INTERPLAY BETWEEN EXTERNAL AND INTERNAL FACTORS IN THE USE OF GENAI FOR THE CONSTRUCTION OF PROOFS: SHORTS IN THE ESSENTIAL ALGEBRAIC CYCLE

Francesco Contel and Annalisa Cusi

Sapienza University of Rome

In this paper, we explore secondary students' interaction with GPT-4o in the context of activities of proof construction through algebraic language with the aim of investigating how the interplay between internal and external factors affects the represent-manipulate-interpret essential algebraic cycle. A qualitative micro-analysis of chats, students' interviews and their written work was conducted. The findings show that bias and hallucinations in the GenAI tool interact with students' epistemic beliefs, producing different forms of short circuits of the algebraic cycle. Implications are raised for the design of GenAI systems that support algebraic reasoning as meaning-making rather than procedural execution.

INTRODUCTION

Generative artificial intelligence (GenAI) tools are increasingly explored in research and instructional practices as means to support teaching and learning activities. The present study belongs to a broader line of research devoted to the investigation of human-GenAI interactions in the activity of mathematical problem solving. Within this perspective, the student is conceived as a user of the GenAI resource, and the GenAI resource is conceived as an artefact whose purpose is to scaffold the user's mathematical processes. Previous studies have stressed how complex these interactions are on both the user and the GenAI's end, focusing on the impact of students' beliefs (Yoon et al., 2024) and perceptions (Zhuang, 2025) on their use of GenAI tools and on the severe limitations of these tools in providing students with support both at cognitive and metacognitive level (Contel & Cusi, 2025). In this paper we further explore these aspects by focusing on the interplay between internal and external factors in driving human-GenAI interactions. In particular, we analyse how this interplay affects students' conjecturing and proving activities using algebra as a reasoning tool.

THEORETICAL FRAMEWORK

The theoretical framework consists of two components. The first one is the construct of instrumental genesis (IG) (Rabardel, 2022), which we use to model the human-GenAI interaction. The novelty of mathematical activities mediated by GenAI artefacts brought us to reconceptualise the process of IG in this context: in Cusi and Contel (2026) we proposed to conceive IG as a process of noise reduction, adopting the physical process of Brownian motion as a metaphor to represent this kind of dynamics.

2026. In H. Krzywacki, M. Hähkiöniemi, E. Haataja, M.S. Hannula & P. Hästö (Eds.). *Proceedings of the 49th Conference of the International Group for the Psychology of Mathematics Education: Research Reports*, pp. 291-298. PME.

According to this metaphor, we interpret IG as a process subject to both deterministic (the drift) and stochastic factors (the noise). Within this reconceptualised model, we distinguish between internal and external factors (with reference to the user's perspective) affecting the IG process. *Internal factors* could act both as noise and drift terms in the use of GenAI to support mathematical problem solving (Cusi & Contel, 2026). On one hand, users' beliefs about mathematics and about GenAI tools could act as noise terms, negatively affecting their interpretation of the task they are facing and of the support they receive. On the other hand, internal factors such as the users' metacognitive competences could act as *drift* terms in supporting the user's monitoring of their interaction with the GenAI. The *external factors* that we consider in this study are GenAI's bias and hallucinations, which mainly act as noise terms in the IG process. Recent studies have highlighted the role played by GenAI's biased representation of mathematical knowledge and practice (Frieder et al., 2025). As regards *hallucinations*, namely the production of seemingly plausible yet factually unsupported content, we distinguish between *factuality hallucination*, associated to a discrepancy between generated content and verifiable facts, and *faithfulness hallucination*, associated to a generated content that diverges from the user's input or to a lack of self-consistency within the generated content (Huang et al., 2024).

