immediately sending all this raw data back to a public cloud or private server for storage or processing would require considerable bandwidth, availability, and power consumption<sup>12</sup>. In many industrial applications, especially highly distributed systems located in remote areas, constantly sending large amounts of data to a central server is not possible. Even if you had the bandwidth and sufficient infrastructure, which would be incredibly costly to deploy and maintain, there would still be substantial latency in data transmission and analysis. Mission-critical industrial applications must be able to analyze raw data as quickly as possible.

In order to reduce latency, reduce data communication and storage costs, and increase network availability, IIoT applications are moving AI and machine learning capabilities to the edge of the network to enable more powerful preprocessing capabilities directly in the field. More specifically, advances in edge computing processing power have enabled IIoT applications to take advantage of AI decision-making capabilities in remote locations. Indeed, by connecting your field devices to edge computers equipped with powerful local processors and AI, you no longer need to send all of your data to the cloud for analysis<sup>13</sup>. In fact, the data created and processed at the far-edge and near-edge sites is expected to increase from 10% to 75% by 2025<sup>14</sup>, and the overall edge AI hardware market is expected to see a CAGR of 20.64% from 2019 to 2024<sup>15</sup>.

---

<sup>8</sup> Frost, A. (2019) Traffic Technology Today. *AI-based traffic monitoring system developed by researchers* <https://www.trafficechnologytoday.com/news/congestion-reduction/ai-based-traffic-monitoring-system-developed-by-researchers.html>

<sup>9</sup> Pozanco, A., Fernandex, S., & Borrajo, D. (2016) "Urban Traffic Control Assisted by AI Planning and Relational Learning" *CEUR Workshop Proceedings (CEUR-WS.org)* <http://ceur-ws.org/Vol-1678/paper6.pdf>

<sup>10</sup> Wu, H.H., Su, J.W., & Chen, C.L. (2016, September 28) "Automatic optical inspection system design for golf ball." *Proc. SPIE 9971, Applications of Digital Image Processing XXXIX, 99712S.* <https://doi.org/10.1117/12.2237286>

<sup>11</sup> Bechtold, B. (2018) Cisco. *Beyond the Barrel: How Data and Analytics will become the new currency in Oil and Gas* <https://gblogs.cisco.com/ca/2018/06/07/beyond-the-barrel-how-data-and-analytics-will-become-the-new-currency-in-oil-and-gas/>

<sup>12</sup> Reese, B. (2019) GigaOm. "AI at the Edge: A GigaOm Research Byte" *Knowingly, Inc.* <http://files.iccmedia.com/pdf/190206arm.pdf>

<sup>13</sup> Dickson, B. (2018) PCMag. *When the Cloud Is Swamped, It's Edge Computing, AI to the Rescue* <https://uk.pcmag.com/features/94303/when-the-cloud-is-swamped-its-edge-computing-ai-to-the-rescue>

<sup>14</sup> Van der Muelen, R. (2018) Gartner. *What Edge Computing Means for Infrastructure and Operations Leaders* <https://www.gartner.com/smarterwithgartner/what-edge-computing-means-for-infrastructure-and-operations-leaders/>

<sup>15</sup> Markets and Markets (2019). "Edge AI Hardware Market by Device (Smartphones, Cameras, Robots, Automobile, Smart Speakers, Wearables, and Smart Mirror), Processor (CPU, GPU, ASIC and Others), Power Consumption, Process, End User Industry, and Region - Global Forecast to 2024" <https://www.marketsandmarkets.com/Market-Reports/edge-ai-hardware-market-158498281.html>

## Choosing the Right Edge Computer for Industrial AIoT

When it comes to bringing artificial intelligence to your industrial IoT applications, there are several key issues you need to consider. Even though most of the work involved with training your AI models still takes place in the cloud, you'll eventually need to deploy your trained inferencing models in the field. AIoT edge computing essentially enables AI inferencing in the field rather than sending raw data to the cloud for processing and analysis. In order to effectively run AI models and algorithms, industrial AIoT applications require a reliable hardware platform at the edge. To choose the right hardware platform for your AIoT application, consider the following factors.

