

Forming Computational Social Scientists in the Era of Generative AI

Dr Jon Cardoso-Silva

 J.Cardoso-Silva@lse.ac.uk



Data Science
Institute

19 May 2026

The GENIAL study

(Part 1/4)

When do Generative AI tools act as a catalyst for learning?

Who I am and what I teach




- I am a Assistant Professor (Education) based at the **Data Science Institute** * at the London School of Economics and Political Science(LSE).
- I teach data science courses that involve programming, web scraping, data engineering, version control.
- My pedagogical approach is for project-based learning, continuous feedback cycles, and authentic assignments - everything is assessed through coursework, with no exams.

LSE AI and Education Fellow (2025–2027)

I am one of the 10 [LSE Fellows in AI and Education](#), a programme with an ambitious goal to test out how to embed Generative AI in the teaching & learning practices of our disciplines.

* (*I'm moving to the LSE Department of Methodology as an Associate Professor in September*)

Data from our study (2023-2025)

Although the opinions are mine, some of the data and preliminary findings come from the  **GENIAL** study.

- A **collaborative project** across several LSE departments
 - Data Science Institute
 - Department of Statistics
 - Department of Management
 - School of Public Policy
- **Timeline:** July 2023 to April 2024
- **Funding:** internal (LSE Eden Centre for Educational Enhancement, LSE Data Science Institute)

We asked students to share their **chat logs** and in the case of my data science courses, I also collected the **git histories** of their assignments.



Participating courses

Case study	Autumn Term (Sep–Dec 2023)	Winter Term (Jan–Mar 2024)
Undergraduate courses	DS105A – Data for Data Science Quant	DS105W – Data for Data Science Quant
	DS202A – Data Science for Social Scientists Quant	DS202W – Data Science for Social Scientists Quant
	ST207 – Databases Quant	MG317 – Leading Organisational Change Qual
Postgraduate courses	—	ST456 – Deep Learning Quant
	—	PP422 – Data Science for Public Policy Quant
	—	MG4B7 – Leading Organisational Change Qual

Cohorts: 48 active participants (out of 200+) / ~160 active participants (out of 300+)

Attitudes: Tools

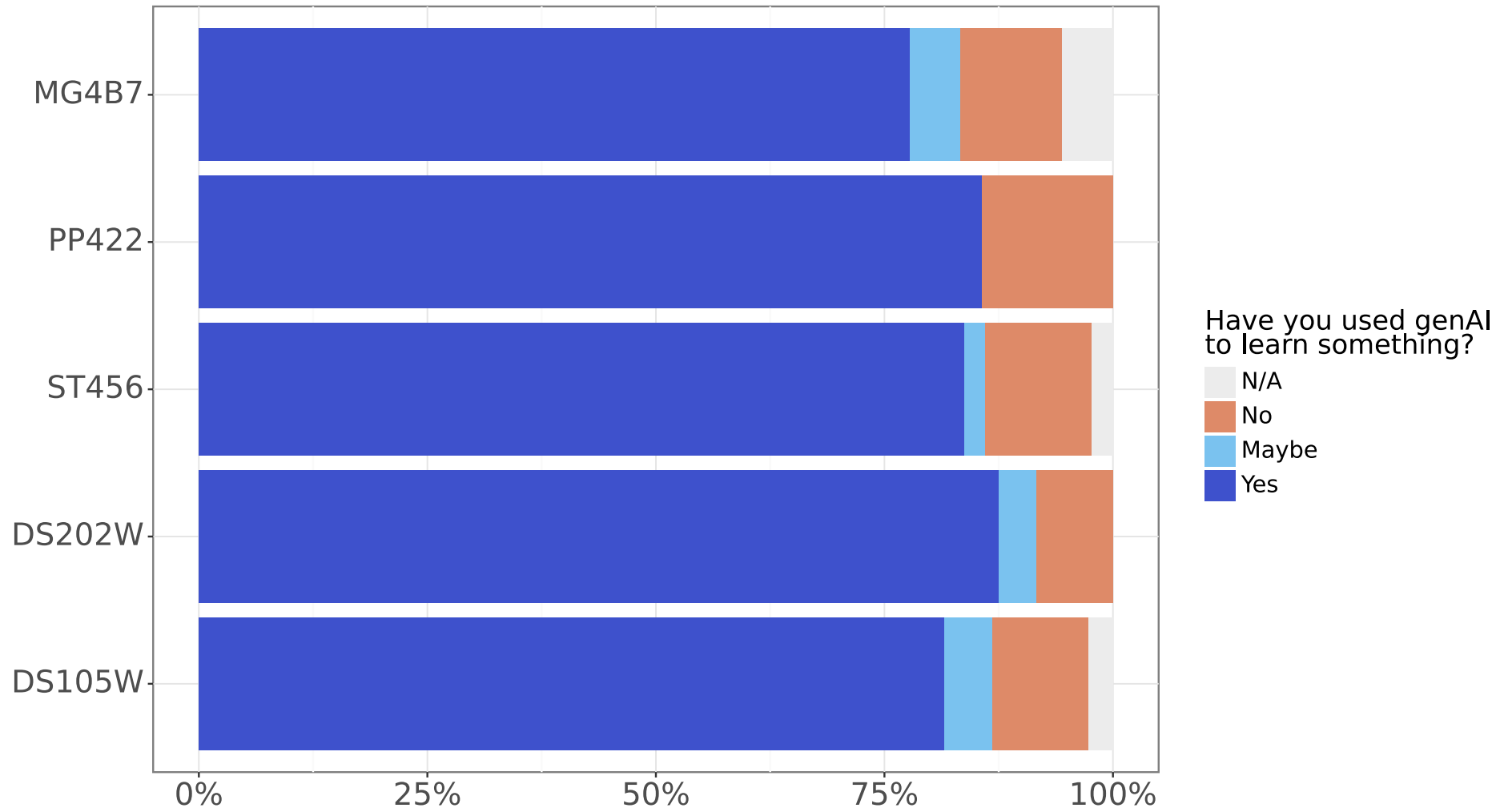
	DS105A		DS105W		DS202A		DS202W		ST207		MG317	
	Heard about it	Have used it	Heard about it	Have used it	Heard about it	Have used it	Heard about it	Have used it	Heard about it	Have used it	Heard about it	Have used it
ChatGPT	16	15	38	37	21	19	24	22	17	17	11	11
Grammarly	16	8	37	23	21	15	22	14	17	11	12	11
Bing AI	9	5	15	3	15	4	15	4	9	7	4	1
DALL·E	7	4	13	6	9	5	10	7	9	5	6	3
Bard			15	7	12	2	14	4	7	3	4	3
GitHub Copilot			14	2	13	1	17	8	11	5	2	1
Midjourney	9	3	7	3	5	1	7	3	5	2	2	2
Claude					2	1					4	2
Consensus.ai					1	1						
Quillbot					1	1						
Stable Diffusion	1	1										

