GSAW 2024
EUMETSAT ML Framework System and AI/ML Applications

Gianni Casonato – Michel Horny

27.02.2024

Agenda

ML Framework System Short Summary

- Background
- Scope and Objectives
- High Level Architecture
- ML-Ops capabilities
- HW and SW layers

AI/ML Applications for EUM Use Cases

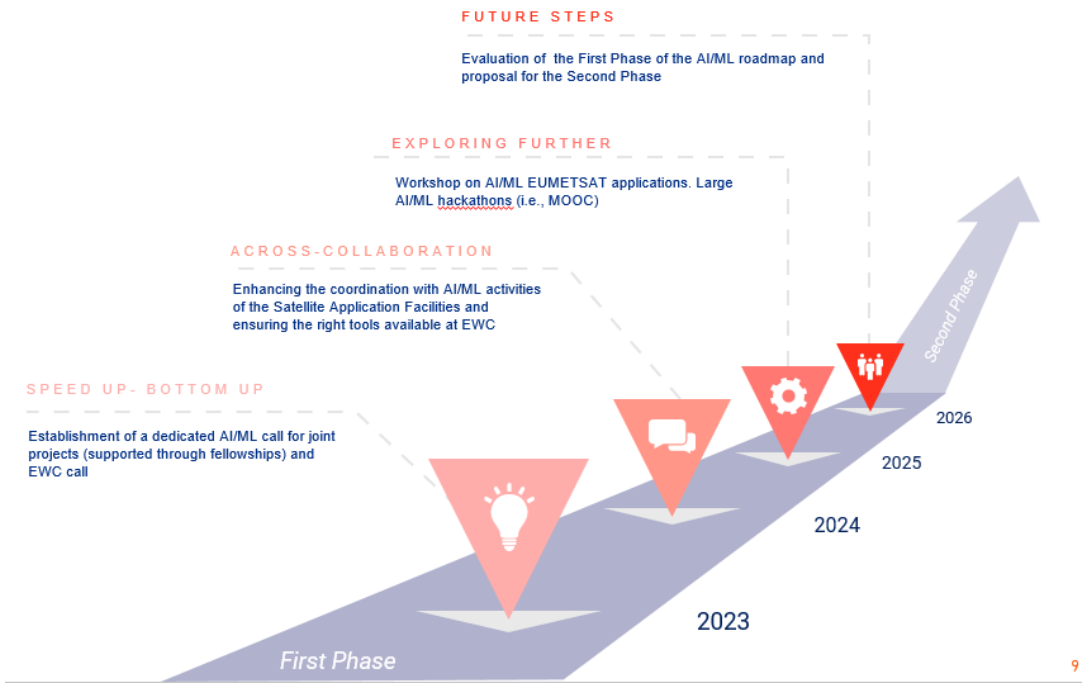
- Supported applications and models
- Examples

Summary and Way Forwards

- Future steps

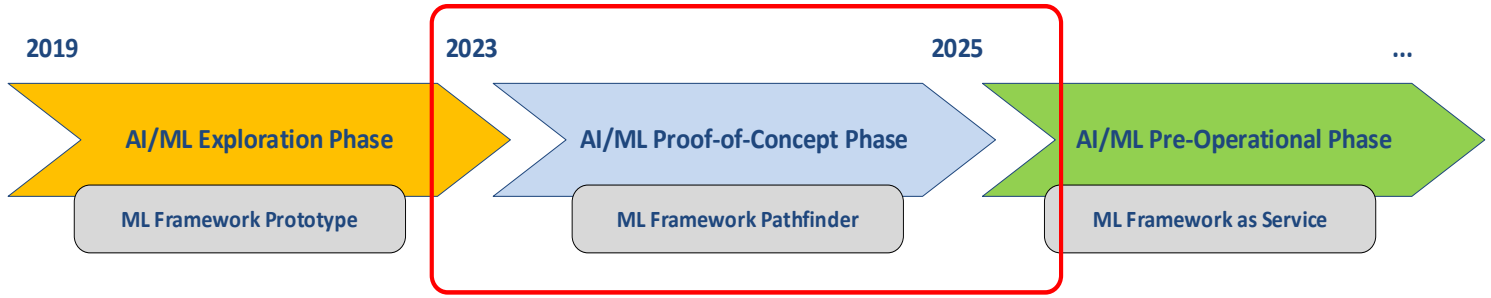
EUMETSAT ML Framework Background (1)