The second component of our framework is constituted by the tools that supports the analysis of students' processes when they face proving tasks that require them to explore a situation, develop conjectures, and justify them through algebraic language. Bell's (1996) notion of the *essential algebraic cycle* provides an effective representation of the three coordinated processes involved in this kind of activity: *representing* a situation algebraically, *manipulating* symbolic expressions, and *interpreting* the resulting expressions in relation to the proving goal. To analyse the processes involved in the essential algebraic cycle, we use two theoretical lenses. The first lens concerns the notion of *conceptual frames* (Arzarello et al., 2001), understood as organised sets of knowledge, expectations, and strategies that govern how algebraic expressions are produced, transformed, and interpreted. Coordinating appropriate conceptual frames supports the use of algebraic language as a thinking tool: it enables students to interpret algebraic expressions within the context of the proving activity, while also guiding symbolic manipulation toward mathematically meaningful goals. This latter aspect connects directly with the second lens, concerning the role of *anticipatory thinking* (Boero, 2001), defined as the process through which problem solvers foresee the intermediate or final symbolic expressions that would best serve the proving goal and orient their transformations accordingly. Anticipating what would count as a useful symbolic expression supports efficient transformations and informs re-interpretation of results.

Given this framework, our research question addresses how the interplay between external and internal factors interfere with the processes involved in the essential algebraic cycle: *How does the interplay between GenAI's bias and hallucinations (external factors) and students' beliefs about the use of algebraic language for proving*

(internal factors) affect the realisation of the essential algebraic cycle when using algebra as a tool for proving?

RESEARCH CONTEXT

In the data collection process, we involved 19 grade 9 and 10 students. The participants were volunteers, informed by their teachers and by the researchers of the purpose of the study. Each participant worked individually on two tasks with the support of GPT-4o. Researchers prompted the chatbot asking it to act as a tutor supporting the students' metacognitive processes in problem solving activities involving the use of algebraic language as a thinking tool. The users were free to use the chatbot at their will, the only request being that of thinking aloud, as their interaction with GPT-4o was audio- and video-recorded. The researchers acted as non-participant observers, intervening only to elicit evidence of the students' reasoning processes or in case of serious impasse moments. Each interaction lasted roughly 100 minutes.

In the task considered in this study, students are first asked to explore the behaviour of the difference between the squares of two consecutive natural numbers by producing examples and formulating conjectures. Then they have to prove the conjectures they formulate. Among the conjectures that may reasonably emerge are "the difference is always an odd number" and "the difference is equal to the sum of the two consecutive numbers considered". To justify such conjectures, a strategy is to represent the difference between the squares of two consecutive numbers by activating the preceding-successive frame, for instance by denoting the numbers as n and n-1 and writing the difference as $n^2-(n-1)^2$. This expression must then be transformed under the guidance of clear anticipatory thinking, associated with the activation of the even-odd frame: the goal is to reach a form such as 2x+1 or 2x-1, making the oddness of the difference explicit. After having manipulated the expression obtaining 2n-1, this expression should be rewritten as n+(n-1), thereby highlighting that the difference also coincides with the sum of the two consecutive numbers.

RESEARCH METHOD

The study adopts a qualitative, micro-analytic approach to examine how the interaction between GPT-4o (in the following, GPT) and the students unfolded when facing the proving task. Three types of data were collected: (a) chat transcripts, capturing the full interaction between each student and GPT; (b) the transcripts of semi-structured interviews, conducted before, during, and after the interaction, focusing on the student's interpretation of the task, their expectations regarding the use of algebraic language, their intentions during the interaction, and their reflections on the meaning of the outcomes and on the whole interaction with GPT; and (c) students' handwritten work produced while interacting with GPT (e.g., algebraic expressions, intermediate manipulations, informal annotations).

The analysis was structured around three analytical foci: (1) the type of support provided by GPT with respect to the essential algebraic cycle (object of analysis: the

chats); (2) the role of GPT's bias and hallucinations and their effects on the proving process (object of analysis: the chats); and (3) the role of the student's beliefs and their effects on the realisation of the essential algebraic cycle (object of analysis: both the chats and the interviews). Bell's notion of the essential algebraic cycle permeated the analysis and was used to trace how the proving activity evolved. The constructs of conceptual frames (Arzarello et al., 2001) and anticipatory thinking (Boero, 2001) supported the examination of students' use of algebraic language, helping identify whether and how the represent, manipulate, and interpret processes were actually realised during the interaction and revealing points of blockage or tensions and their underlying reasons. Finally, the notions of bias and hallucination were used both to characterise the tutoring behaviour of GPT and to describe the external factors that intervened on students' processes, to highlight how such factors contributed to triggering blockages, redirecting activity, or consolidating existing epistemic beliefs.