- ChatGPT dominated as the most used tool.
- Not every participant filled out the initial survey

Attitudes: AI for learning?

~80% of students reported having used a genAI tool to learn something

The proportions of 'Yes' are similar across courses

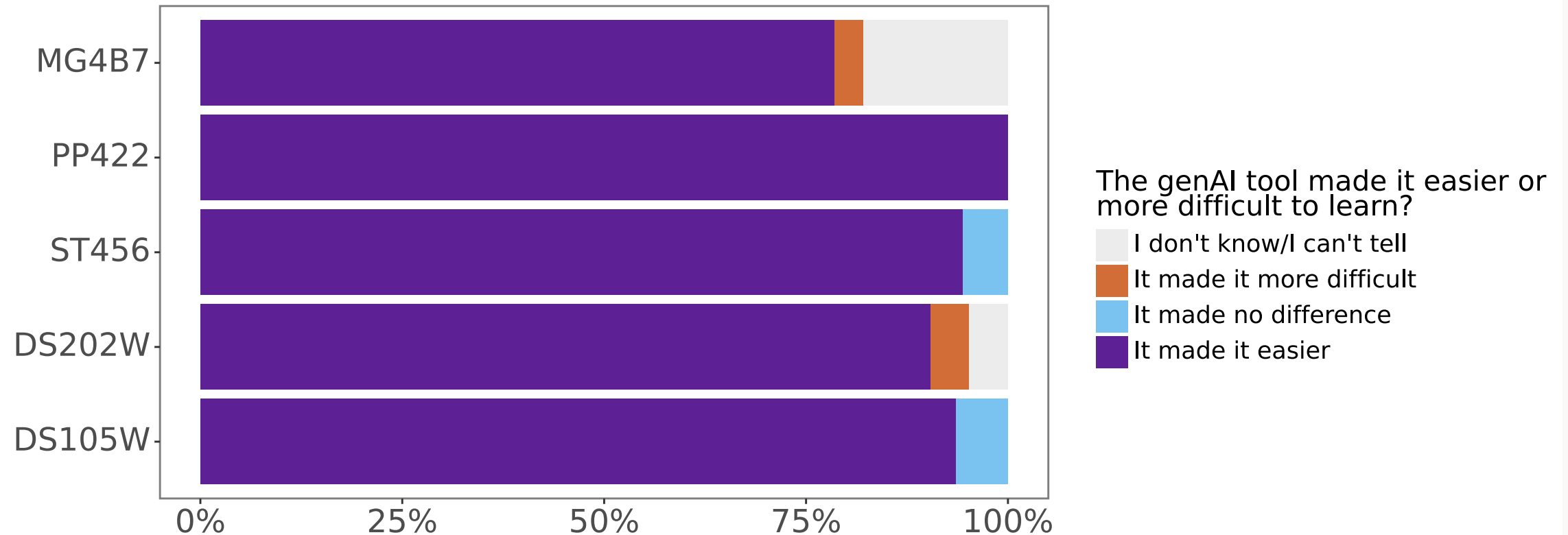


PP422 has fewer students

Attitudes: Any good?

Of those who reported using genAI for learning, most found it useful

GenAI use is similar across courses



Note: For students enrolled in multiple courses, we allocated them to just one of their reported courses.

Students are very optimistic about GenAI and the tools had already been seemingly fantastically helpful for their learning.

There's nothing to learn from such a biased sample, right?