- A corporate EUMETSAT AI/ML Roadmap was defined in 2022 with a view until 2026 for identifying and developing AI/ML application areas within the organisation and in synergy with member states.
- It provides the umbrella for several AI/ML related projects and studies planned for next years, providing common objectives, and high level coordination and consistency



EUMETSAT ML Framework Background (2)

- As part of the EUMETSAT AI/ML Roadmap, the ML Framework (MLF) initiative was started, with a focus in 2023-2025 timeframe for identifying and supporting EUM internal operational use cases.
- MLF Prototype Phase
  - Run until early 2023 for starting looking at EUM internal use cases and developing a supporting HW/SW MLF.
- MLF Pathfinder Phase
  - Kicked off mid 2023 and on-going for enhancing the MLF resources as well as provide a more stable ML team support for EUM internal use cases proof-of-concept. Coordination with EUM AI/ML Roadmap for supporting initiatives on that side

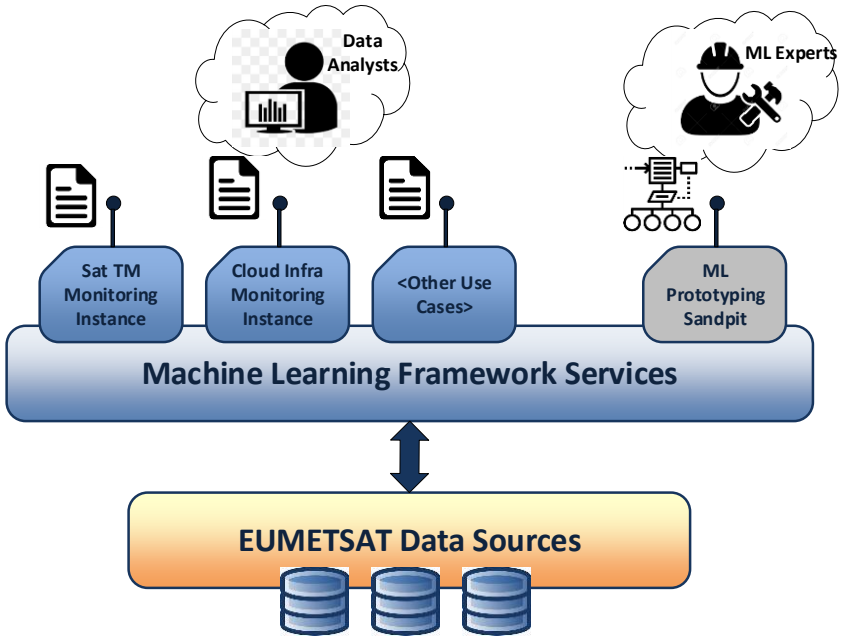


ML Framework Objectives

- MLF objectives are
  - EUM internal use cases proof-of-concept, for helping teams in assessing if/how AM/ML can be helpful in their activities
  - Make/buy future decisions support on AI/ML systems
  - System specification consolidation for a future operational ML Framework procurement
- Current Pathfinder phase timeframe
  - From Q3-2023 to Q4-2025

