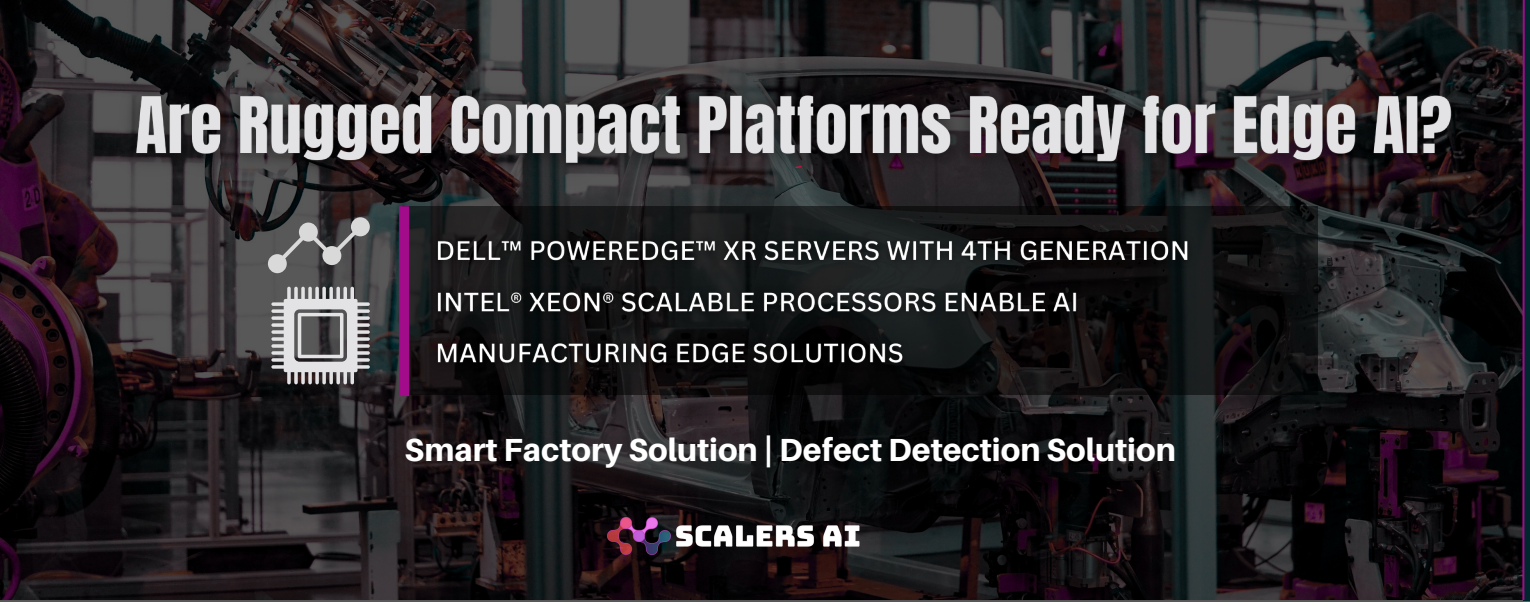


Are Rugged Compact Platforms Ready for Edge AI?

DELL™ POWEREDGE™ XR SERVERS WITH 4TH GENERATION INTEL® XEON® SCALABLE PROCESSORS ENABLE AI MANUFACTURING EDGE SOLUTIONS

Smart Factory Solution | Defect Detection Solution

SCALERS AI

Delivering Business Transformation at the Network Edge

AI at the network edge is a game changing technology that opens up new possibilities for various industries, including improved Internet of Things (IoT) applications, smarter industrial systems, and faster decision making in areas such as public safety.
It integrates artificial intelligence capabilities into edge devices and networks, which are closest to the data source. This approach enables real-time processing of data and eliminates the latency associated with sending data to centralized data centers for processing. The combination of low latency and high processing power allows for faster and more efficient analysis of data, enabling organizations to make **real-time decisions** and **unlock new business value**.

Perhaps no system is better suited to deploy AI at the network edge than **Dell™ PowerEdge™ XR platforms**. With the latest release they offer dramatic improvements on AI inference & decode of video streams with compact scalable rugged form factor that meets the rigorous requirements for enterprise-class computing outside the data center.