ANALYSIS

In what follows, we present a detailed analysis of the interaction between a student, S1, and GPT. While several interactions were examined in the study, this case was selected to exemplify how the analytical framework introduced above can be applied to how external and internal factors shape students' engagement with the essential algebraic cycle. The excerpt of the chat on which we focus starts with S1 exploring the problem and developing a conjecture. She then turns to GPT for support:

3. S1: I would like to know if, taking a natural number and squaring it minus the square of the previous number, the result is equal to the sum of the two numbers.
4. GPT: Great question! To start with, how can you express "the square of a number minus the square of the previous number" in algebraic language?
5. S1: What do you mean by "express in algebraic language"?
6. GPT: I mean writing your idea using symbols and letters instead of words. For example, if we call the natural number n, how can we write the square of n and the square of the previous number?
7. S1: n = natural number c = number preceding n n^2 - c^2 = n + c
8. GPT: You've already written a good equation! Now, how could you rewrite c as a function of n, knowing that it is the preceding number?
9. S1: n = natural number c = n-1 n^2 - (n-1)^2 = n + c
10. GPT: Perfect! Now, can you try to expand the square of $(n-1)$ and simplify the expression on the left side of the equation?

[*S1 manipulates the expression constructed in this way:* $n^2-n^2+2n-1=n+c \rightarrow n=c+1$]

11. S1: So n = c + 1
12. GPT: Correct! Now expand $(n-1)^2$ and substitute it into the equation. What do you get?

[*S1 asks for clarification on the request; GPT explains that performing the requested procedure will reveal whether the equality is satisfied; S1 substitutes n-1*

into the expression n=c+1, obtaining the identity 0=0, and communicates this to GPT]

17. S1: If I substitute the expression n-1 for c, I get 0 = 0, and I don't understand how that would help me prove it.
18. GPT: Excellent observation! If you get $0 = 0$, it means that the equality is verified for every natural number n. This is a good sign that your conjecture might be correct. But before we conclude, let's go back to the key step for a moment: Did you correctly expand $(n-1)^2$ into $n^2 - 2n + 1$? If so, can you rewrite the entire equation and simplify the terms?

[*Prompted by the researcher's question, "Do you think you have proven the conjecture?", S1 declares, "Yes, because if you say that zero equals zero is a great way to prove it, then I think so!", then, when asked, "But are you convinced?", adds, "This 0=0, no, but... I was more convinced by n = c+1, if I may ask?"]*

19. S1: Are n = c+1 and 0 = 0 sufficient to prove my theory?
20. GPT: Not yet! To have a true proof, you must show step by step that your initial equation is always true for every natural number. [*GPT performs the required steps, showing that the expression n-(n-1)^2 can be transformed into 2n-1, equivalent to n+(n-1)*]

The analysis of the chat highlights an interaction in which GPT's support is characterised by a systematic bias in favour of symbolic manipulation from the outset (4). In constructing the expression representing her conjecture, S1 does not activate the previous/successive frame, generating an expression that only partially adheres to the conjecture to be represented (7).

In interventions 8, 10 and 12, GPT guides S1 to the substitution, but this indication is not accompanied by an explanation of its meaning within the problem. The misalignment between the suggested procedure and the development of the meaning of the process through which S1 is guided is exacerbated when GPT invites S1 to "simplify" by improperly using the term equation (8 then 12). This lexical choice, which is out of focus (faithfulness hallucination) with respect to the nature of the activity, leads S1 to activate procedures typical of solving equations, such as deriving one variable as a function of the other (11), shifting the focus from the structure of the relationship to the algorithmic treatment of the expression.