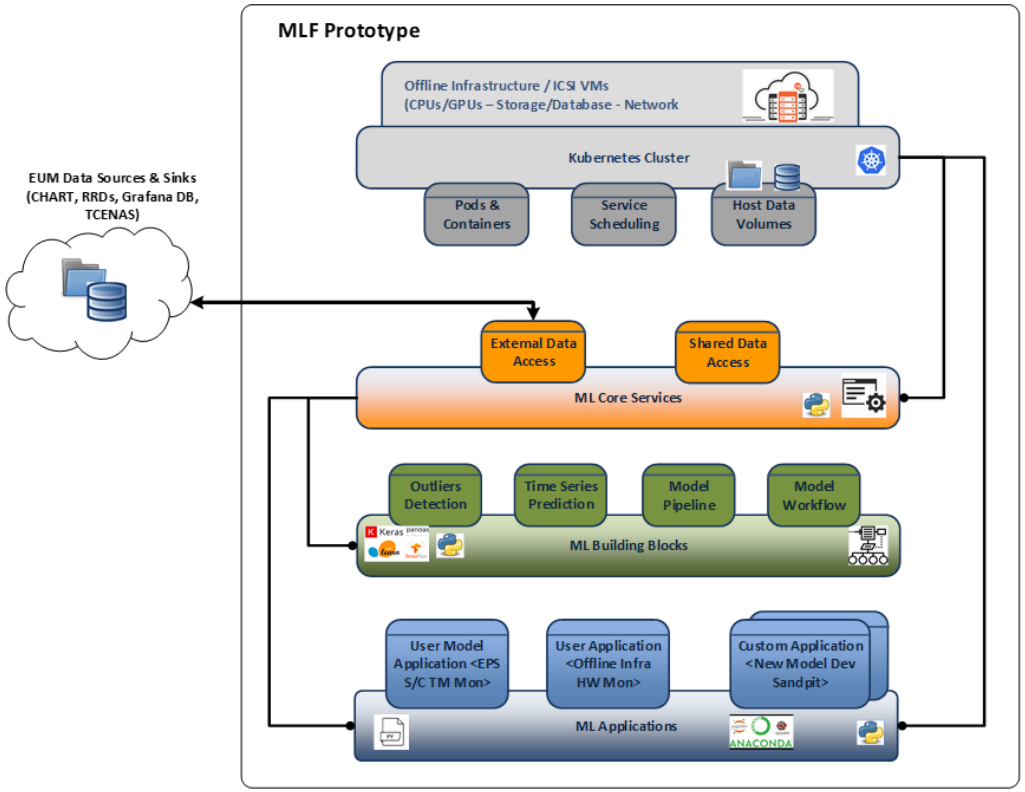
ML Framework System Concept

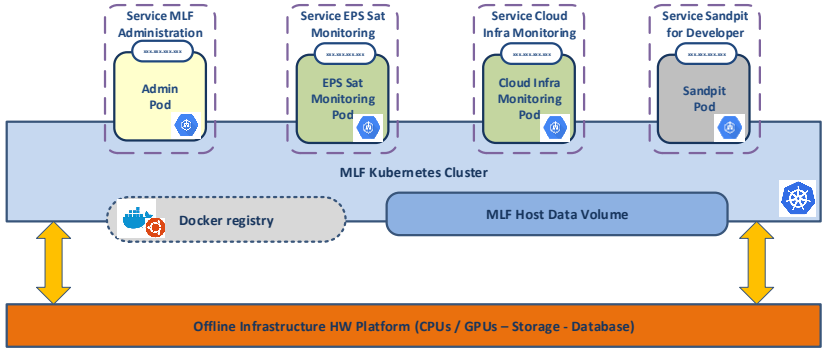
- “Production” mode automatic supporting ML models execution
- “Expert” mode for interactive ML development
- Direct access to EUMETSAT data sources
- Flexible and scalable framework usable for a wide range of AI/ML applications and models (from Neural Network to Large Language Models)
- Decoupling from underlying HW infrastructure



