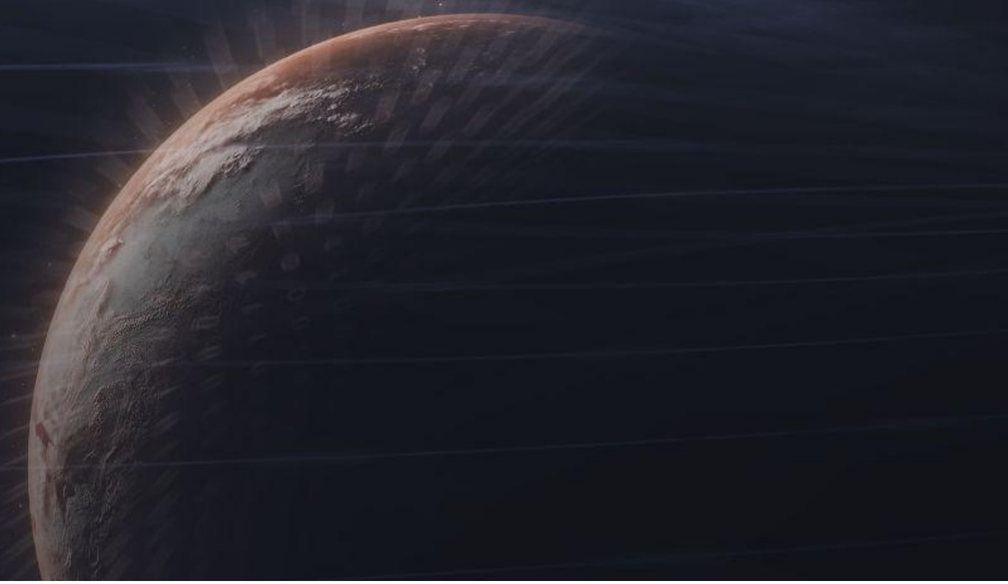
LUMI user case: Top Data Science

Kai Lehtinen, Kseniia Khakalo

March 6th 2025

Top Data Science

- Specialized AI, machine learning and software services for multiple industries
- Highly experienced team of data scientists, software engineers and business professionals
- Located in Espoo, Finland
- Excellent customer track record
  - Finland, Japan, Germany, Sweden, South Korea



SONY · Metsä · GE · storaenso · Changing worlds with you. NTT docomo · DENSO Crafting the Core · LAHTI

JYVÄSKYLÄ · SEGASammy · Metso:Outotec · brose Excellence in Mechatronics · EARTHBRAIN · J-QuAD DYNAMICS · J-POWER

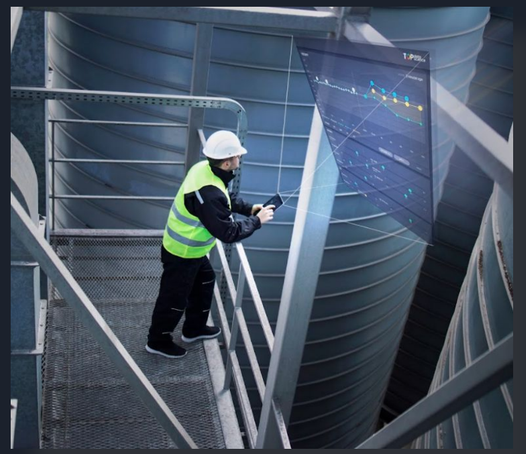
TOP data science
A MORPHO COMPANY

Speciality Areas

Industrial Computer Vision

Services that enable business process automation for higher productivity, quality and safety. Areas of expertise e.g.

- Factory & Process Automation
- Quality Assurance
- Safety & Security



Business and Process Optimization

Enabling optimizing business & process performance to achieve your business and sustainability targets

- Forestry and biomaterials industry
- Manufacturing

Software Frameworks

Software frameworks and services to extend AI-based automation through application development

