

**Service ID** S00262

**Location** Remote



## Monitoring of deployed AI models for agrifood applications

### Provider service

GRADIANT

### Link to content

<https://agrifoodtef.eu/services/monitoring-deployed-ai-models-agrifood-applications>

### Type of Sector

Arable farming, Food processing, Greenhouse, Horticulture, Livestock farming, Tree Crops, Viticulture

### Accepted type of products

Design / Documentation, Software or AI model

### Type of service

Performance evaluation

### Description

Our monitoring service provides comprehensive visibility into your own current model deployments. Through an intuitive dashboard interface, stakeholders can track and analyse their model's performance in real-time. The service continuously monitors three critical aspects: computational resource utilisation (such as processing power and memory usage), incoming data streams and patterns, and the model's output results and accuracy. This visibility allows teams to proactively identify potential bottlenecks, optimise resource allocation, and ensure the model performs as expected. The dashboard presents complex technical metrics in an easy-to-understand visual format, enabling both technical and non-technical team members to make informed decisions about their deployment. By providing these monitoring capabilities, the service helps maintain optimal performance and reliability of your model within the test environment while reducing the time needed to identify and resolve potential issues. The deployment of an AI model for testing purposes could also be done by Gradient on another service (see related services) and then be monitored through this service.

## How can the service help you

Before implementing this monitoring service, teams often struggle with limited visibility into their own AI model's behaviour within their test or production environment.

They had to manually piece together information about resource usage, data flows, and model outputs, which was time-consuming and could lead to delayed problem detection. After deploying our monitoring service, you gain immediate insights through a centralised dashboard that transforms complex technical data into actionable information. For instance, if your model suddenly requires more computational resources than usual, you'll know it immediately, rather than discovering it through system slowdowns.

If the input data patterns change unexpectedly, you'll see this reflected in real-time visualisations instead of finding out through degraded model performance. The service particularly helps technical leads and project managers who need to ensure their models perform reliably.

Rather than waiting for end-of-day reports or investigating issues after they've impacted performance, teams can now

## How the service will be delivered

The monitoring service offers various customisation options to align with your specific needs. You can set custom thresholds for different performance indicators and trigger alerts when these thresholds are crossed. These notifications can be delivered through email or integration with common messaging platforms.

Alert severity levels and notification recipients can be tailored to match your team's escalation procedures. The service requires that the compute infrastructure in which your model is deployed have connectivity with the monitoring service deploying location. If your model uses proprietary monitoring interfaces, some additional integration work may be necessary.

Additionally, while the service can track most common performance metrics, specialised or highly custom metrics may require additional development time and could incur extra costs. For security reasons, certain customisations involving direct access to the underlying infrastructure or raw system data are not available. However, our team can work with you to find alternative solutions that meet your monitoring requirements while maintaining system security and stability.

## Service customisation

The monitoring service is delivered as a solution that tracks the performance of your already-deployed AI model while your AI application is being tested and validated. In this service, Gradient provides the monitoring capabilities; the actual deployment and hosting of your AI model must be handled by you or through another service (see related services).

Implementation of the monitoring system typically takes 2-3 business days, which includes dashboard setup, integration with your existing model deployment, and initial configuration based on your specific monitoring needs. To begin service implementation, customers need to provide basic information about their model deployment, including access credentials, any specific metrics they want to monitor, and their preferred alert thresholds.

Our team will work with you to ensure all necessary permissions and configurations are properly set up. The primary output you receive is access to a real-time monitoring dashboard, customised to your needs.