1. Processing Requirements for Different Phases of AI Implementation
2. Edge Computing Levels
3. Development Tools
4. Environmental Concerns

### Processing Requirements for Different Phases of AI Implementation

Generally speaking, processing requirements for AIoT computing are concerned with how much computing power you need and whether you need a CPU or accelerator. Since each of the following three phases in building an AI edge computing application uses different algorithms to perform different tasks, each phase has its own set of processing requirements.

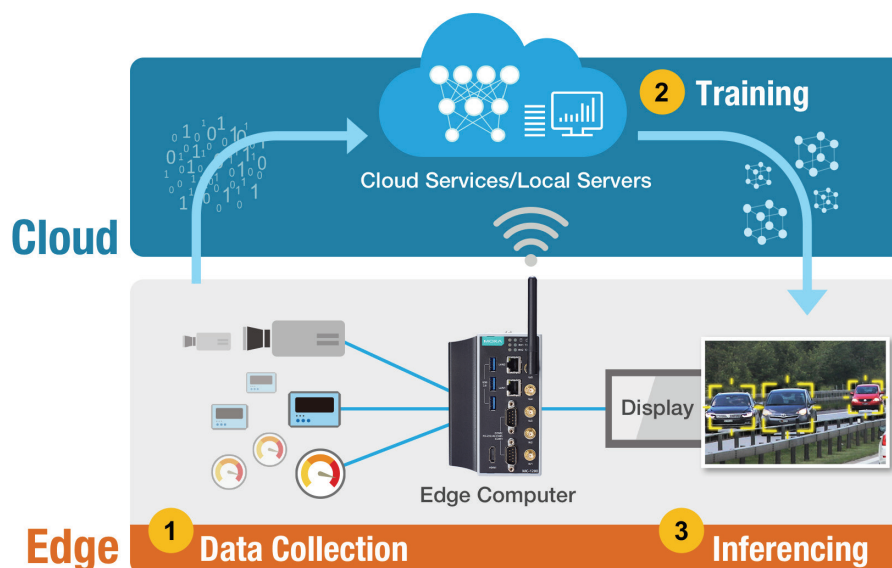


Figure 1: Three Phases in Building AIoT Applications

- **Data Collection**

The goal of this phase is to acquire large amounts of information to train the AI model. Raw, unprocessed data alone is not helpful because the information could contain duplications, errors, and outliers. Preprocessing collected data at the initial phase to identify patterns, outliers, and missing information also allows you to correct errors and biases. Depending on the complexity of the data collected, the computing platforms typically used in data collection are usually based on Arm® Cortex® or Intel® Atom®/Core™ processors. In general, I/O and **CPU** specifications (rather than the GPU) are more important for performing data collection tasks.

- **Training**

AI models need to be trained on advanced neural networks and resource-hungry machine learning or deep learning algorithms that demand more powerful processing capabilities, such as powerful **GPUs**, to support parallel computing in order to analyze large amounts of collected and preprocessed training data. Training your AI model involves selecting a machine learning model and training it on your collected and preprocessed data. During this process, you also need to evaluate and tune the parameters to ensure accuracy. Many training models and tools are available for you to choose from, including off-the-shelf deep learning design frameworks such as PyTorch, TensorFlow, and Caffe. Training is usually performed on designated AI training machines or cloud computing services, such as AWS Deep Learning AMIs, Amazon SageMaker Autopilot, Google Cloud AI, or Azure Machine Learning, instead of in the field.

