



# Automated Response Evaluation for Large-Scale AI Training

## Challenge

- Our client needed to manage **70,000** annotations with detailed analysis of long-form AI-generated prompts and responses.
- Word or Excel proved inadequate for handling the prompt lengths.

Do these challenges look familiar when training an LLM?



## Solution

Argos Multilingual created the **Response Quality Assessor**, a platform that:



Unified interface for prompts, responses, and metrics.



Automated quality checks for consistent annotations.



Quick management of large-scale tasks.



One-click task distribution for project managers.

## Result

- Streamlined assessments with reduced oversight.
- Increased speed, accuracy, and scalability.
- Centralized evaluation improved consistency.
- Managed **70,000** tasks with **10-12** quality checks.
- Scalable for multiple linguists and project management.