Dell™ PowerEdge™ XR5610 Server supports 50 simultaneous streams running AI defect detection in a single CPU config with Dell™ PowerEdge™ XR Portfolio offering scalability to 4 CPUs at near-linear scale

1.4x Gen on Gen Performance Improvement Using Intel® Deep Learning Boost & Intel® OpenVINO™

Using Dell™ PowerEdge™ XR Server & 4th Gen Intel® Xeon® Scalable Processors with edge AI throughout manufacturing lines allows you to find defects early in the manufacturing process, reducing commodity input costs, improving yields, as well as lowering emissions.

- Steen Graham, CEO at Scalers AI™

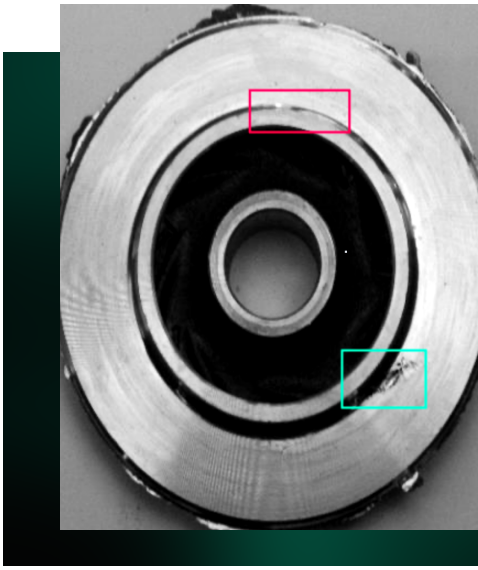
How AI at the Edge Transforms Quality Inspection in Manufacturing

AI at the network edge is revolutionizing quality inspection in manufacturing by enabling real-time analysis of production data. By integrating artificial intelligence capabilities into edge devices on the factory floor, manufacturers can perform in-process inspections and detect defects as they occur. This allows for immediate corrective action to be taken, reducing the need for time-consuming and expensive post-production inspections. The low latency and high processing power of AI at the network edge also enables advanced techniques such as computer vision and machine learning to be applied to the inspection process, leading to more accurate and efficient quality control.

By leveraging AI at the network edge, manufacturers can improve their overall quality, increase production efficiency, and stay ahead of the competition in an increasingly digitalized world.

CASE STUDY

Scalers AI™ Impellers Defect Inspection at the Edge



Impellers are rotating components used in various industrial processes, including fluid handling in pumps and fans. Quality inspection of impellers is crucial to ensure their reliable performance and durability. Improving the quality of impellers contributes to the profitability of the business by reducing the costs associated with defective products and improving customer satisfaction. In addition, it also improves sustainability by reducing wastage of energy and raw material.

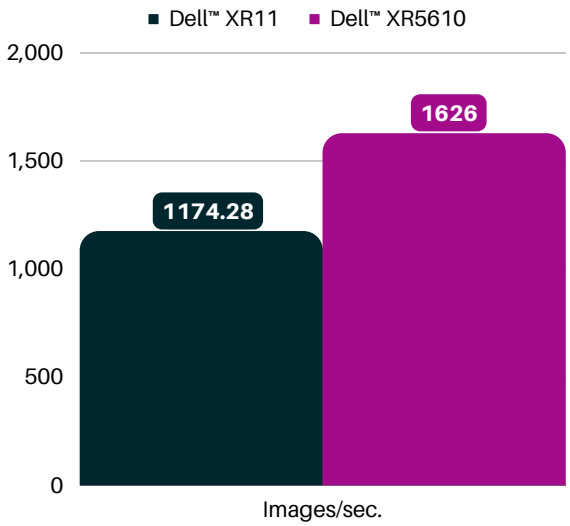
In a manufacturing facility producing impellers, AI at the network edge can be leveraged to improve the quality inspection process. By using computer vision and machine learning algorithms, the AI system can accurately inspect each impeller for defects and deviations from specifications in real-time. The system can quickly identify and flag any impellers that do not meet the quality standards, allowing for immediate corrective action to be taken. This results in a significant improvement in the overall quality of the products, increased production efficiency, and a reduction in waste and costs. By implementing AI at the network edge for quality inspection, the manufacturing facility can maintain a competitive edge and continue to produce high-quality impellers for its customers.