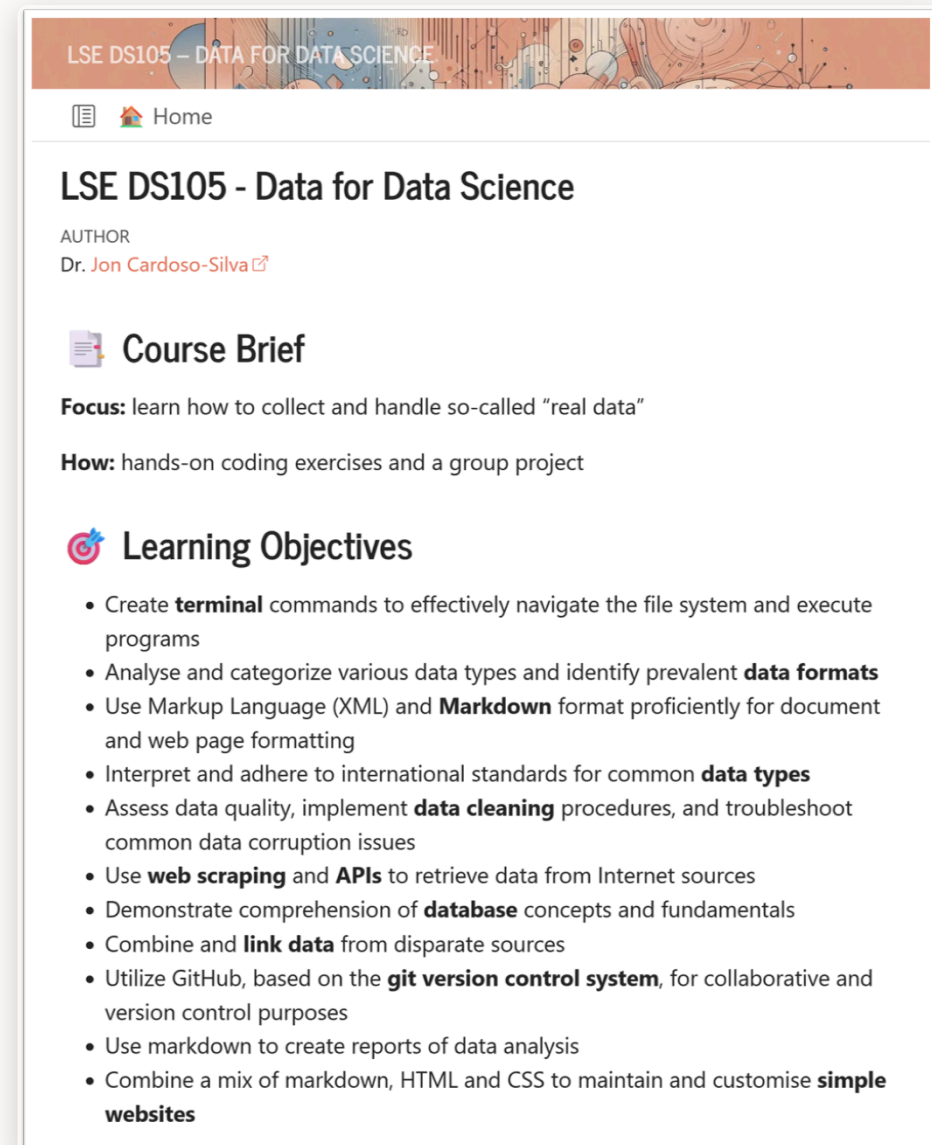
Well..

Not every student uses GenAI productively (Part 2/4)

A bit of context about the DS105 course

- Lots of coding 🧑💻
- As you may know, programming is an iterative process and frustration is a **desirable difficulty** we want our students to experience.
- I want students to:

1. Understand the problem
2. Break it into smaller, actionable pieces
3. Figure out how to start
4. Write a first draft
5. Test it out
6. Fix 'bugs'
7. Figure out what is the next step
8. Go back to Step 3



LSE DS105 – DATA FOR DATA SCIENCE

Home

LSE DS105 - Data for Data Science

AUTHOR
Dr. [Jon Cardoso-Silva](#)

Course Brief

Focus: learn how to collect and handle so-called "real data"

How: hands-on coding exercises and a group project

Learning Objectives

- Create **terminal** commands to effectively navigate the file system and execute programs
- Analyse and categorize various data types and identify prevalent **data formats**
- Use Markup Language (XML) and **Markdown** format proficiently for document and web page formatting
- Interpret and adhere to international standards for common **data types**
- Assess data quality, implement **data cleaning** procedures, and troubleshoot common data corruption issues
- Use **web scraping** and **APIs** to retrieve data from Internet sources
- Demonstrate comprehension of **database** concepts and fundamentals
- Combine and **link data** from disparate sources
- Utilize GitHub, based on the **git version control system**, for collaborative and version control purposes
- Use markdown to create reports of data analysis
- Combine a mix of markdown, HTML and CSS to maintain and customise **simple websites**

Source: LSE DS105 website (lse-dsi.github.io/DS105)