- **Inferencing**

The final phase involves deploying the trained AI model on the edge computer so that it can make inferences and predictions based on newly collected and preprocessed data quickly and efficiently. Since the inferencing stage generally consumes fewer computing resources than training, a **CPU or lightweight accelerator** may be sufficient for your AIoT application. Nonetheless, you will need a conversion tool to convert the trained model to run on specialized edge processors/accelerators, such as Intel® OpenVINO™ or NVIDIA® CUDA®. Inferencing also includes several different edge computing levels and requirements, which are discussed in the following section.

## Edge Computing Levels

Although AI training is still mainly performed in the cloud or on local servers, data collection and inferencing necessarily take place at the edge of your network. Moreover, since inferencing is where your trained AI model does most of the work to accomplish the application objectives (i.e., make decisions or perform actions based on newly collected field data), you need to determine which of the following levels of edge computing you need in order to choose the appropriate processor.

- **Low Edge Computing Level**

Transferring data between the edge and the cloud is not only expensive, but also time-consuming and results in latency. With low edge computing, you only send a small amount of useful data to the cloud, which reduces lag time, bandwidth, data transmission fees, power consumption, and hardware costs. An Arm®-based platform without accelerators can be used on IIoT devices to collect and analyze data to make quick inferences or decisions.

- **Medium Edge Computing Level**

This level of inference can handle various IP camera streams for computer vision or video analytics with sufficient processing frame rates. Medium edge computing includes a wide range of data complexity based on the AI model and performance requirements of the use case, such as facial recognition applications for an office entry system versus a large-scale public surveillance network. Most industrial edge computing applications also need to factor in a limited power budget or fanless design for heat dissipation. It may be possible to use a high-performance CPU, entry-level GPU, or VPU at this level. For instance, the Intel® Core™ i7 Series CPUs offer an efficient computer vision solution with the OpenVINO toolkit and software based AI/ML accelerators that can perform inference at the edge.

- **High Edge Computing Level**

High edge computing involves processing heavier loads of data for AI expert systems that use more complex pattern recognition, such as behavior analysis for automated video surveillance in public security systems to detect security incidents or potentially threatening events. High Edge Compute Level inferencing generally uses accelerators, including a high-end GPU, VPU, TPU, or FPGA, which consumes more power (200 W or more) and generates excess heat. Since the necessary power consumption and heat generated may exceed the limits at the far edge of the network, such as aboard a moving train, high edge computing systems are often deployed in near-edge sites, such as in a railway station, to perform tasks.

## Development Tools

Several tools are available for various hardware platforms to help speed up the application development process or improve overall performance for AI algorithms and machine learning.

### Deep Learning Frameworks

Consider using a deep learning framework, which is an interface, library, or tool that allows you to build deep learning models more easily and quickly, without getting into the details of the underlying algorithms. Deep learning frameworks provide a clear and concise way for defining models using a collection of pre-built and optimized components. The three most popular include:

- **PyTorch**

Primarily developed by Facebook's AI Research Lab, PyTorch is an open source machine learning library based on the Torch library. It is used for applications such as computer vision and natural language processing, and is a free and open-source software released under the Modified BSD license.

<https://pytorch.org/>

- **TensorFlow**

Enable fast prototyping, research, and production with TensorFlow's user-friendly Keras-based APIs, which are used to define and train neural networks.

<https://www.tensorflow.org/>

- **Caffe**

Caffe provides an expressive architecture that allows you to define and configure models and optimizations without hard-coding. You can set a single flag to train the model on a GPU machine, and then deploy to commodity clusters or mobile devices.

<https://caffe.berkeleyvision.org/>

### Hardware-based Accelerator Toolkits

AI accelerator toolkits are available from hardware vendors and are specially designed to accelerate artificial intelligence applications, such as machine learning and computer vision, on their platforms.

- **Intel® OpenVINO™**

The Open Visual Inference and Neural Network Optimization (OpenVINO) toolkit from Intel is designed to help developers build robust computer vision applications on Intel® platforms. OpenVINO also enables faster inference for deep learning models.

