

Hope this prompt finds you well: An Exam Using AI

Amedeo Andriollo
Finance Department (DRM)
Université Paris Dauphine - PSL.

June 11th, 2026

A Challenge

- Starting point: Business Analytics with Python, Fall 2025.
 - Stick to the norm: **take-home** single exam, within 7 days.
 - ↪ Somewhat “useless” course: limited class participation, an exam solved with AIs **w/out learning**.
 - Why this outcome (=limited learning value with the traditional format)?
 - The delivering/teaching skills? The information? Maybe...
 - ↪ Also a single “trivial” exam: tasks too **easy to prompt** that requires **low level of auditing**.
 - Educators should not have the “**luxury**” to ignore **AI**, as students use AI:
 - During their studies (e.g., coding, personalized notes, routine tasks).
 - After their studies: many jobs implicitly require them to use AI.
 - Fresh start: Empirical Asset Pricing, Winter 2026.
 - Use of AI needs to be incorporated, but it is not core of the course (time constrained)...
...however, part of the course was about coding with Python, thus an opportunity:
- ↪ An **in-class** exam allowing AI use: **prompt optimization** with a **high level of auditing**.

Designing the Exam

- Tasks to evaluate: transferring financial knowledge to empirical application.
 - In practice, two dimensions: coding and output interpretation (6Qs in total).
 - Innovation as double-edge sword: open book/code & limited use of AIs.
 - Questions/tasks need to be **tractable and replicable** for all studs:
 - Extensive margins: AI users and non-AI users.
 - Intensive margins: human and machine logic.
 - **Trade-offs**:
 - difficulty/time vs. optionality of using AI.
 - AI complementarity vs. substitutability.
- ↪ Calculator analogy: evaluating math intelligence shifted from arithmetic to algebra.
- How to restrict the AI usage?
 - **Time constraint**: tighter as mostly AI speeds up productivity.
 - **How much** to use it: number of characters per prompt & number of prompts.
 - ↪ AI usage optimization problem.
 - **Evaluation and replicability** (to some degree): prompts/conversations need to be recorded.
 - ↪ AI usage (and inputs) graded as well.

The In-Class Exam (1)

- The typical set of exam submissions:
 1. Python script + written answers, all in the same file.
 2. A markdown `.md` file containing the full AI conversation log used during the exam.
 3. The file(s) uploaded to the AI.
- AI usage rules:
 - Students can use **at most 5 prompts** in total: 1 mandatory setup prompt, 3 prompts, and 1 mandatory transcript prompt.
 - Each prompt may contain **at most 600 characters**.
 - Students can upload to the chat only text or Py files (e.g., `.txt`, `.md`, or `.py`). Uploads of other file formats are not allowed (e.g., `.pdf`).
 - The Markdown transcript and the uploaded files are part of the exam materials.
 - Students' code and written answers must remain coherent with their submitted AI transcript. Any use of undisclosed AI chats, accounts, or devices is prohibited. If the submitted materials are inconsistent, the exam may be referred for academic integrity review and/or followed by a short oral verification.

The In-Class Exam (2)

- Examples of the two mandatory prompts:

- **First prompt:**

“For this exam, follow these instructions for the next 4 prompts in this chat. Use only the information contained in this chat and in any file I upload here. Do not rely on memory, prior conversations, custom instructions, or personal context outside this chat. Do not assume knowledge of my previous work unless it appears in this chat or in the uploaded files. Keep answers concise and task-focused. When I later ask for the transcript, reproduce all prompts and responses from this chat exactly, in chronological order, without omission, rewriting, summarization, reordering, or cleanup.”

- **Fifth and final prompt:**

“Produce a markdown transcript of this chat for my exam submission. I need to download it as a .md file. Include all prompts and responses in chronological order, with timestamps (if available). Do not summarize or rewrite. After the transcript, list all uploaded files shown in this chat and reproduce their full contents exactly. If needed, split the output into consecutive markdown parts labeled Part 1, Part 2, etc. Return only the markdown content. Please create the .md file.”

Aftermath

- Problems during the exam: **compliance with the rules** (of course!).
 - Multiple chats and/or devices/files (e.g., cellphone, .png).
 - Paid vs. Unpaid version.
 - Manual modification of the conversation transcript.
 - About the students:
 - AI tools did not always converge in time to deliver usable answers.
 - No teaching materials (specifically) on AI **covered before** the exam.
 - Much time was spent **auditing and converting** AI outputs into coherent human answers.
 - Grades were, though, generally good (median: 11/15, normally distr.)!
- ↪ Has **inequality** between “good” and “bad” students increased? Yes, but not trivially.
- How to prepare properly next time:
 - More time.
 - Enforcing **uniform** compliance (especially, paid vs. unpaid is tricky).
 - Homework for the educator: more transparent and less arbitrary grading criteria.
 - **Forcing to use a certain model/version of AIs** → Viable also to **learn how to audit!**

Thanks for your attention!
And thanks a lot to my students!!