Product Inspection with Dell™ PowerEdge™ XR5610

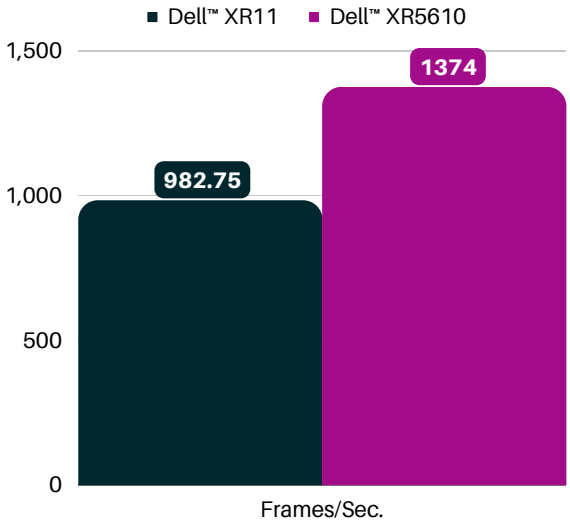
Dell™ PowerEdge™ systems are designed to process and analyze large amounts of data in real-time, making them ideal for use in product inspection scenarios. It is now possible to deploy state of the art computer vision applications leveraging AI models for object detection, defect classification and complex analytics and visualization. In addition, these systems are powerful enough to support rapid customization through the use of techniques such as transfer learning at the edge.
To learn more about how to create custom models using transfer learning follow this link:
https://infohub.delltechnologies.com/section-assets/dell-r760-transfer-learning

Dell™ PowerEdge™ XR11 & XR5610 Performance Insights

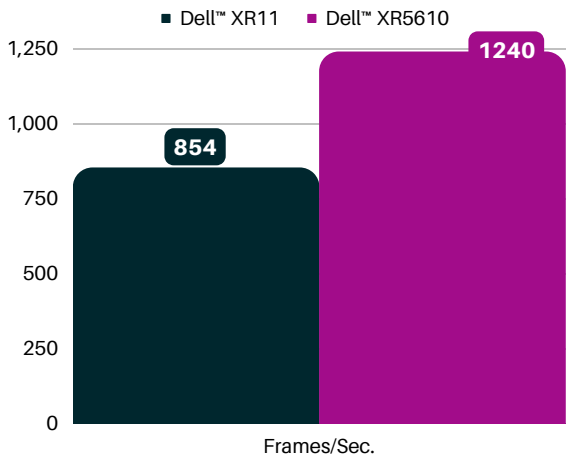
Inference Performance



Decode + Inference Performance

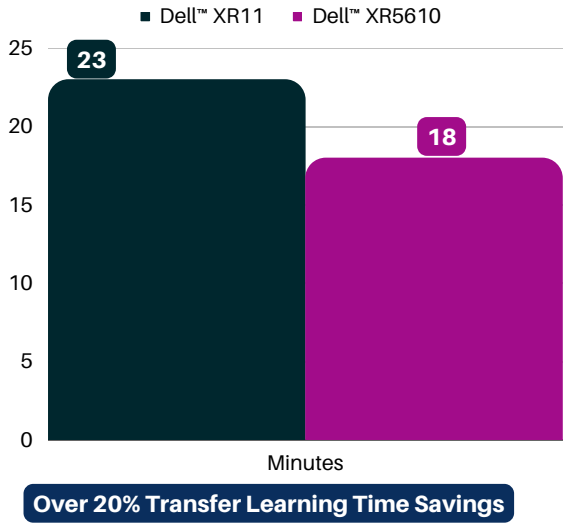


Application (Decode, Inference, App Logic)



Transfer Learning Time (Training Time)