<https://software.intel.com/content/www/us/en/develop/tools/opencvino-toolkit.html>

- **NVIDIA® CUDA®**

The CUDA Toolkit enables high-performance parallel computing for GPU-accelerated applications on embedded systems, data centers, cloud platforms, and supercomputers built on the Compute Unified Device Architecture (CUDA) from NVIDIA.

<https://developer.nvidia.com/cuda-toolkit>

## Environmental Considerations

Last but not least, you also need to consider the physical location of where your application will be implemented. Industrial applications deployed outdoors or in harsh environments—such as smart city, oil and gas, mining, power, or outdoor patrol robot applications—should have a wide operating temperature range and appropriate heat dissipation mechanisms to ensure reliability in blistering hot or freezing cold weather conditions. Certain applications also require industry-specific certifications or approvals, such as fanless design, explosion proof construction, and vibration resistance. And since many real-world applications are deployed in space-limited cabinets and subject to size limitations, small form factor edge computers are preferred.

Moreover, highly distributed industrial applications in remote sites may also require communications over a reliable cellular or Wi-Fi connection. For instance, an industrial edge computer with integrated cellular LTE connectivity eliminates the need for an additional cellular gateway and saves valuable cabinet space and deployment costs. Another consideration is that redundant wireless connectivity with dual SIM support may also be needed to ensure that data can be transferred if one cellular network signal is weak or goes down.

To see how real-world industrial applications enable and benefit from AIoT edge computing, let's examine the following two examples.

## Case in Point 1: How AIoT Keeps Mass Transit On Track

All trains—whether in an inter-city railway line or municipal mass transit system—run on metal tracks that need to remain upright and properly spaced according to a standard gauge at all times. If the tracks become uneven, trains could derail. That's why you always see some sort of support, known as railroad ties or ballasts, laid perpendicularly beneath the tracks. To ensure a smooth ride, railroad tracks need to be securely fastened to the ties by spikes, screws, or bolts. Due to constant friction and vibration between fast-moving train wheels and the tracks, as well as damage from the natural environment, track fasteners degrade and break over time. Consequently, timely detection and repair of track fasteners is crucial to ensuring the safety of any railway line.

A large metropolitan railway in East Asia needed a more efficient way to inspect the vast number of fasteners used to stabilize thousands of miles of tracks throughout its entire mass transit system. Located in the Ring of Fire where many earthquakes occur, the transit system



cannot take any chances on the safety of its infrastructure since constant tremors compound the regular wear and tear from rolling stock and high passenger traffic. Usually, after train service ends on one of the lines, the railway operator dispatches human maintenance engineers to perform manual visual inspection of the tracks and check for loose fasteners. If a loose or damaged track fastener is detected, the fastener must be repaired before train service recommences on the railway line.

Since visual inspection of railway tracks during non-operating hours is time-consuming and human fatigue may lead to data omission, the transit system decided to deploy an AI edge computing solution that could accelerate track fastener inspection with computer vision. More specifically, the transit operator wanted a customized AI inference model with object recognition for track fastening systems that could detect track fastener defects while the trains are moving and perform maintenance between journeys. AI inferencing for track fastener inspection also requires the edge computer to have powerful computing performance and storage expansion for video data, compact size and fanless design for installation in small cabinets, wide operating temperature range, and EN 50155 compliance for use on rolling stock.