- Flexibility
- Interoperability
- Scalability
- Cloud & edge

Customized AI Solutions

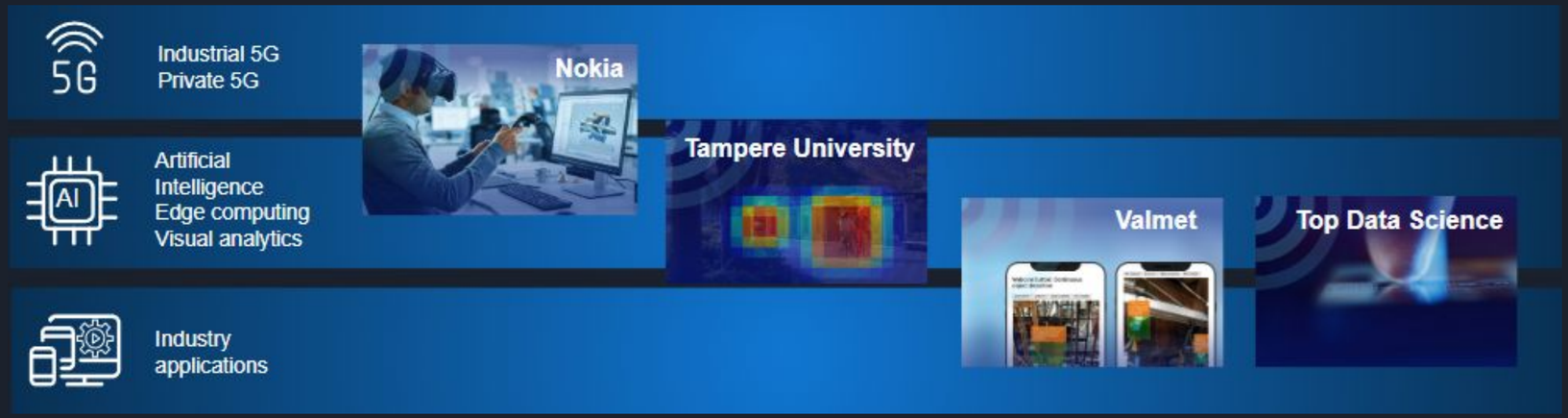
Practical problem solving by combining customer domain knowledge and our AI knowhow

- Fast prototyping
- Sharing best practices
- From POC to production

TOP data science
A MORPHO COMPANY

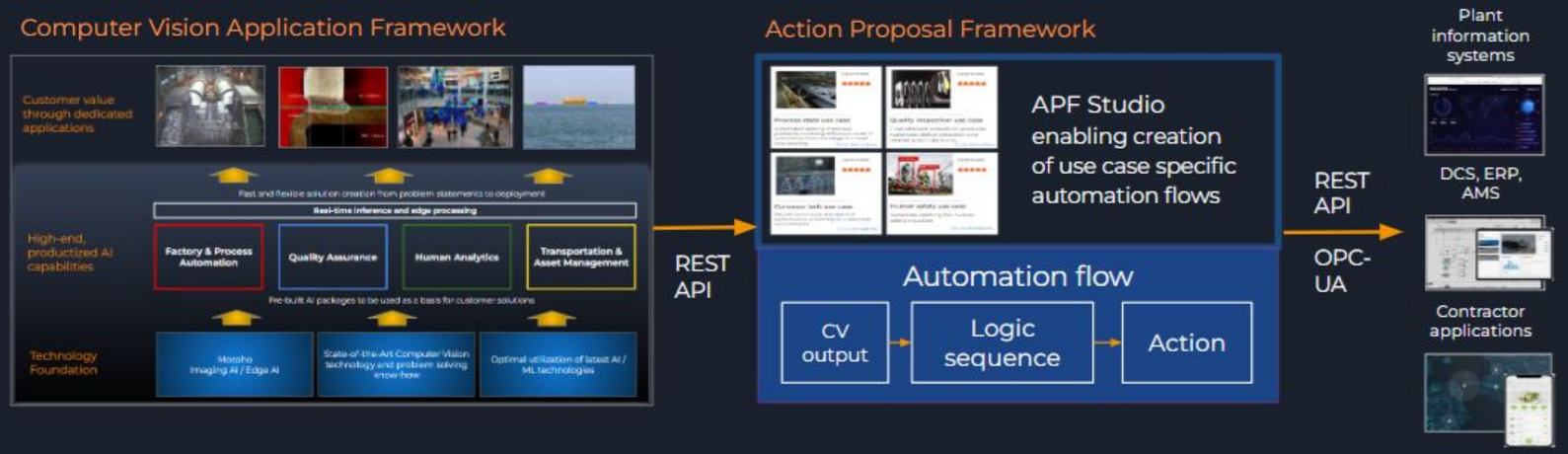
AISA project - co-innovation with leading organizations

AISA (AI-based Situational Awareness) project was a 3-year (2021-2024) co-innovation project led by **Nokia** and partially funded by **Business Finland**



Our focus in AISA project

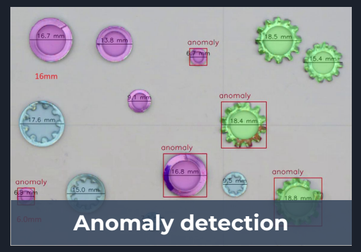
- Computer vision (CV) framework to enable fast implementation of scalable CV applications
- Extending the automation support towards action logic