Throughout the chat, GPT repeatedly prompts the formal rewriting of the statement and the algebraic development of the expressions, implicitly assuming that these transformations constitute the preferred and sufficient path for the proof. This approach produces an interaction in which the manipulate phase is hyper-activated, while the interpret phase is not supported as an autonomous activity of the student, but is carried out by the tutor, who certifies the outcome (10, 12, 18). This generates what we have called a *short circuit of the essential algebraic cycle*. This short circuit becomes more evident when S1 obtains the identity. Although this is a formally correct result, it is not interpreted in relation to the initial conjecture: the student repeatedly states that she

"does not understand the meaning" and does not see how this result can contribute to a proof (17). GPT validates the outcome without questioning the process that led to it, stating that it constitutes "an excellent observation" and effectively replacing S1 in both the manipulations and the validity check (18). In this way, the essential algebraic cycle is formally completed, but an authentic interpretative process is never activated by S1.

S1's beliefs about the role of algebraic language, which emerged during the interview conducted at the end of the interaction with GPT, contribute to amplifying this short circuit. On the one hand, she attributes a greater status of "truth" and precision to letters than to numbers (in the interview she states "the letter gave me a greater sense of truth"), while on the other hand she expresses a problematic relationship with proof, which she perceives as something distinct from the "procedure" and as a critical and poorly internalised moment. In the absence of support that challenges these beliefs, GPT ends up implicitly reinforcing them. The overall result is, in fact, the consolidation of a conception of algebra as the activation of procedures and of proof as mere symbolic manipulation aimed at producing "formal clues" (n=c+1 and 0=0), to which no meanings related to the problem under consideration are attributed. During the interview, S1 states:

> "At first, I didn't understand it because I thought that 0=0 had no meaning, because I thought it was wrong as a result, maybe that I had done something wrong... then when he [GPT] explained that it was a great way to prove that the equality was verified, then I think I understood."

The role of GPT here is characterised by a procedural and authoritative bias: GPT insists on algebraic transformations as the preferred path, provides explanations oriented towards formal correctness and, in moments of uncertainty on the part of S1, intervenes by validating the outcome rather than relaunching the interpretation as an open problem. However, from the faithfulness hallucination we see that it does not even construct justifications in a coherent and mathematically rigorous language. In this way, it not only fails to reconstruct the essential algebraic cycle but also contributes to stabilising a conception of algebra as a system of correct manipulations to be accepted on the basis of authority (human or algorithmic), rather than as a language for reasoning and giving meaning to mathematical relationships.

DISCUSSION AND CONCLUSION

The analysis of the interaction between S1 and GPT revealed how the interplay between external factors stemming from the GenAI tutor and internal factors associated with students' epistemic beliefs can short-circuit the essential algebraic cycle during a proving task. The examination of further cases confirmed this tendency and provided a broader picture of the observed phenomena. We complement the observations of S1's case reporting highlights from the cases of other two students, S4 and S6, working on the same task.

In S4's case, the short in the essential algebraic cycle emerged primarily in the form of cognitive outsourcing: although the student autonomously identified and formalised a conjecture, the activation of interpreting processes remained fragile, leading GPT to progressively assume both the manipulation and the semantic validation of the symbolic expressions. The interaction was further complicated by a faithfulness hallucination in which GPT, confronted with an incorrect proof developed by S4 and presented in image form, responded with coherent discourse that ignored the content of the student's argument, ambiguously suggesting correctness. The analysis of S4-GPT interaction highlighted the effects of the interplay between different external and internal factors not only in terms of a short circuit in the essential algebraic cycle, but also in reinforcing externally-based proof-schemes (Harel & Sowder, 1998) and authority-dependent validation: (1) GenAI's biased representation of mathematical practice; (2) GenAI's limited capacity to handle multimodal data; (3) S4's beliefs about the use of algebra in developing mathematical proofs, intended in a ritualistic way; (4) S4's a priori trust in GenAI's capabilities.

In S6's case, the interaction with GPT is strongly affected by his difficulties in interpreting what the problem is asking, which lead him to ask many clarifications about the terms used, including those introduced by ChatGPT itself. This initial confusion prevents him from exploring examples to look for a pattern. ChatGPT's support, starting from the first exchange, guides him exclusively toward the algebraic representation of the statement, followed by algebraic transformations aimed at simplification. This short circuit in the essential algebraic cycle, which reduces it to a process in which representation and manipulation phase coincide while the interpretation phase is completely absent, reduces the proving activity to symbolic rewriting, treating "proof" as the derivation of an algebraic identity rather than the justification of a conjecture. Hallucinations involving misinterpretation of correct manipulations further disoriented the student, who ultimately framed proof as empirical verification of formulas obtained through symbolic manipulations, reinforcing an empirical proof scheme (Harel & Sowder, 1998). Again, it is possible to highlight the role played by external and internal factors in fostering a short circuit in the algebraic cycle and the student's development of erroneous mathematical beliefs.