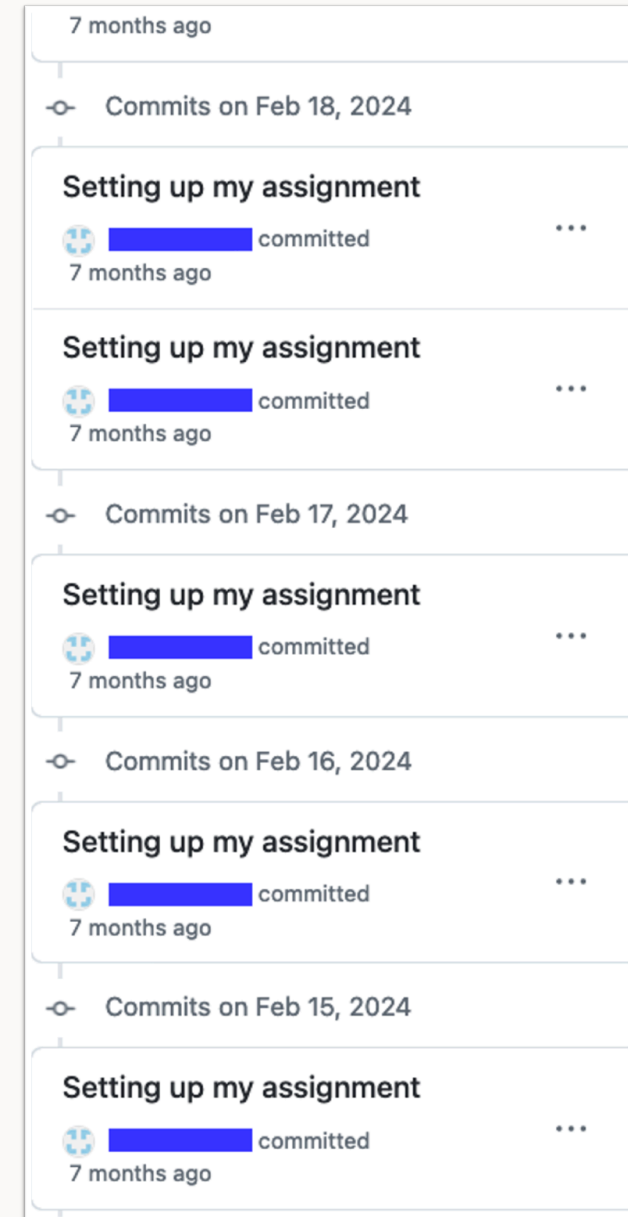
Two students, same assignment, same tool

	Student A	Student B
Background	2nd Year BSc, International Social and Public Policy	2nd Year BSc, Economics
Prior coding	Took the Python pre-sessional (struggled with it)	Had prior experience with Python
How they used ChatGPT	Used ChatGPT to build the solution for the assignment	First reviewed the week's content with ChatGPT, then asked for help

How Student A documented their drafts

Student A's logs

- All draft updates had the same description 🙌
- The student's metacognition process was not engaged. The purpose of the git commits was not salient to them and became a mechanical step to complete.



How Student B documented their drafts

Student B's logs

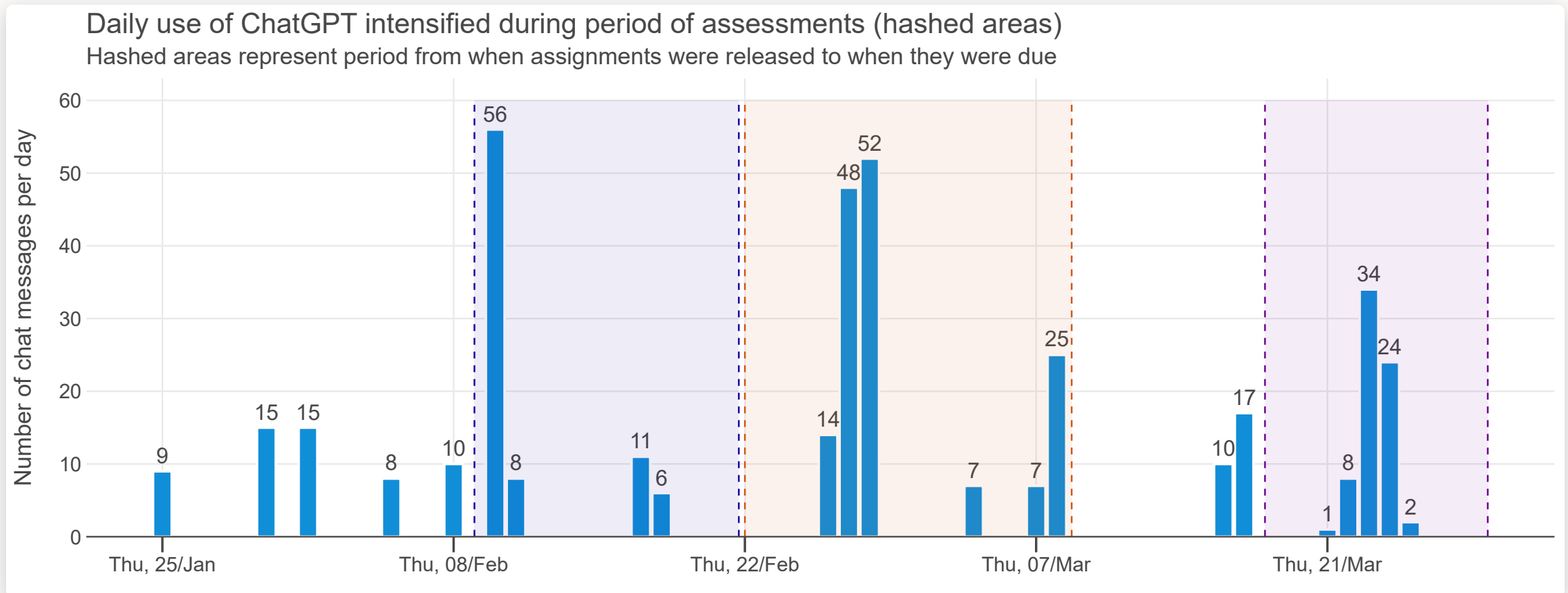
- Added meaningful descriptions to their commits
- We know what their purpose was with every update
- This makes it possible to decode their thinking process more easily

The screenshot displays a commit log with the following entries:

- Update README.md after submission 1**
Verified
[redacted] committed
7 months ago
- submit first draft (can probably just treat it as finished) of assignment**
[redacted] committed
7 months ago
- finish task 2**
[redacted] committed
7 months ago
- complete task 1**
[redacted] committed
7 months ago
- setting up my unique assignment**
[redacted] committed
7 months ago
- set up candidate number**
[redacted] committed
7 months ago

A section header "Commits on Feb 10, 2024" is visible between the second and third entries.