ML Framework System Architecture

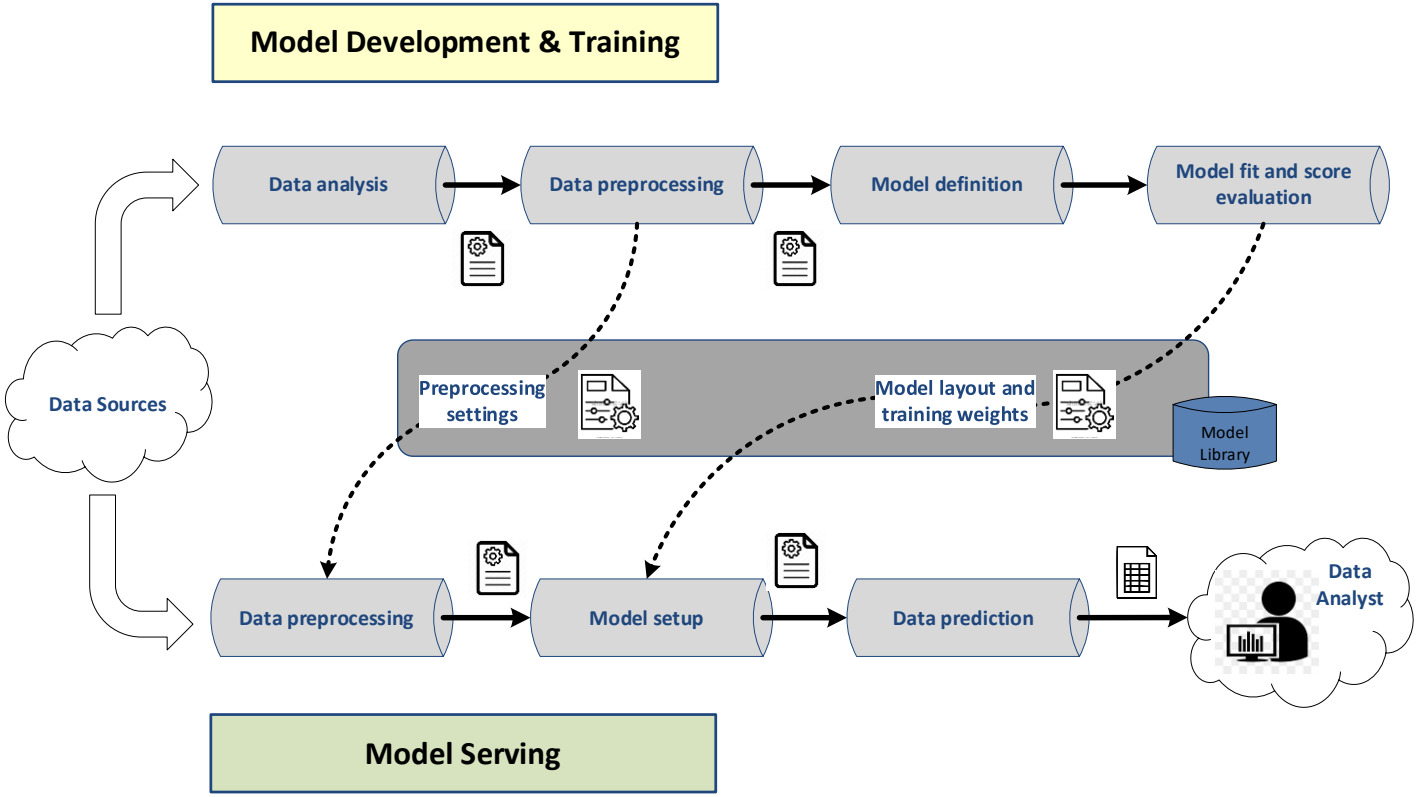
- Layered design
  - Core services layer for low-level capabilities
  - Building blocks layer for AI/ML specific capabilities
  - Applications layer for defining and run independently each AI/ML model
- Underlying platform layer on Kubernetes, hosting MLF applications in Docker images
- MLF System and AI/ML Apps developed in Python





ML Framework – ML-Ops Pipeline

- Level-1 ML-Ops capabilities supported, with manual model versions deployment and re-training and automatic model serving capabilities. Upgrade to Level-2 (full automatic pipeline) planned in MLF Pathfinder phase
- Seamless Integration with EUM data sources (Telemetry Database, Product Archive, Logs Database)
- Data analysis and model setup/tuning and training support



ML Framework – HW and SW Specs

- HW layer on K8S cluster
  - Pool of CPU workers (4x 8 vCores, 32 GB)
  - GPU support (2x NVIDIA A40 RTX, 48GB)
  - Storage (1 TB)

- SW layers base on Python
  - EUM in-house SW development for SW layers and AI/ML Apps
  - Python libraries for Neural Network models (Keras, Tensorflow, Scikit-Learn), for Image Detection (Yolo), and for Large Language Models (LangChain, HuggingFace, ChromaDB, etc.)

