

Using Generative AI to Give Feedback on Mathematical Proofs

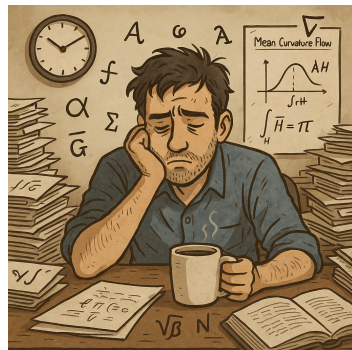
Aron Gohr, Marie-Amelie Lawn

Kevin Gao, Stephen Heslip, Inigo Serjeant

October 31, 2025

The Problem

- First-year students often struggle with proof-writing.
- Feedback on proofs is hard to scale: it needs judgement, attention, and consistency.
- In large cohorts, feedback can easily become delayed or uneven.



Our aim

Use generative AI to provide useful formative feedback on undergraduate mathematical proofs.

What We Built

- We built a **multi-stage workflow**, not a single prompt.
- It uses the question, course context, and an exemplary solution.
- One stage analyses the student's proof; a later stage writes feedback.
- The workflow is modular, so stages can be tuned separately and fixed parts can be precomputed.

Prompt example

```
Write an exemplary solution to: {question} → {solution}
Rewrite for clarity and conciseness: {solution} → {solution}
Evaluate student work against solution: {input} vs {solution} →
{feedback}
Give concise, helpful feedback on: {question}, {input}, and
{course context} based on {feedback} → {feedback}
```

Initial Evaluation

- We first tested the system on **Introduction to University Mathematics** at Imperial.
- 65 question/answer pairs with IUM questions and natural errors in the answers
- Each submission was graded independently by three human markers.
- We compared LLM feedback with human marking and human-written feedback.
- The structured workflow outperformed generic prompting and was often strong on detailed error detection.
- The main weaknesses were verbosity and overcriticality-now mostly fixed!

From Pilot to Real Use

- What began as an evaluation project has now been used for **two years**.
- Over time, we became much more confident that it gives genuinely useful support.
- The approach has now been **generalised across all Pure Year 1 lectures**.

What changed?

The key point is not only that the system worked in principle, but that it proved useful enough to sustain and extend in real teaching.

Embedding the Workflow in Lambda Feedback

- We embedded the workflow in **Lambda Feedback**, the platform used to deliver feedback to students.
- This moved us from an experimental pipeline to a tool inside existing teaching infrastructure.
- Since our questions were fixed, we could precompute model solutions.
- That made deployment much more practical and helped scale feedback across courses.
- The platform also imposed constraints: no dialogue, and handwriting required a separate workflow.

Significance

The project was not only about prompt design, but also about deployment in a form that was actually usable in practice.

Takeaway

- Generative AI can give useful formative feedback on undergraduate proofs.
- Workflow design matters: structure, context, and deployment all make a big difference.
- After two years of use, we are now much more positive about its value.
- The right picture is not replacement of human feedback, but scalable support for students and teaching teams.

Thank you – questions?



<https://d2ff9b6c8a2179ba11.gradio.live/>