Student B was a Resourceful user of AI




- Note also that the use *intensifies* when there is an assignment. This was common across ALL participants.

Student B challenged the AI

Student B's logs

- ChatGPT often produced overcomplicated code
- The student went back to the teaching material and told the AI:

“your stuff is not in line with what I learned”
- They were in control of their learning

 **USER said:** (2024-02-10 18:45:07.971380949)


ok, how do i get that code to output the title of the webpage?

 **ASSISTANT:**

Comment: Here, ChatGPT's output is a modified version of a complex code it had previously produced, with accompanying commentary. The code it produces will run without any errors, but just as before, it won't return anything because it is in the format of a Python class, not a 'runnable' code like the way we taught in the course.

Importantly, this time, the final sentences in the output clarified that because this is a Python class, this type of code cannot be run as you would any regular script on a Jupyter Notebook. Instead, one would have to create a specific file and run it from the terminal (not how we've been teaching).

It seems unlikely that the student tried that suggestion. A little over 2 minutes later, there is another question:

 **USER said:** (2024-02-10 18:47:27.346262932)

`sel.css('title')` worked when the other stuff didn't

This is very interesting, and it only makes sense in the context of the course! The student provided a piece of code adapted from the teaching material, saying that it worked, whereas the complex code produced by ChatGPT didn't.

 **ASSISTANT:**

From now on, ChatGPT will stop suggesting `Scrapy` classes. Instead, it will use the style of coding suggested by the user, which aligns with the material taught in the course.

Student A was misled

The task involved writing code to:

- ~~Navigate to a page~~
- ~~Grab the entire content of that page~~
- Select just the relevant information
- Clean it up
- Store it

```
In [546...
# URL AND ALT LINKS - ALUMNI
alumni1 = ''
alumnisoup3 = BeautifulSoup(alumni3, 'html.parser')
alumni_tags3 = alumnisoup3.select(img_selector1)
for img_tag in alumni_tags3:
    alumni3URL = img_tag.get('src')
    alumni3alt = img_tag.get('alt')
```

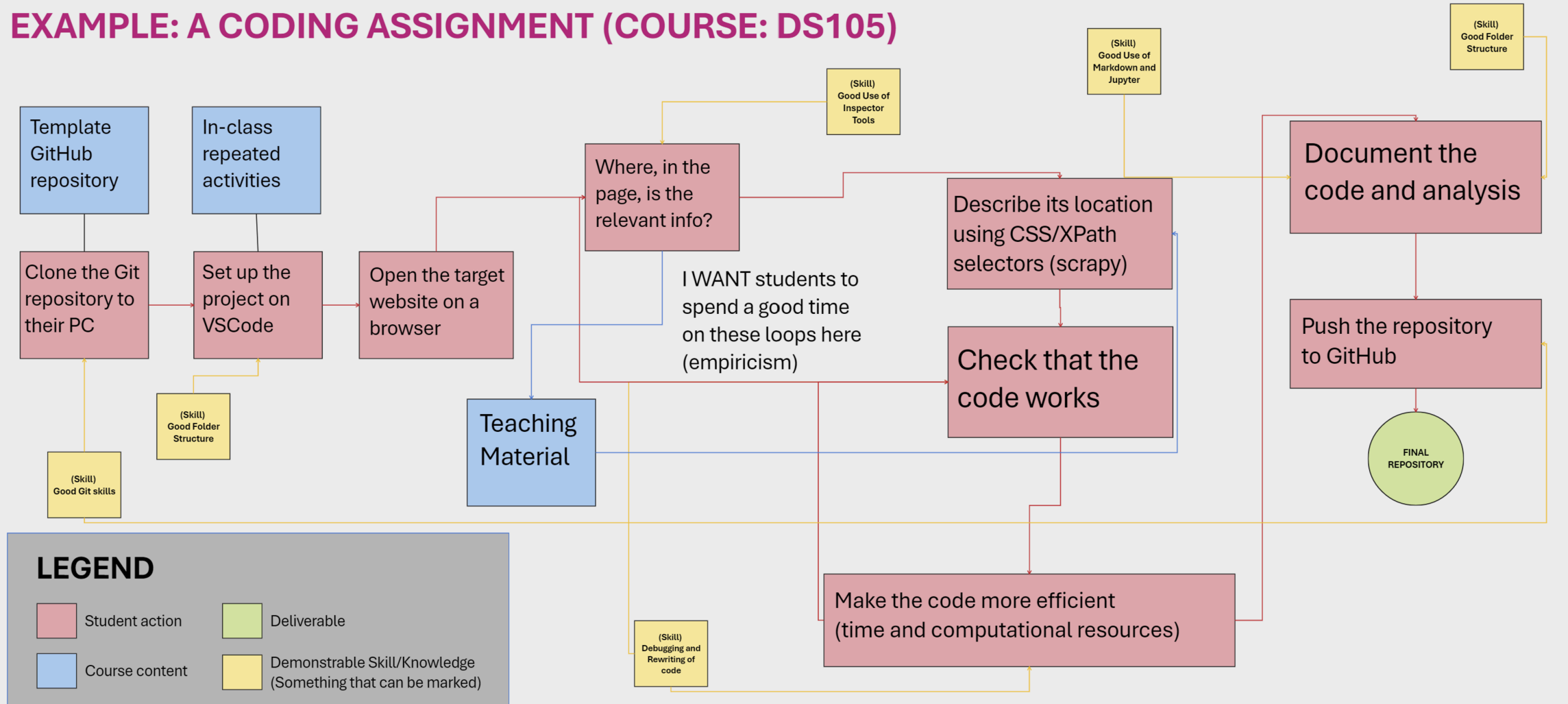
Student A missed two crucial steps. They were misled by an inaccurate use of GenAI, clearly swayed by the chatbot's authoritative tone.