ML Framework - Models

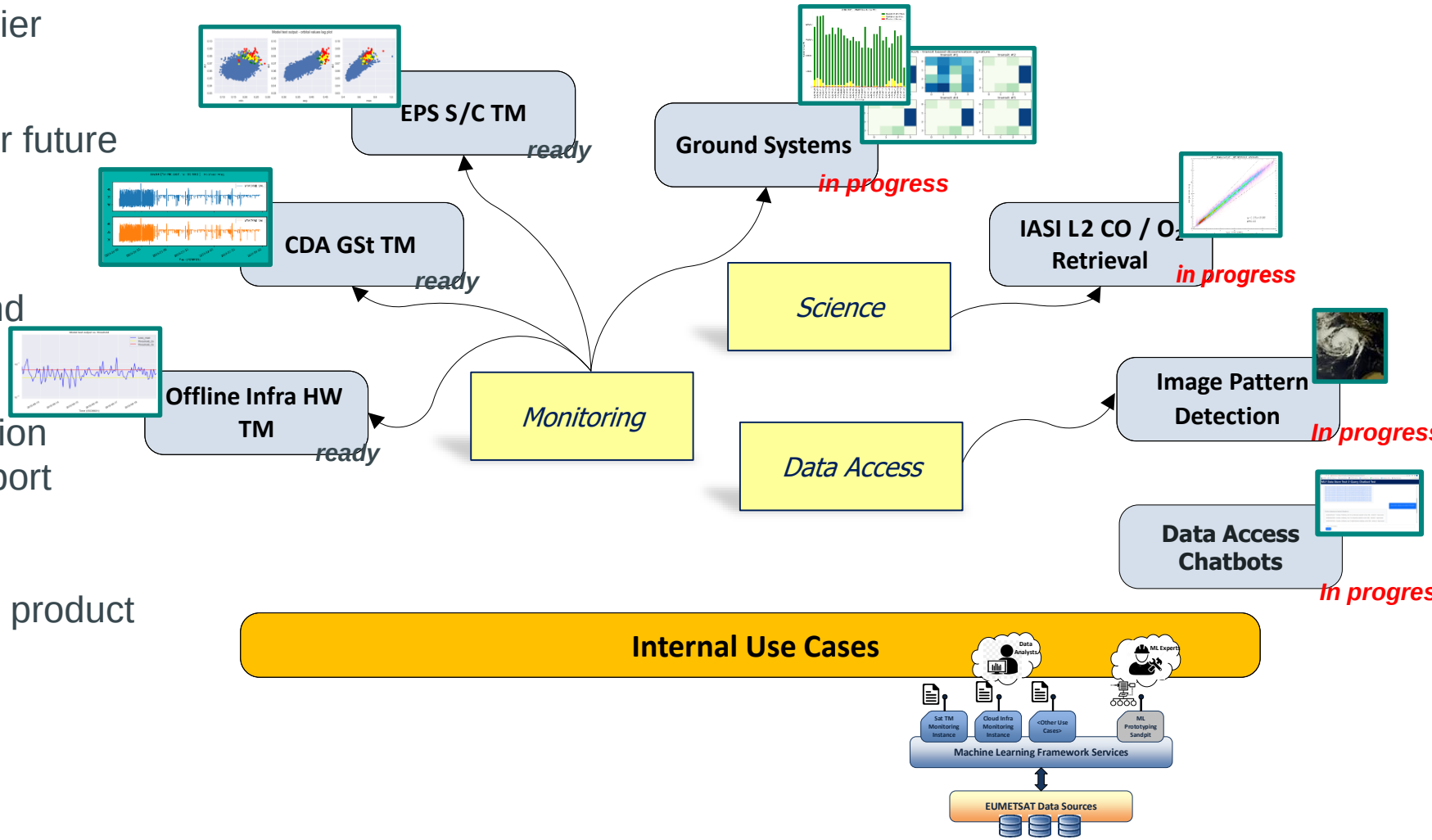
- Deep Neural Network based
  - Autoencoders for anomaly detection in various flavours
    - multi-parameters, multi-data sources, influencer-follower
    - outlier detection and timeserie prediction
  - Convolutional Neural Networks for science data retrievals
  - Deep networks for image object detection

