

Module 1: Evaluatie van agenten

Agent Evaluatie & Security

Systematisch testen van agent-gedrag: van ad-hoc testing tot geautomatiseerde eval-pipelines.

Lessen in deze module

Waarom agenten moeilijk te testen zijn 20 min [video]

Non-determinisme, lange output en subjectieve kwaliteit: waarom traditionele testing niet werkt voor agenten.

Eval-frameworks en LLM-as-judge 38 min [video]

Automatische evaluatie met LLM-as-judge, pairwise comparison en rubric-based scoring.

Testsets en regressietests bouwen 41 min [exercise]

Hands-on: bouw een testset met edge cases en implementeer automatische regressietests.

Leerdoelen

- Non-determinisme, lange output en subjectieve kwaliteit: waarom traditionele testing niet werkt voor agenten.
- Automatische evaluatie met LLM-as-judge, pairwise comparison en rubric-based scoring.
- Hands-on: bouw een testset met edge cases en implementeer automatische regressietests.