But Student A improved!

After receiving feedback on their poor performance and on their unhelpful use of GenAI, **Student A improved tremendously**, scoring scores equivalent to distinction scores on the subsequent two graded assignments.

Process-based assignments

EXAMPLE: A CODING ASSIGNMENT (COURSE: DS105)



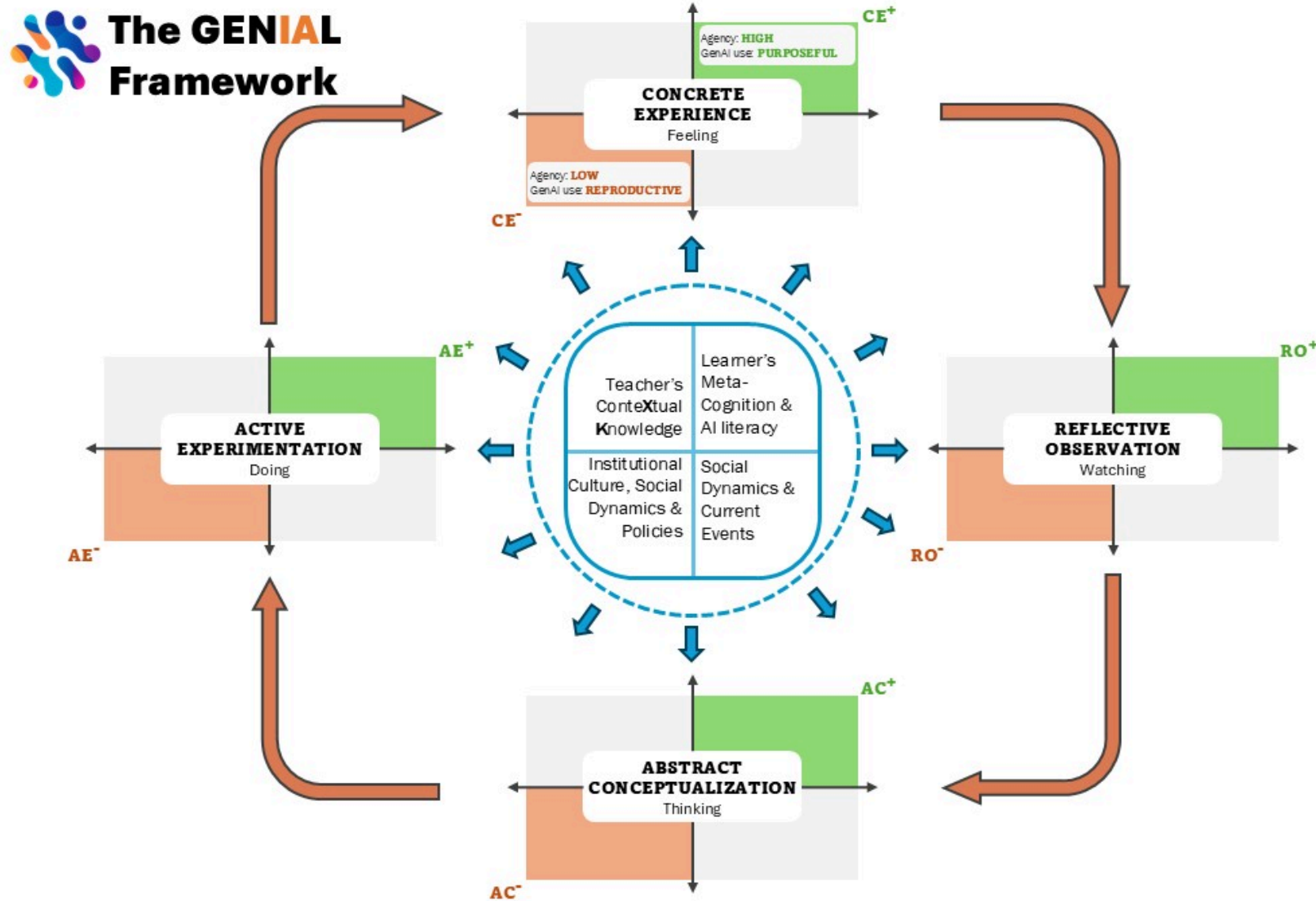
Bridging the Divide

Despite what I had initially thought, the pedagogical tools I adopted in the course had not failed me. They were actually what helped Student A!