- Large Language Models (LLM) based
  - Chatbots on specific EUM knowledge areas
    - Retrieval Augmented Generation approach using local LLM instances

- Further areas of investigation to be covered in 2024
  - “Smart” monitoring and troubleshooting (e.g. automatic data analysis and correlation with possible root causes, or with link to historical similar cases and resolutions)
  - Active agents, for supporting heavy operational tasks on large/complex systems (e.g. spacecraft constellation operations)

ML Framework AI/ML Apps Areas (so far…)

- **Satellite, ground stations and system operations**
  - Monitoring support (e.g. outlier detection, anomaly pattern identification, etc.)
  - Autonomous tasking (e.g. for future constellation missions)
- **Science Products**
  - Image anomaly detection and analysis
  - Image quality monitoring
  - Science parameters estimation
  - Algorithm development support
- **User Data Access**
  - Chatbots for Data Store and product retrieval access

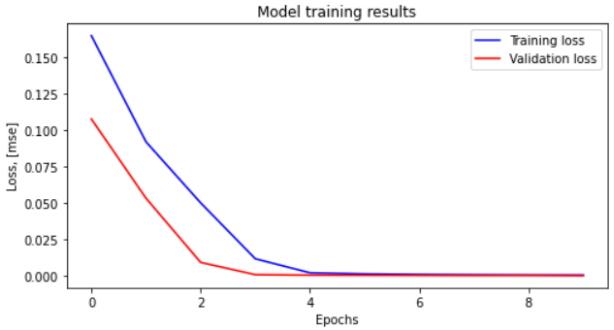


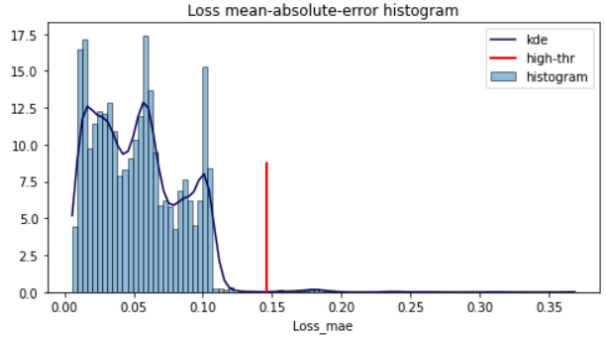
ML Framework AI/ML Apps – Some Figures

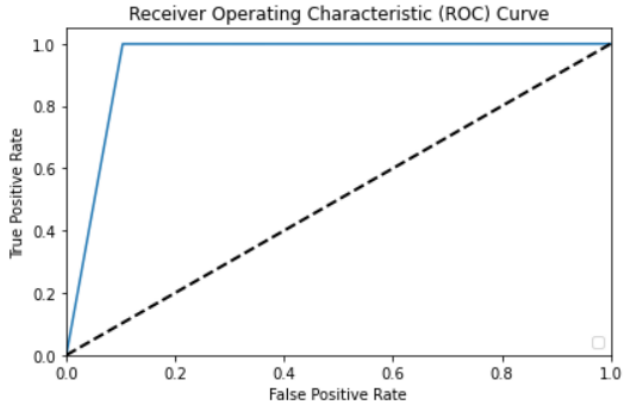
- 4 AI/ML apps deployed in production and 3 in preparation
- 5 outlier detection apps
  - Spacecraft TM: 2 S/C * 5 TM pairs → 10 parallel models
  - Ground Station TM: Prime/Backup Station * 4 TM → 8 parallel models
  - Ground System Events: ~10 Application * 5 Events → 50 parallel models
- 1 LLM based app
  - Data Access Chatbot: 7B local model + domain specific knowledge base
- 1 image detection app
  - Remarkable events detection on images: 2 event types → 2 parallel models

