AI Evals for Engineers & PMs

How do you know your AI actually works?

Find failure patterns, build evaluations you can trust, and improve accuracy, reliability and cost.

November: 21 Nov - 18 Dec 2026 / **EUR 500**

December: 19 Dec 2026 - 15 Jan 2027 / **EUR 1,000**

Saturdays 10:00-12:00 / Chisinau (UTC+2) / English

Included in your enrollment

- Live sessions and 10+ hours of live office hours.
- Lifetime access to course recordings, materials and private Discord.
- Future live editions of this course at no extra course fee.
- 200+ page course reader, prompts, datasets and evaluation code.
- Six assignments with worked solutions and recorded walkthroughs.
- Completion certificate after instructor review of required coursework.

With Ion Mosnoi

11+ years in AI/ML. Production AI work with startups and enterprises; research in multi-agent coordination and computer vision in France and Finland.

Plan around 3-5 hours weekly for core sessions and practice, plus optional office hours. Saturday lectures: 10:00-12:00, Europe/Chisinau (UTC+2). Office-hour times are arranged separately.

WEEK 01

AI fundamentals & build an agent

- Understand tokens, context limits, model behavior and structured outputs.
- Build an agent loop with tool calls, stop conditions and bounded retries.
- Define allowed actions, permissions, clarification and human escalation.
- Record messages, tool inputs and outputs; learn why final answers alone are not enough to evaluate an agent.

Lab outcome: Build a working agent and inspect its first complete trace.

WEEK 02

Error analysis

- Build a trace viewer and review examples with a domain expert.
- Label Good/Bad with a specific reason and the step where the error happened.
- Cluster review notes into error types and prioritize by frequency and impact.
- Define one pass/fail check per error; use code for objective rules and judges where interpretation is needed.

Lab outcome: Review a varied set of traces, create an error taxonomy and write the rules your evals will check.

WEEK 03

Create synthetic evaluation data

- Generate cases from features, roles and scenarios, including ambiguity, missing data and multi-step requests.
- Review synthetic questions, expected behavior and source evidence with a domain expert.
- Separate train, dev and test cases; check coverage by error type and category.
- Calibrate LLM judges against expert labels using false positives, false negatives, precision and recall.

Lab outcome: Build a versioned evaluation dataset and test whether your judge catches the mistakes an expert sees.

WEEK 04

Improve your AI agent

- Use the error clusters to decide whether to change data, retrieval, prompts, models or tool flow.
- Try better examples, clearer instructions, query rewriting, routing or smaller task scopes.
- Compare changes on the same cases. Measure fixes, regressions, cost and latency.
- Test prompt injection, permissions and unsafe actions; automate evals in CI and monitor production drift.

Lab outcome: Run a before/after experiment and explain what improved, what regressed and what to try next.

Practice, support and next steps.

Six assignments

- Inspect an agent trace and make its behavior visible.
- Review traces, write expert notes and group errors.
- Create a synthetic dataset with reviewed labels and safe splits.
- Write one binary evaluator per error.
- Calibrate a judge against domain-expert labels.
- Run a controlled improvement experiment and define release checks.

Who this course is for

Engineers, product managers, domain experts and technical leaders. Basic Python is useful for coding labs; PMs can pair. Use an authorized, anonymized work project or the supplied practice materials.

Team pricing and access

Enroll 2-20 participants together for 10% off. Lifetime access applies to the course/community service; future editions require registration with Ion. Full terms are available before payment.

Earn your completion certificate

Submit at least four assignments, including the final improvement report, for instructor review. This is a completion certificate, not an accredited qualification.

Enroll or ask a question

smartaitomation.com/ai-evals-course
imosnoi@smartaitomation.com