

**Industries:** EdTech, online learning and professional upskilling

**Region:** United Kingdom

## PROJECT TYPE

AI proof of concept, early-stage feasibility, and ROI validation

## TECHNOLOGIES

Python, PyTorch, Scikit-learn, Hugging Face Transformers (DeBERTa), Sentence-transformers, GPT-4 (API), pandas, NumPy, Jupyter, Streamlit, AWS

## DURATION

4 weeks

## TEAM

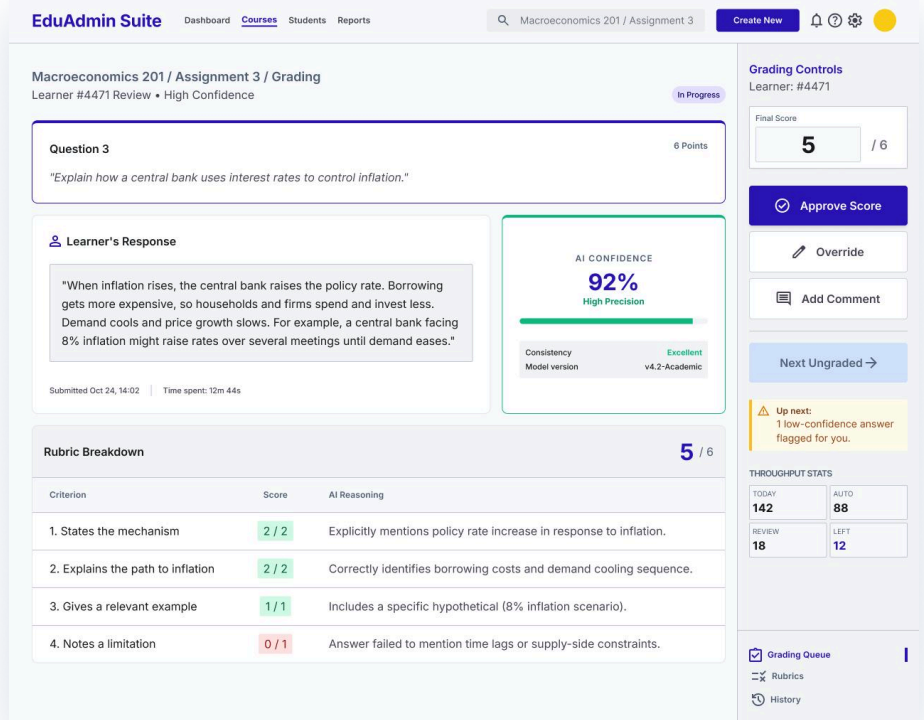
- 1 Lead Data Scientist
- 1 ML Engineer
- 1 Data Engineer
- 1 Project Manager

# An AI proof-of-concept for adaptive learning and automated grading

A UK online learning platform delivers academic and professional courses to learners across the country, with open-text assignments and quizzes at the heart of each course. As enrollment grew, manual grading became a bottleneck, and learners kept following one fixed path regardless of level. Before committing to a full AI rebuild, the platform asked SumatoSoft to prove, on its own data, that machine learning could grade open answers accurately and personalize the path each learner takes. SumatoSoft ran a four-week proof of concept on a single course, and the result sent the client into a full production build.

## Project background

The platform's problem was growth stalling due to the manual process. Each course relied on open-text assignments that a human had to read and score, and as enrollment climbed, the grading queue grew. Turnaround slowed, and graders strained. The cost of scoring rose with every new cohort.



**EduAdmin Suite** Dashboard Courses Students Reports  [Create New](#)

Macroeconomics 201 / Assignment 3 / Grading  
Learner #4471 Review • High Confidence In Progress

**Question 3** 6 Points  
"Explain how a central bank uses interest rates to control inflation."

**Learner's Response**

"When inflation rises, the central bank raises the policy rate. Borrowing gets more expensive, so households and firms spend and invest less. Demand cools and price growth slows. For example, a central bank facing 8% inflation might raise rates over several meetings until demand eases."

Submitted Oct 24, 14:02 | Time spent: 12m 44s

**AI CONFIDENCE**  
**92%**  
High Precision

Consistency Model version Excellent v4.2-Academic

**Grading Controls**  
Learner: #4471

Final Score **5** / 6

[Approve Score](#)

[Override](#)

[Add Comment](#)

[Next Ungraded →](#)

**Rubric Breakdown** 5 / 6

Criterion	Score	AI Reasoning
1. States the mechanism	2 / 2	Explicitly mentions policy rate increase in response to inflation.
2. Explains the path to inflation	2 / 2	Correctly identifies borrowing costs and demand cooling sequence.
3. Gives a relevant example	1 / 1	Includes a specific hypothetical (8% inflation scenario).
4. Notes a limitation	0 / 1	Answer failed to mention time lags or supply-side constraints.

**THROUGHPUT STATS**

TODAY	AUTO
142	88
REVIEW	LEFT
18	12

Grading Queue |  Rubrics |  History

**Up next:** 1 low-confidence answer flagged for you.

Personalization was the second gap. Every learner moved through the same fixed sequence, regardless of their starting level, and completion rates had plateaued. Leadership believed AI could address both problems and was weighing a full rebuild, but the budget was large, and the team had no proof that grading would be accurate or fair, or that personalization would change the numbers. They asked SumatoSoft for evidence before the commitment, on their own data, within a tight window.

