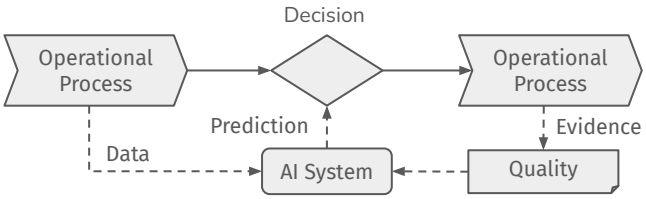
Quality assurance, monitoring and drift detection in AI systems

AI systems require constant monitoring

Quality assurance for AI

In general, we define "good quality" as the ability of a system to meet its functional requirements and user expectations. For AI systems, quality is largely determined by the data they rely on—both during model training and throughout their operational life cycle. Even small, statistically significant changes in the data can noticeably impact the AI's performance. As a result, the quality of an AI system can degrade unexpectedly and rapidly.

Ongoing concern



AI systems are designed to support and automate operational processes by leveraging real-time data to make statistically driven decisions. In practice, the statistical properties of data are constantly evolving - often due to changing user behavior. As a result, the quality of an AI system must be continuously monitored to detect and address changes early. Even minor data deficiencies, if statistically significant, can lead to subtle performance degradation, while more severe data issues can cause a complete system failure.

IT monitoring is not sufficient

AI monitoring encompasses both functional performance (e.g., forecast accuracy) and technical operability (e.g., system availability). While IT monitoring ensures technical operability, it is insufficient for tracking the functional performance of an AI system.

Maintaining and restoring the functionality of an AI system requires in-depth knowledge of data flows, data content, and its relationship to the AI model. Therefore, AI monitoring is primarily a data science responsibility, conducted in close collaboration with the business side. Quality assurance for an AI system is not suitable for delegation to IT operations alone.

Key quality metrics

The quality of an AI system is essentially measured by its **accuracy** and **error rate**. Correctness measures the function, e.g. prediction of user behaviour, against the observed reality, e.g. effective user behaviour. The error rate measures how often the AI system is wrong, e.g. predicts user behaviour that cannot be observed later. These metrics can be used to make statements about the **reliability** of the system.

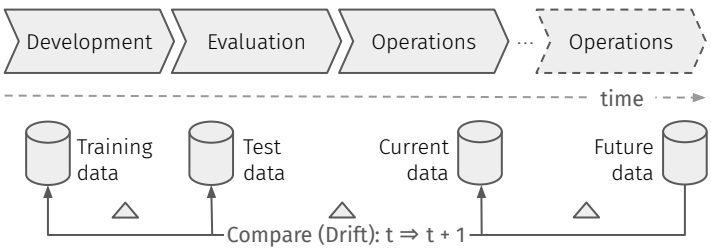
Aspects of AI monitoring

AI systems exhibit three typical changes that influence their functionality over time. Statistical tests that measure the degree of change between two points in time and in relation to a quality aspect of the AI system are used to recognise these changes: Data, Model, Metric. If a change is statistically significant, this is referred to as drift.

Aspect	Statistically significant change
Data Drift	Input Data (features)
Model Drift	Results (predictions)
Metric Drift	Accuracy, Error Rate

AI Monitoring is a process

The quality assurance of an AI system covers the entire life cycle of the system. Statistical ‘snapshots’ of the data used and the ML model used are recorded at all times. If the deviations from the expected values are too large, corrective measures are initiated, such as retraining the models.



MLOps requirements

omega-ml ✅

The MLOps platform is designed to automatically monitor all aspects of quality, recognise drift and alert those responsible