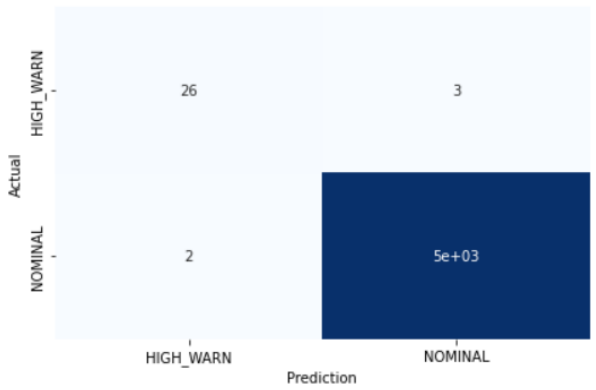
ML Framework AI/ML Apps Example – Outlier Detections

- Timeserie parameter datasets
- LSTM Neural Network Based Autoencoder Models
- Accuracy ~95%



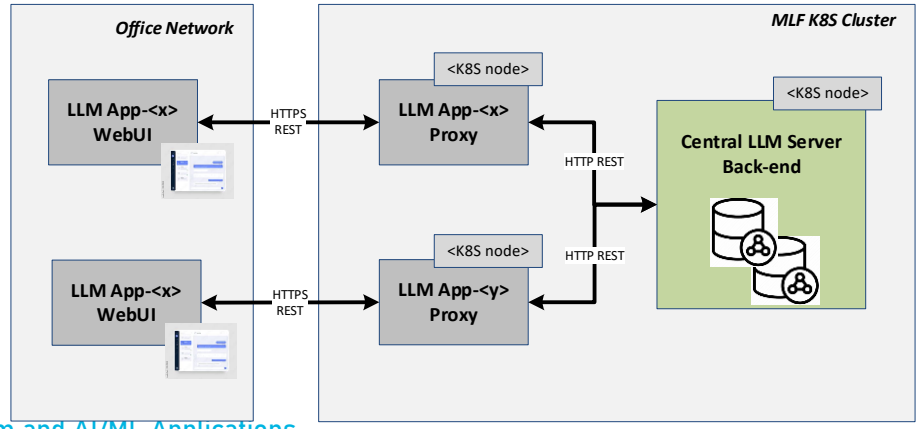


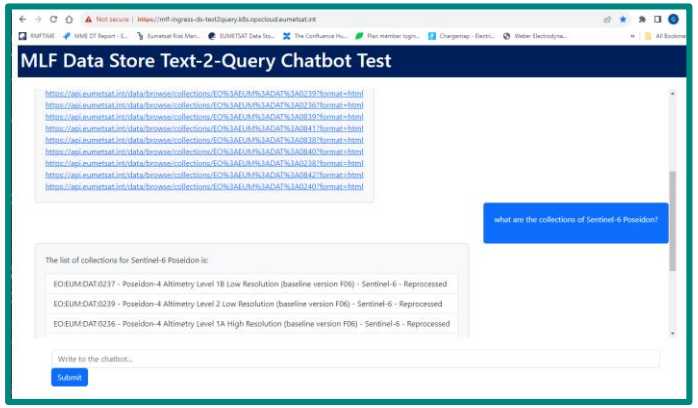




ML Framework AI/ML Apps Example – Chatbots

- Data Access Support Chatbot
- Prompts engineering and Retrieval Augmented Generation for introducing domain specific knowledge
  - User support web pages content export directly used
  - “Text-based” product access capability also supported
- Client-server LLM (API based) in K8S
  - Local LLM instance shared by various applications
  - REST API interface
  - Optimised use of HW resources, with LLM deployments on GPU and apps front-end on CPU





Summary and Way Forwards

- EUM actively engaged in developing AI/ML applications for supporting internal activities
- ML Framework project focused on supporting teams for AI/ML use cases proof-of-concept and on defining an AI/ML architecture for a future operational platform
- MLF Pathfinder phase is currently running 7 AI/ML with others planned in 2024/25, particularly in the “hot” area of Large Language Models, and in improving AI/ML apps results usability from end-user perspective, bridging the gap between raw AI/ML model results and user operational needs

Thank you!

Questions are welcome.