Lower is Better



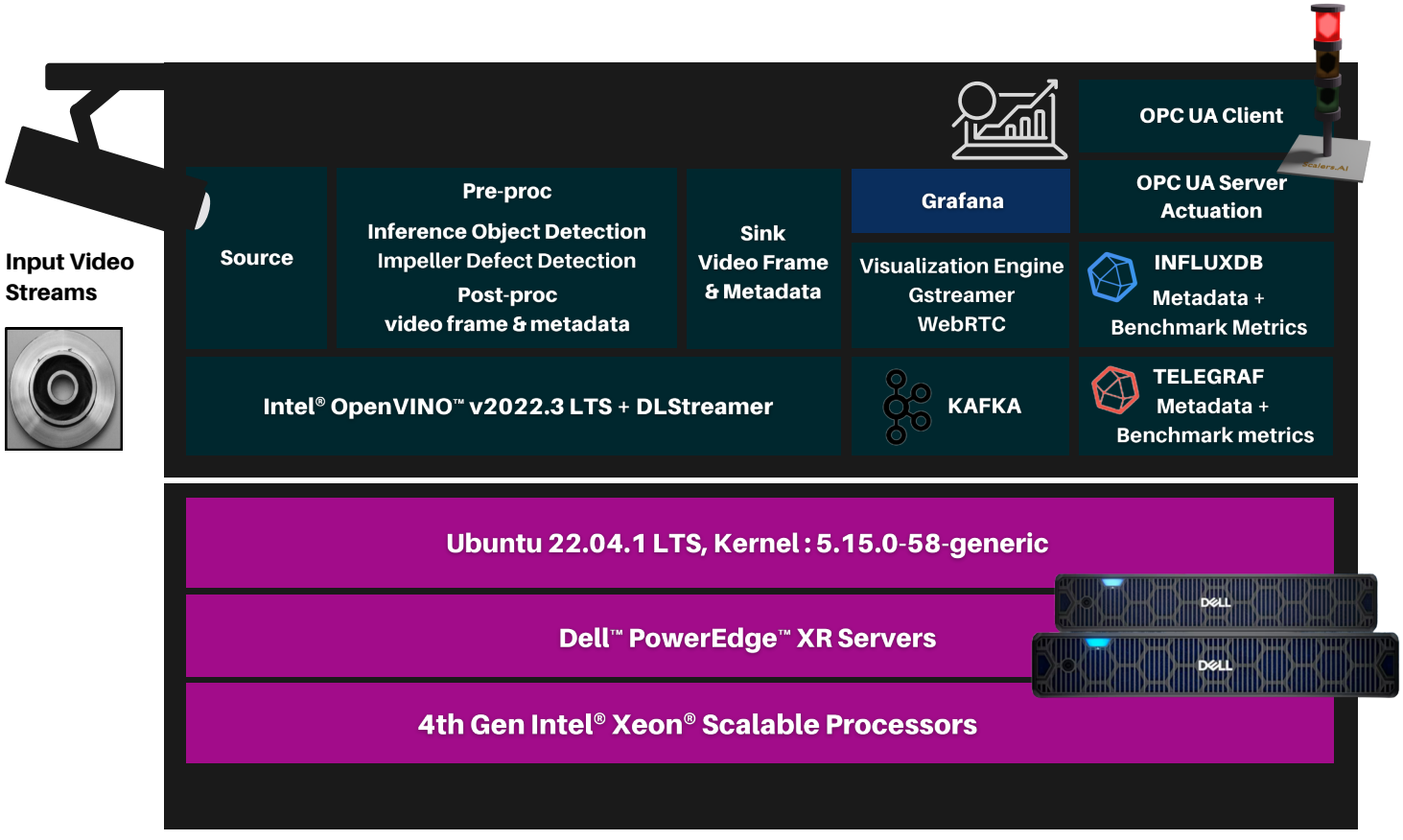
Over 20% Transfer Learning Time Savings

Solution Architecture

In this computer vision AI solution using the RTSP camera, DL Streamer framework, Kafka, Grafana, and InfluxDB, the flow of information would be as follows:

- The RTSP camera captures video footage and streams it to the DL Streamer framework, which is responsible for ingesting and processing the data.
- The DL Streamer framework performs inference on the video data using a trained machine learning model to identify objects or patterns of interest.
- The DL Streamer elements send messages about the inference results to Kafka, a messaging system that acts as a buffer between the DL Streamer and other components of the application.
- Grafana, a visualization tool, retrieves the messages from Kafka and displays them in a dashboard for users to view and analyze.
- InfluxDB, a database for storing time series data, receives the messages from Kafka and stores them for use in application analytics. This data can be used to track trends and patterns over time and to inform decision making within the application.

Overall, this flow of information allows the computer vision AI application to continuously process and analyze video data in **real-time**, and to present the results to users in a **clear** and **interactive** way.