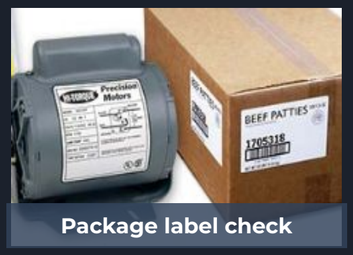


Industrial Computer Vision - Our Approach

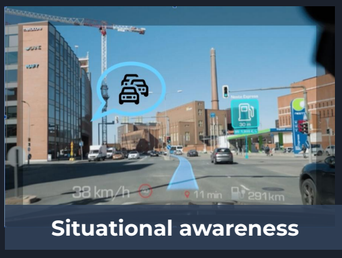
Customer value through specific use cases


Production automation


Anomaly detection


Package label check


Automated measurement


Situational awareness

Fast and flexible delivery from problem statements to deployment

Speciality areas

- Factory & Process Automation
- Quality Assurance
- Human Analytics
- Assets & Automotive

Computer Vision Application Framework

Core competencies

- State-of-the-art Computer Vision technology
- Multi-platform application development
- Inference optimization
- Problem solving know-how

Continually enhancing our capabilities by delivering customer value

Computing resources taxonomy for AI development

AI development requires diverse computing resources, each offering distinct capabilities suited to various stages of model development and deployment.

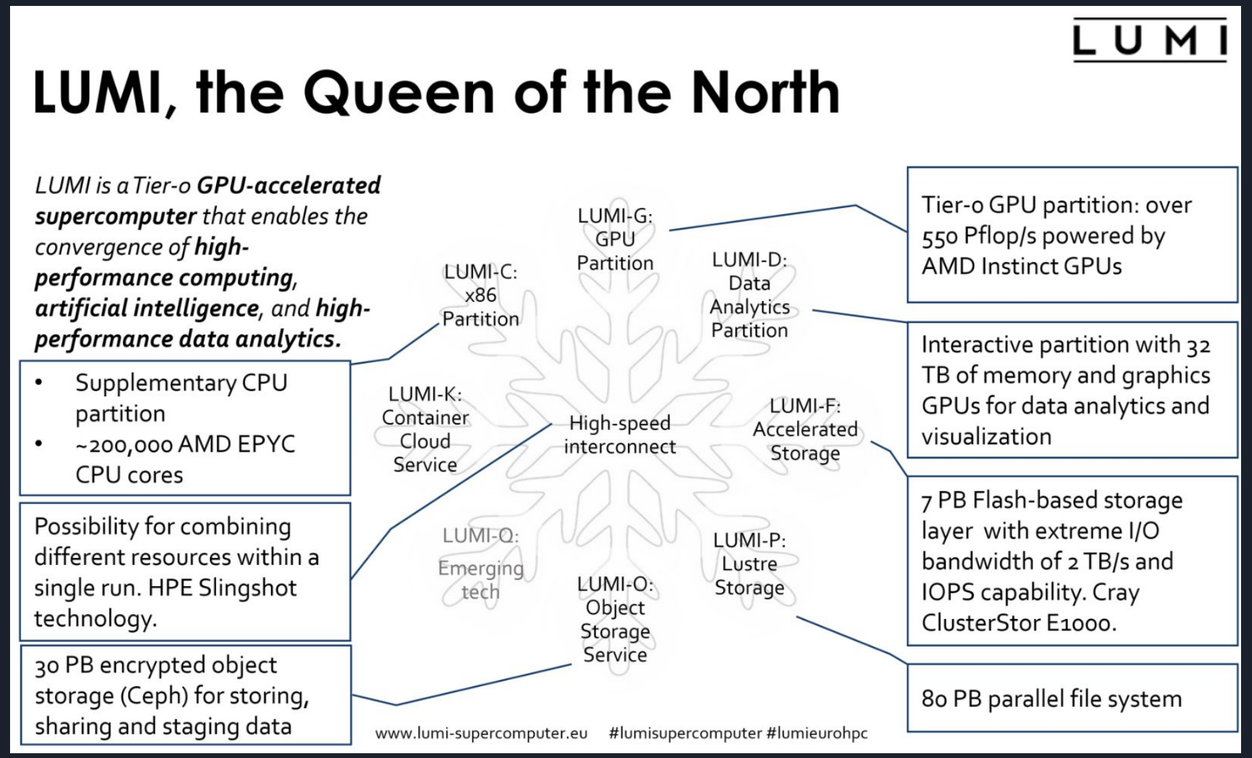
Edge computing	**Embedded systems**	**Workstations and high-performance PCs**	**Cloud-based AI services**	**Supercomputers**
Enables processing data near its source, such as on IoT devices or local servers. This approach minimizes latency and bandwidth usage, making it suitable for real-time AI applications.	Specialized computing platforms integrated into devices like wearables or consumer electronics. They run AI models directly on-device, reducing the need for cloud connectivity and enhancing privacy.	Powerful computers equipped with advanced CPUs, GPUs, and substantial RAM. Provides balance between performance and cost, suitable for developing smaller AI models without relying on cloud services.	Scalable computing resources, including specialized AI accelerators. High-performance computing without upfront HW investments, supporting the training of large AI models.	Large-scale systems designed for high-speed, parallel processing. They provide immense computational power for training complex AI models across vast datasets.