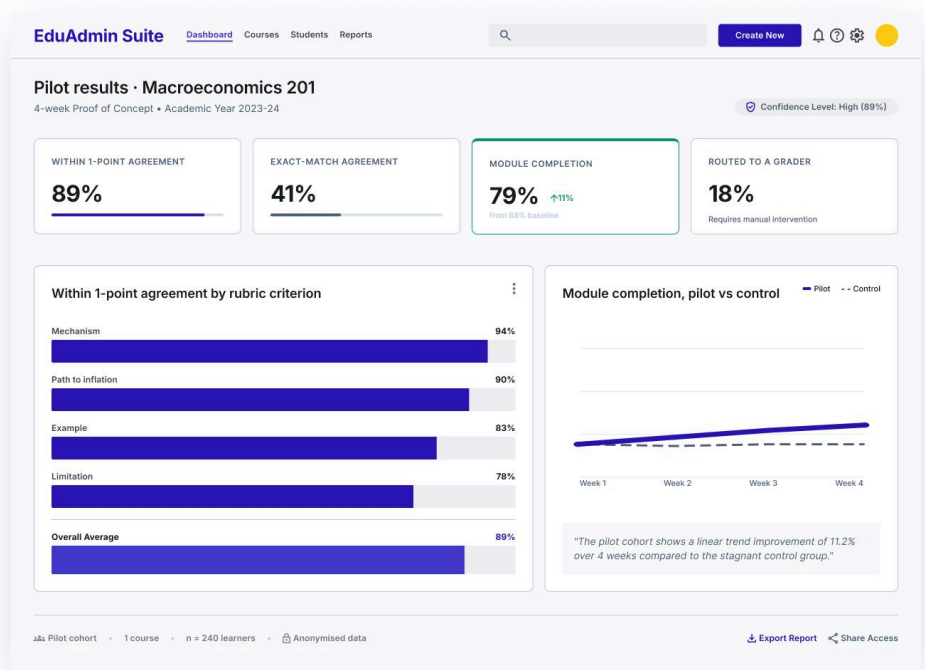
## Project Distinctive Features

- ✓ Tested on the platform's own anonymized student data
- ✓ Measured AI grading against the platform's human graders, with agreement metrics agreed before the work started
- ✓ Kept a human in the loop on every low-confidence score
- ✓ Scoped to four weeks, so the cost of finding out stayed low
- ✓ Delivered a costed production roadmap alongside the evaluation

## Business challenge

Before it committed budget to a full AI rebuild, the platform wanted SumatoSoft to prove that machine learning would work on its courses. Two questions had to be answered. Could a model grade short open-text answers closely enough to a human to take routine work off the queue, and could a model personalize each learner's path in a way that lifted completion? The aim was a clear decision at the end, not a demo.

Four conditions framed the work. The proof of concept had to run on the platform's own anonymized student data rather than a public dataset. Auto-grading had to be measured against the platform's own human graders, on metrics agreed before the work started. Student data had to stay private and compliant throughout. And the engagement had to end in a go or no-go recommendation backed by cost projections, delivered inside a tight window.



## Our solution

SumatoSoft scoped the work to one course and cohort, with two questions. Could a model grade short open-text answers closely enough to a human to take work off the queue, and could a model personalize each learner's path to lift completion? The team agreed on success metrics with the client up front, so the work would end in a decision rather than a demo.

For grading, SumatoSoft fine-tuned a transformer on the platform's human-graded answers and scored it on a held-out set of graders' marks, measuring agreement with within-one-point accuracy and quadratic weighted kappa. The team also compared it against a large language model prompted with the same rubric, and routed low-confidence scores to a human. The grader matched the human score within one point 89% of the time, while exact agreement sat near 40%, enough to prove auto-grading viable while keeping a person on the edge cases.

For personalization, the team built a knowledge-tracing model that estimates what a learner has mastered from their diagnostic and quiz responses and adjusts the difficulty of the next content. Compared with the pilot cohort, the adaptive path increased completion from about 68% to 79% and shortened the time to reach the same checkpoint.

Everything ran in a private sandbox on anonymized data under GDPR controls. SumatoSoft handed over an evaluation report, a sandbox for the client, a costed production roadmap, and a recommendation, along with the data and rubric issues to fix before a full build.

## Customer's benefits

The proof of concept turned a large, unproven bet into a measured decision. On grading, the platform saw that a model could match its human scores within one point about 89% of the time on the pilot course, enough to take the routine answers off the queue while a person handled the rest. On personalization, module completion on the pilot cohort rose from about 68% to 79%. Beyond the numbers, the client got a sandbox its team had used, a costed plan, a short list of fixes, and the confidence to proceed. The four-week spend was small against the cost of the full build it was deciding on, and the result justified moving ahead.

## What's happening with the project right now?

After the proof of concept, the platform engaged SumatoSoft for the full build. The production work brings auto-grading and adaptive paths across the catalog and adds analytics dashboards for instructors. A governed AI tutor that stays inside each course's material is rolling out alongside them. SumatoSoft and the platform are taking it course by course, starting with the highest-volume programs.