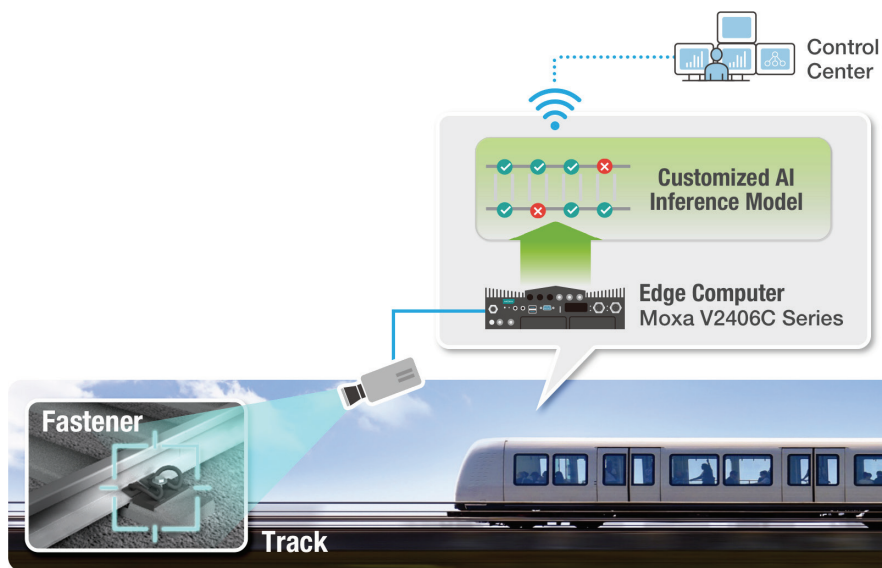


Figure 2: AIoT Track Fastener Inspection System

The first step was to install high-resolution cameras underneath the train carriages, which enabled the system operator to capture real-time video of track fasteners as trains run on the tracks during service hours. Video data is then transmitted to an onboard edge computer for image processing and object recognition of track fastener defects. The train operator selected Moxa's V2406C Series rail computer for its compact-size with an Intel® Core™ i7 processor that provides ample computing power for running the trained AI inferencing model. The V2406C also runs on low power consumption and has a wide operating temperature range of -40 to 70°C. Last but not least, the V2406C supports the Intel® OpenVINO™ toolkit and features two mPCIe slots for Intel® Movidius™ VPU modules to accelerate image recognition computations and edge AI inferencing. By replacing manual visual inspection with real-time AI visual inspection during operating hours, the transit system was able to improve efficiency and reduce maintenance expenses.

## Case in Point 2: How AIoT Helps Autonomous Mining Trucks Boost Safety and Productivity

The growing popularity of autonomous haul trucks in open-pit mining, an application which is expected to triple by 2023<sup>16</sup>, is mainly driven by the ability of autonomous hauling systems to reduce accidents, fuel consumption, and operating costs, while also increasing machine life and overall productivity. Automating the trucks not only enables mining companies to move human workers to a control room, where they can oversee operations from a safe distance, but also optimizes overall production by shortening truck exchanges and eliminating shift changes<sup>17</sup>.

Surface mining operations depend on heavy-duty dump trucks, called “haul trucks,” to transport rocks and debris from excavation sites. Due to the heavy weight and large volume of rocks and other materials that need to be moved in mining operations, haul trucks are often massive vehicles in their own right. For example, some of the largest haul trucks used in open-pit mining are designed to carry payloads of 400 tons or more<sup>18</sup>. Traditionally, these giant vehicles are operated by human drivers in quarries located in dangerous, extreme outdoor environments, such as deserts or mountains, where explosives are used to excavate mineral resources and ore from the Earth’s surface. Besides the inherent dangers involved with open-pit mining, human truck drivers often need to work 12-hour shifts or longer<sup>19</sup>, which results in fatigue and a greater risk of human error. In recent years, leading mining companies around the world have been increasingly looking towards autonomous technology and AI to help improve occupational safety and productivity.

