



MACHINE LEARNING

# ORIGIN AI

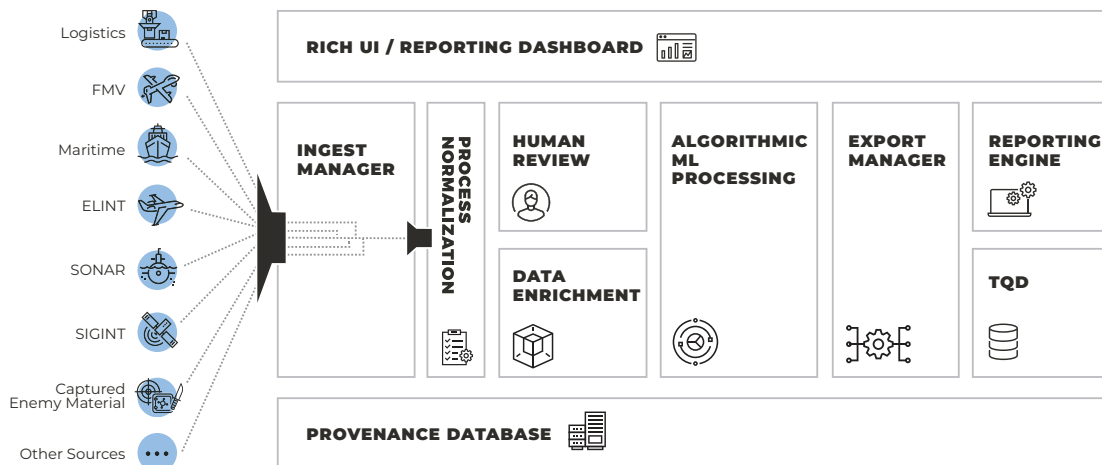
Produce training quality data for machine learning models from video, sensor data, and imagery.

OriginAI is a machine learning-enhanced data pipeline designed for large volumes of sensor and imagery data to deliver a scalable, modular, extensible, and infrastructure-agnostic framework for producing high-quality training data.

The ability to quickly convert raw source material into superior training quality data (TQD) is a critical enabler of the development and advancement of artificial intelligence/machine learning (AI/ML) in defense applications.

OriginAI supports production of TQD from full motion video (FMV) & wide area motion imagery (WAMI). It ingests, organizes, curates, processes, declassifies, and exports FMV & WAMI for use by the AI/ML community. OriginAI is adaptable to future National Mission Initiatives (NMI) & Component Mission Initiatives (CMI), utilizing cloud object storage and Docker container-based services.

## ORIGIN AI ARCHITECTURE



[NOVETTA.COM](http://NOVETTA.COM) | (571) 282-3000 | [contact@novetta.com](mailto:contact@novetta.com)

This content is proprietary and confidential information of Novetta. It is not intended to be distributed to any third party without the written consent of Novetta.

**KEY ORIGIN AI CHARACTERISTICS****RESPONSIVENESS**

The requirements for training data change faster than the models themselves. A DevSecOps CI/CD pipeline allows for near-immediate updates based on new requirements or analyst feedback.

**SCALABILITY**

As AI in the defense applications becomes widespread, the demand for training and re-training data will increase. OriginAI scales to meet high-volume ingestion demands while maintaining data provenance across repositories.

**FLEXIBILITY**

OriginAI's container-based architecture enables deployment across virtualized, on-premises, cloud, and hybrid environments.

**OWNERSHIP**

OriginAI enables mission owners to create TQD that meet their specific requirements, tracking provenance throughout the pipeline.

**EFFICIENCY**

Human analysis and adjudication is the most time-consuming aspect of developing TQD. OriginAI has incorporated ML techniques to draw analyst's attention to the elements of the data that require the most intense adjudication. ML increases efficiency while enabling humans to ensure high quality.

**DATA TYPES****FMV**

- Electro Optical
- Infrared
- Tactical
- UAV
- Med
- Alt
- P8
- Maritime

**WAMI**

- Electro Optical
- Infrared
- HADR
- Near-Real Time

**ADDITIONAL SENSORS**

- Synthetic Aperture Radar
- High Altitude
- Hyperspectral Imaging
- Mini-Drone (GBOSS) (IVAS)
- Handheld
- Ground Motion Target Indicator

**ORIGIN AI PROCESS FOR CREATING TRAINING QUALITY DATA**

- Identify data to meet algorithm objectives and customer requirements.
- Catalog data by content, type, quantity, architecture, and format.
- Triage data by removing non-conforming file formats; evaluate for content richness against the environment and mission, separating rich data from sparse data.
- Evaluate and characterize data for the environment and first-level ontology class to facilitate identification of datasets that meet testing or training requirements.
- Scrub information that if unintentionally released could have negative customer impacts.
- Structure data for optimal training requirements for labelling tools and ML/AI algorithms.