These cases suggest that different forms of short circuits in the essential algebraic cycle may occur when proving is mediated by GenAI, depending on where the misalignment between external and internal factors originates. In S1 the disruption centred on the interpretation phase, in S4 on the control and validation of the proving process, and in S6 on the semantic framing of the task itself. Across these cases, GPT systematically displayed procedural and authoritative bias in its representation of proof through algebraic language, while hallucinations and lexical choices played a role in redirecting students' activity or stabilising existing beliefs about algebra and proof.

These findings open several directions for future research. From a cognitive perspective, it would be valuable to characterise more systematically the spectrum of

proving behaviours elicited by GenAI tutors and how such tools interact with students' proof-schemes and beliefs. From an educational perspective, these observations raise important questions regarding how GenAI tools should be integrated into instructional settings in ways that foster algebraic reasoning as a meaning-making activity rather than as a sequence of authorised manipulations.

References

Arzarello, F., Bazzini, L., & Chiappini, G. (2001). A model for analyzing algebraic thinking. In R. Sutherland, T. Rojano, A. Bell & R. Lins (Eds.), *Perspectives on school algebra* (pp 61–81). Kluwer.

Bell, A. (1996). Problem-Solving Approaches to Algebra: Two Aspects. In N. Bernarz, C. Kieran & L. Lee (Eds.), Approaches to Algebra.

Boero, P. (2001). Transformation and anticipation as key processes in algebraic problem solving. In R. Sutherland, T. Rojano, A. Bell & R. Lins (Eds.), *Perspectives on school algebra* (pp 99–119). Kluwer.

Contel, F., & Cusi, A. (2025). Investigating the Role of ChatGPT in Supporting Metacognitive Processes During Problem-Solving Activities. *Digital Experiences in Mathematics Education*, *11*, 167–191. https://doi.org/10.1007/s40751-024-00164-7

Cusi, A., & Contel, F. (2026). A reconceptualisation of the instrumental genesis process in the context of human-LLM interactions: exploring students' perspective on their use of GPT-4o as a tool to support self-assessment. In E. Geraniou, C. Crisan & M. Mavrikis (Eds.), *Digital Technology and Artificial Intelligence in Mathematics Education Assessment*. Routledge. https://doi.org/10.4324/9781003533764-8

Frieder, S., Bayer, J., Collins, K. M., Berner, J., Loader, J., Juhász, A., Ruehle, F., Welleck, S., Poesia, G., Griffiths, R., Weller, A., Goyal, A., Lukasiewicz, T., Freer, C., Buzzard, K. & Gowers, T. (2025). Data for mathematical copilots: Better ways of presenting proofs for machine learning. arXiv preprint arXiv:2412.15184

Harel, G., & Sowder, L. (1998). Students' proof schemes: results from exploratory studies. In E. Dubinsky, A. Shoenfeld e J. Kaput (Eds.), Research Issues in collegiate Mathematics Education III (pp. 234–283). American Mathematical Society.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B. & Liu, T. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, *43*(2), 1–55. https://doi.org/10.1145/3703155

Rabardel, P. (2002). People and technology – a cognitive approach to contemporary instruments. https://hal.archives-ouvertes.fr/hal-01020705

Yoon, H., Hwang, J., Lee, K., Roh, K. H., & Kwon, O. N. (2024). Students' use of generative artificial intelligence for proving mathematical statements. *ZDM – Mathematics* Education, 56(7), 1531–1551. https://doi.org/10.1007/s11858-024-01629-0

Zhuang, Y. (2025). Lessons from Using ChatGPT in Calculus: Insights from Two Contrasting Cases. Journal of Formative Design in Learning, 1–11.