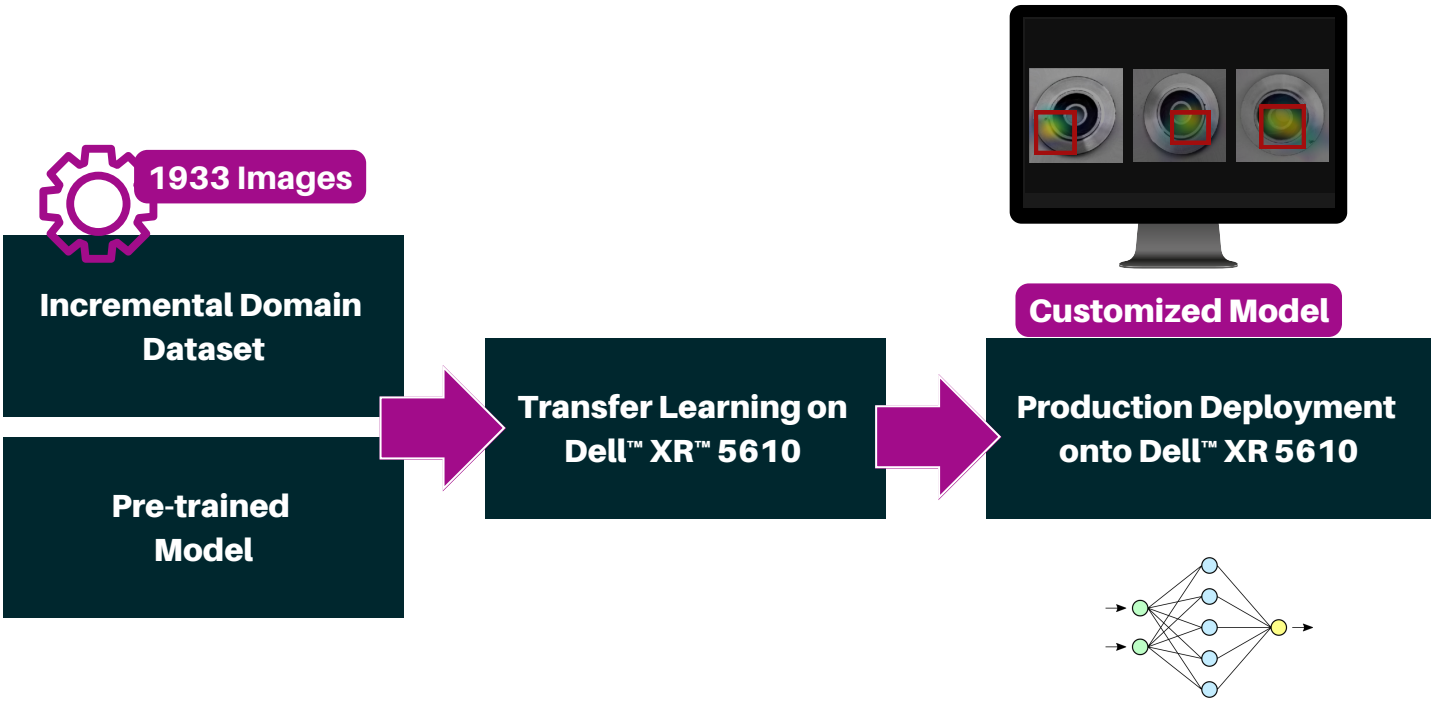
DASHBOARDS

Defect Detection Solution

Dell™ XR Performance Insights

Transfer Learning

Transfer learning is the process of taking a pre-trained machine learning model and adapting it for use on a new task or dataset. It allows enterprises to take advantage of the large amounts of data and computational resources that were used to train the original model, rather than starting from scratch. Transfer learning is often used in computer vision applications such as image classification, object detection, and segmentation, where the pre-trained model is fine-tuned on a new dataset using a smaller number of labeled examples. Using pre-trained models with transfer learning can save organizations thousands of hours of development time.



Customizing for Impeller Detection

In the case of YOLOv5, we start with a model that has already been trained to detect a wide range of objects in images and then fine-tune it to detect defects in impellers. First we create or obtain a dataset of images that contain impellers and adjusting the model's parameters to optimize its performance on this specific task. To fine-tune the YOLOv5 model for impeller defect detection, we start by freezing the weights* of the pre-trained model and then adding a few additional layers on top of it. We then train these layers using your impeller detection dataset, while leaving the weights of the pre-trained layers unchanged. This allows the resulting model to take advantage of the knowledge and features learned by the pre-trained model, while adapting the model to the specific task of impeller defect detection.