LUMI Experiments: Background

- Business Finland funding within AISA project
- Triton HPC experience in Aalto University
  - Slurm Workload Manager

Objectives:

- Study how LUMI works
- Outsource tasks which would benefit from massive parallelization
- Run tasks which require running on multiple GPUs in parallel



Human analytics on LUMI

Use case: **Human pose estimation**

- For some projects we need to pre-process videos and extract human pose estimation skeletons for further processing

- Data:
  - Multiple video clips

Video files

Lumi runs array of jobs in parallel

Trajectories are saved on disk

TOP data science
A MORPHO COMPANY

Custom running environment in LUMI

Issue:

- Human pose estimation code uses **onnxruntime-gpu** library.
- LUMI has **AMD MI250X GPU** nodes.
- To be able to use **onnxruntime** with **AMD GPU** it should be built from the source.
- Isolated running environment can be achieved with **Docker** containerisation.

Steps to solve:

1) Create docker image with preinstalled libraries
2) Push docker image to the registry
3) Pull the image in Lumi with singularity software

```
$ singularity pull docker://khakaloksenia/onnx_rocm
```

4) Specify the image as running environment when submitting the jobs

```
$ srun --partition=<partition> --account=<account> singularity exec
onnx_rocm_latest.sif cat /etc/os-release
```

Dockerfile.rocm

```
# -------------------------------------------------------------
# Copyright (c) Microsoft Corporation. All rights reserved.
# Licensed under the MIT License.
# -------------------------------------------------------------
# Dockerfile to run ONNXRuntime with ROCm integration
#--------------------------------------------------------------------------

FROM rocm/pytorch:rocm5.4_ubuntu20.04_py3.7_pytorch_1.12.1

ARG ONNXRUNTIME_REPO=https://github.com/Microsoft/onnxruntime
ARG ONNXRUNTIME_BRANCH=main

WORKDIR /code

ENV PATH /opt/miniconda/bin:/code/cmake-3.27.3-linux-x86_64/bin:${PATH}

# Prepare onnxruntime repository & build onnxruntime
RUN git clone --single-branch --branch ${ONNXRUNTIME_BRANCH} --recursive ${ONNXRUNTIM
    /bin/sh onnxruntime/dockerfiles/scripts/install_common_deps.sh &&\
    cd onnxruntime &&\
    /bin/sh ./build.sh --allow_running_as_root --config Release --build_wheel --updat
    ONNXRUNTIME_VERSION=$(cat ./VERSION_NUMBER) --use_rocm --rocm_home=/opt/rocm &&\
    pip install /code/onnxruntime/build/Linux/Release/dist/*.whl &&\
    cd ..
```

Conclusions and future steps

Conclusions:

- Overall a great experience!
- Good documentation
- Good flexibility
- Effective for running parallel jobs which are small enough to fit to 1 GPU node

Future steps:

- Interactive data analytics and visualization with LUMI-D nodes
  - Try prototyping with LUMI
- Experiments with larger models (like LLMs) which do not fit to our local servers
  - Some open source models like DeepSeek R1 can be used for data processing to create custom datasets

Summary

The primary advantage for us was to understand what LUMI offers, and how it should be strategically positioned for our AI solution development.

This helps us address customer problems and deliver tailored solutions that go beyond what the normal computational setup offers.

TOP data science
A MORPHO COMPANY

Empower Your Business with AI

We help our customers to unlock value with innovative use of AI technologies

Thank you!

Kai Lehtinen, COO

Kseniia Khakalo, Lead Data Scientist

Phone Number
+358 40 5912369

Email Address
kai.lehtinen@topdatascience.com

Website
www.topdatascience.com