- Even if not apparent to students, the mapping and the **process-driven approach** to the assessment helped me identify more easily where the student had gone on a wrong path.
- The **continuous feedback** mechanism helped them course-correct more effectively and still learn in time. The student even engaged more in class (they were someone who *wanted* to learn!)
- I can now use what I have learned from this interaction and conduct a **backwards (re)design** of the assignment again, for the future.
- The process of marking student work becomes a rewarding task of discovery and less of a chore (*but it's still very laborious though...*)

What to build from that (Part 3/4)

As an educational researcher



Maps student interactions with GenAI onto **Kolb's experiential learning cycle**.

Each stage is coded for whether the student:

- Exercised agency (+)
- Had the stage disrupted (-)
- Skipped it entirely

Five engagement patterns along an agency spectrum:

Resistive → Receptive → Resourceful → Reflective → Riffing

Student A's traces code as Receptive. Student B's code as Resourceful.

Two levels of coding

Level 1: low inference

Codes each exchange on observable features:

- What did the student ask? (9 prompt types, from verbatim paste to metacognitive reflection)
- What did the AI reply? (code, explanation, redirect, clarification)
- What did the student do with it? (adopted, modified, overrode, ignored)
- How much time passed before the next action?

Level 2: high inference

Groups exchanges into learning cycles (one sub-task) and asks:

- Did the student try something before consulting the AI? (**CE**)
- Did they evaluate the AI's response? (**RO**)
- Did they connect it to course material? (**AC**)
- Did they produce and test their own version? (**AE**)

Each stage gets a quality label: **+**, **-**, or **skip**.

Default for any unobserved stage is "skipped." The coder needs positive evidence to code it otherwise.

Hopes for reliability

Two coders. Training round on students 1–5, calibration round on students 6–10. Target $\kappa > 0.60$.

The best comparable in the published literature (Oliveira et al.'s DRIVE framework) reports $\kappa = 0.44$ on its hardest categories.

The research assistant codes Level 1 and does not know the grades. The derived variables that enter the regression come from his codes, not mine.

I code Level 2. The dual role (instructor and researcher) is documented.

A new cohort

	DS105W (Data Science)	MG317 (Management)
Exam?	No	No
Assessment	Coursework only	Coursework only
AI access	Enterprise Claude (LSE) + personal tools	Enterprise Claude (LSE) + personal tools
Process data	Chat logs, git histories, reflections	Chat logs, reflections

- I have collected data from a new cohort in the Winter term of 2025/26 (Jan-Mar 2026). We'll soon start the systematic coding of that data.
- Goals: quantitative analysis of the nature of the interactions now that we have a reliable coding scheme, and deep forensic analysis of the most interesting cases.

83
consenting
students in
DS105W
(80% of
cohort)

53 shared
chat logs

9,721
student-AI
exchanges

1,689 git
commits

Three submission points across the term (W04, W06, W11)

What I am building next (Jul-Dec 2026)

- What I'm building next in the second strand of the fellowship: a system that **proactively reaches out to students** when they look stuck, rather than waiting to be asked.
- If the tutor works and grades go up, those increases look identical to grade inflation from the outside. Without process data, you cannot tell “the tutor helped them learn” from “the tutor helped them perform.” I will have to think about what to do about grade inflation.

What the literature is saying (Part 4/4)

What evidence do we have?

GenAI Creates Performance Paradoxes

Bastani et al. (2025): Students using unstructured GPT-4 showed **48% improved practice scores** but **17% decreased exam performance**.

The 'quick wins' obtained earlier in the learning journey might not translate to *real*/deep learning. In fact, it might make it worse. This is what happened to Student A.

Contextual Pressures Drive Problematic Dependency

Abbas et al. (2024): **Time pressure** and **academic workload** are significant predictors of ChatGPT dependency.

When under pressure, students might be more likely to resort to a more 'Receptive' style of engagement with GenAI.

Usage Patterns Determine Learning Outcomes

Lehmann & Cornelius (2024): **Substitutive use** increases coverage of material but decreases understanding. **Complementary use** does the opposite.

The authors argue that it's *how* one uses the tools that matter.

Two recent meta-analyses

Deng et al. (2025) on 69 ChatGPT studies

Pooled effect on academic performance: **$g = 0.71$**

Of 51 studies in that estimate:

- 9 allowed ChatGPT during the post-test
- **33 did not report** the post-test condition
- 9 prohibited it

The positive finding may reflect ChatGPT's output quality, not student learning (Yan et al., 2025)

Maier et al. (2026) on programming, pre-registered

10 studies measuring exam scores after a GenAI-assisted learning phase: **$g = 0.14$ (n.s.)**

Exam-environment moderator (~50% of variance):

- AI **not available** during exam: **$g = -0.06$**
- AI **available** during exam: **$g = 0.76$**

The two results agree once you ask what was being measured.

Half a million grades tell the same story

Chirikov (2026): 507,076 grades across 319 courses at a US research university, 2018 to 2025.

Writing and coding courses saw a **13 percentage point** increase in A grades after ChatGPT's release (~30% above 2022 baseline)

A triple-differences design ties the effect to courses where homework counted for more. Where homework weight was low, the effect was near zero.

If students were learning more, the improvement should have appeared in supervised exams too. It did not.