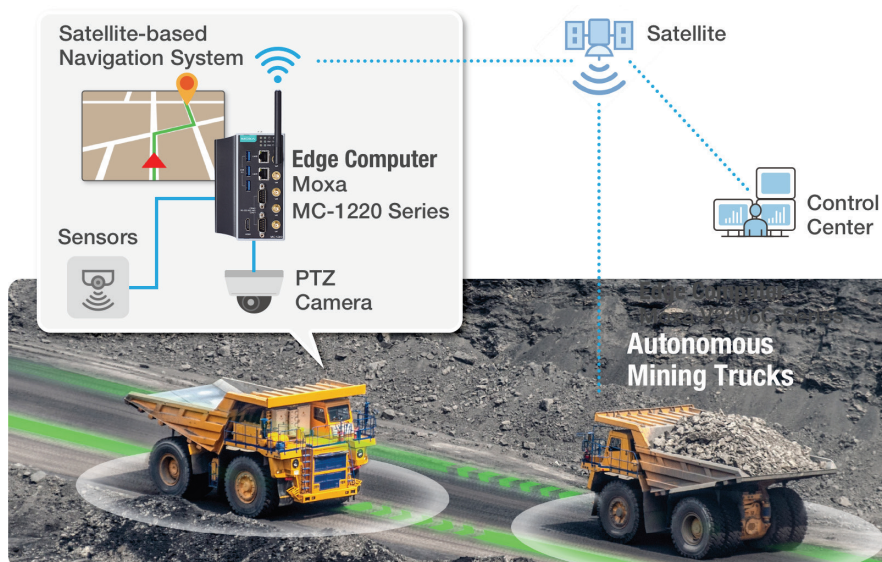


Figure 3: Autonomous Haulage Systems (AHS)

<sup>16</sup> Mining Technology (2020, Mar 17). “Australia dominates global autonomous truck use with numbers set to triple” <https://www.mining-technology.com/comment/australia-autonomous-haul-trucks-use/>

<sup>17</sup> Robinson, D. (2019, Oct 31). NS Synergy. “How autonomous technology could make mining safer, cheaper and more valuable” <https://www.nsenegybusiness.com/features/autonomous-mining-technology-zyfra/>

<sup>18</sup> Mining Technology (2019, Jun 11). “The world’s biggest mining dump trucks” <https://www.mining-technology.com/features/feature-the-worlds-biggest-mining-dump-trucks/>

<sup>19</sup> Duffy, A. (2013, Feb 7). Australian Mining. “Q and A with a dump truck driver” <https://www.australianmining.com.au/features/q-and-a-with-a-dump-truck-driver/>

As with self-driving commercial vehicles, autonomous hauling systems involve training and deploying AI models to enable haul trucks to safely traverse rugged terrain and move rocks across the excavation site. These autonomous haulage systems (AHS) also rely on computer vision and navigation technology to enable autonomous haul trucks to identify obstacles and move into the proper position to collect excavated rocks from excavators and dump the debris in designated locations. By installing a high-performance edge computer such as the Moxa MC-1220 series to connect PTZ cameras and sensors on each autonomous haul truck in the fleet, mining companies can obtain real-time video data from the excavation site as well as the exact position of each truck.

The MC-1220 provides high-performance Intel® Core™ i7 processors for video analysis and self-driving systems, as well as Wi-Fi and cellular connectivity to transmit preprocessed field data to the control center. Since mining trucks need to traverse rugged terrain, solid metal casing and high shock and vibration tolerance are also required. What's more, extreme outdoor quarry environments also necessitate a wide operating temperature range. The MC-1220 is not only Class 1, Div. 2 certified for safe, explosion-proof operation in hazardous mining locations, but also ensures reliable performance from 40 to 70°C.

## Conclusion

As the aforementioned case studies illustrate, enabling AI capabilities at the edge allows you to effectively improve operational efficiency and reduce risks and costs for your industrial applications. Choosing the right computing platform for your industrial AIoT application should also address the specific processing requirements at the three phases of implementation: (1) data collection, (2) training, and (3) inference. For the inference phase, you also need to determine the edge computing level (low, medium, or high) so that you can select the most suitable type of processor.

By carefully evaluating the specific requirements of your AIoT application at each phase, you can choose the best-suited edge computer to sufficiently and reliably perform industrial AI inferencing tasks in the field.

### Disclaimer

This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied by law, including implied warranties and conditions of merchantability, or fitness for a particular purpose. We specifically disclaim any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document.