Once we have fine-tuned the model for impeller defect detection, we can deploy it into a video AI pipeline running the inference through the new model. The model outputs a set of bounding boxes around any impeller defects it detects in the image, along with a confidence score for each detection.

Today, with Dell™ PowerEdge™ XR platforms & Intel® DL Boost we are able to deliver an AI enabled software defined manufacturing architecture that allows factories to enhance their processes to adapt to customers ever changing tastes while reducing costs and lower carbon footprint

- Chetan Gadgil, CTO at Scalers AI™

The Foundation for Software Defined Manufacturing

Why a Software Defined Architecture is Better than the Traditional Bespoke Approach?

Bespoke inspection machines are designed to perform a specific set of tasks and are often tailored to meet the requirements of a single use case. While they may offer a level of accuracy and efficiency, they are limited in their capabilities and lack the versatility to adapt to changing market demands. A software-defined approach, on the other hand, utilizes edge AI servers to support multiple use cases, enabling businesses to quickly respond to changing market demands.
This approach allows for a smooth transition path, reducing the time and effort required to adopt new technologies.

By leveraging edge AI, businesses can improve their overall quality, increase production efficiency, and stay ahead of the competition in an ever-evolving market. The flexible and adaptable nature of edge AI solutions makes them a superior choice for quality inspection in modern manufacturing environments.

Conclusion

Dell™ PowerEdge™ XR 5610 servers, equipped with 4th Gen Intel® Xeon® scalable processors are well suited to handle edge AI applications with both AI inference & training at the edge. The rugged form factor, extended temp, and scalability to four sockets enables compute to be deployed in the physical world closer to the point of data creation allowing for **near-real-time insights**. These industry specific solutions can run remote with workloads previously thought to require cloud resources such as Transfer Learning or retaining the model, making it possible to do this at the edge, **faster**, and **more secure**.

Dell™ XR portfolio includes 2U dual socket Dell™ XR7620 and 2U sled-based XR8000XR8000 along with 1U Single Socket Dell™ XR5610, offering flexibility and scalability with highly configurable options for CPU, memory and I/O needed for Edge and Telecom applications.

APPENDIX

Our Solution Testing Methodology

- The workload and test cases were designed to maximize CPU utilization, ensuring that it was at least 90% throughout the scenario.
- Two Dell™ servers with different CPU models were used in the testing: Dell™ PowerEdge™ XR11 with single socket Ice Lake CPUs and Dell™ PowerEdge™ XR5610 with single Socket 4th Gen Intel® Xeon® scalable processors CPUs.
- The testing was done using the Intel® OpenVINO™ benchmark and DLStreamer benchmark, and system performance was monitored with Linux System tools.
- The AI model used in the testing was YOLOv5s and computation was in int8 format.
- The testing included scenarios with and without the use of Intel® OpenVINO™ optimization software.
- The tests for inference only comparisons using Intel® OpenVINO™ Benchmarking Tool were run with 32 steams. For tests with DLStreamer benchmark and Scalers AI™ Impeller defect detection solution , an input of 50 streams were used with a source video resolution of 1080p and a bitrate of 8624 kb/s.

Performance varies by use case, model, application, hardware & software configurations, the quality of the resolution of the input data, and other factors. This performance testing is intended for informational purposes and not intended to be a guarantee of actual performance of an AI application.

About Scalers AI™

Scalers AI™ specializes in creating end-to-end artificial intelligence (AI) solutions for a wide range of industries, including retail, smart cities, manufacturing, and healthcare. The company is dedicated to helping organizations leverage the power of AI for their digital transformation. Scalers AI™ has a team of experienced AI developers and data scientists who are skilled in creating custom AI solutions for a variety of use cases, including predictive analytics, chatbots, image and speech recognition, and natural language processing. As a full stack AI solutions company with solutions ranging from the cloud to the edge, our customers often need versatile common off the shelf (COTS) hardware that works well across a range of workloads. Additionally, we also need advanced visualization libraries including the ability to render video in modern web application architectures.

Fast track development with access to the solution code

Save hundreds of hours of development with the solution code.
As part of this effort Scalers AI™ is making the solution code available.

Reach out to your Dell™ representative or contact Scalers AI™ at contact@scalers.ai for access.