Chirikov (2026) calls this **task displacement**: AI improved what students submitted without improving what students knew. This is the effect of seeing Student A at institutional scale.

Converging evidence from other fields

Study	Domain	Finding
Bastani et al. (2025)	Maths (RCT, ~1000 students)	GPT Base group scored 17% worse on exam
Fan et al. (2025)	Essay writing	Better essays, same knowledge test scores
Kosmyrna et al. (2025)	Essay writing (EEG)	Up to 55% less neural connectivity
Shaw & Nave (2026)	Reasoning tasks	Followed wrong AI ~80% of the time
Akgun & Toker (2025)	Factual recall	ChatGPT advantage at 3 weeks: gone

These come from cognitive psychology, learning science, neuroscience, labour economics, and programming education, and none of them cite each other.

These problems predate ChatGPT

Bransford & Schwartz (1999)

Distinguished two ways to measure what education produces:

1. **Sequestered problem solving**: isolate the learner, remove all resources, test recall. Most exams work this way.

vs

2. **Preparation for future learning**: watch how students approach new problems. Do they ask better questions? Improve when given a chance to revise?

The gap between what students hand in and what they retain has existed for decades.

Bjork & Bjork (1992)

Conditions that **slow practice down** produce better long-term retention.

Three mechanisms:

- spacing
- retrieval effort
- interleaving

GenAI removes all three 🙄. It answers on demand, provides complete solutions, and narrows the interaction to a single loop.

Students feel fluent, but the fluency comes from the tool rather than from anything they will retain.

What I am betting on

- Assessment is the major lever to shape student behaviour. If we want to encourage productive use of GenAI, we need to redesign our assessments so they reward it.
- I don't like the idea of exams (“sequestered problem solving”). Instead I favour instruments that are more aligned with the authentic practices of the discipline being taught.

Higher-order tasks

AI substitution fails on more complex work. On Akgun & Toker's (2025) hardest task (evaluate a policy document, propose improvements), ChatGPT provided no measurable benefit and AI-generated content dropped from 72% to 27%.

Observed components

One option: a short live coding session. A student who built the pipeline can modify it on the spot; a student who pasted output has to reconstruct what the code does first.

Rubrics that reward the process

DS105W uses criterion-based rubrics (Pass / Good / Really Good / WOW) covering technical judgement, interpretation, and communication alongside code correctness.

References

- Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, 21(1), 10.
<https://doi.org/10.1186/s41239-024-00444-7>
- Akgun, M., & Toker, S. (2025). Short-Term Gains, Long-Term Gaps: The Impact of GenAI and Search Technologies on Retention. *arXiv.org*.
<https://doi.org/10.48550/ARXIV.2507.07357>
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakçı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), e2422633122.
<https://doi.org/10.1073/pnas.2422633122>
- Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. K. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of william k. estes* (Vol. 2, pp. 35–67). Erlbaum.
- Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In *Review of research in education* (Vol. 24, pp. 61–100). American Educational Research Association.
<https://doi.org/10.2307/1167267>

Thank you

1.  CARDOSO-SILVA et al. (2025). The GENIAL Framework.

Mapping Student-GenAI Interactions onto Experiential Learning: The GENIAL Framework

Jonathan Cardoso-Silva^a, Dorottya Sallai^b, Casey Kearney^c, Francesca Panero^{d,e}, Marcos E. Barreto^a

^aData Science Institute, London School of Economics and Political Science, London, UK;

^bDepartment of Management, London School of Economics and Political Science, London, UK;

^cSchool of Public Policy, London School of Economics and Political Science, London, UK;

^dDepartment of Methods and Models for Economy, Territory and Finance, Sapienza University of Rome, Italy;

^eDepartment of Statistics, London School of Economics and Political Science, London, UK

ABSTRACT

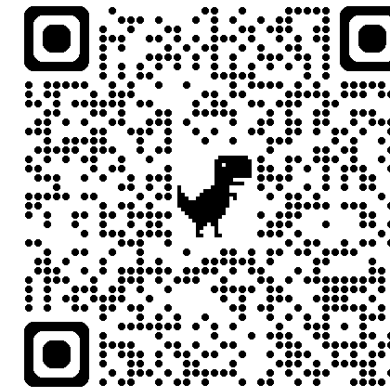
This study proposes a conceptual framework for understanding how generative artificial intelligence (GenAI) affects university students' learning journey. The use of GenAI tools in academic settings left scholars at a lack of theoretical tools to capture the impact of how chatbots impact learning. Current studies rely heavily on self-reports and controlled experiments, missing how students engage with chatbots while learning. Building on literature on experiential learning and styles of student engagement with GenAI as theoretical underpinnings, the GENIAL framework sheds light on the dynamics of AI-human interaction in all stages of the experiential learning cycle, allowing us to identify how the cycle of learning is affected by the use of GenAI tools. Our findings, based on the analysis of 200 course

Cardoso-Silva, J., Sallai, D., Kearney, C., Panero, F., & Barreto, M. E. (2025). Mapping Student-GenAI Interactions onto Experiential Learning: The GENIAL Framework. *SSRN Electronic Journal*. (Under Review)



Sallai, D., Cardoso-Silva, J., et al. (2024). Approach Generative AI Tools Proactively or Risk Bypassing the Learning Process in Higher Education. *LSE Public Policy Review*, 3(3), 7.

Blog series:
jonjoncardoso.github.io/blog



 J.Cardoso-Silva@lse.ac.uk