



The International Journal of Mathematics and Sciences Education

Publication details, including instructions for authors and subscription information:

<https://nakiscience.com/index.php/IJMSEd>



The International Journal of
Mathematics and Sciences Education



Generative artificial intelligence in university mathematics teaching: Theorems-in-action inferred from practicing professors' discourse

Denilson Andrés Silgado-Tuñón^{1*}, Patricia
Sureda², José Iván López-Flores³,
Emmanuel Magallanes Ulloa⁴

¹Universidad Autónoma de Zacatecas, México-Zacatecas,
denilson.silgado@uaz.edu.mx

²Universidad Nacional del Centro de la Provincia de Buenos
Aires, Argentina-Buenos Aires,
psureda@niem.exa.unicen.edu.ar

³Universidad Autónoma de Zacatecas, México-Zacatecas,
jlopez@uaz.edu.mx

⁴Universidad Politécnica de Zacatecas, México-Zacatecas,
emagallanes@upz.edu.mx

To cite this article:

Silgado-Tuñón, D.A., Patricia Sureda, P., López-Flores, J.I., & Ulloa, E.M. (2026). Generative artificial intelligence in university mathematics teaching: Theorems-in-action inferred from practicing professors' discourse. *International Journal of Mathematics and Sciences Education*, 4(2), 110–136.

To link to this article:

<https://nakiscience.com/index.php/IJMSEd>

Published by:

Nasir Al-Kutub Indonesia

Residential Street Kila Rengganis, Block I, Number 11, Labuapi, Indonesia, 83361

Generative artificial intelligence in university mathematics teaching: Theorems-in-action inferred from practicing professors' discourse

Denilson Andrés Silgado-Tuñón^{1*}, Patricia Sureda², José Iván López-Flores³, Emmanuel Magallanes Ulloa⁴

¹Universidad Autónoma de Zacatecas, Zacatecas, México, denilson.silgado@uaz.edu.mx

²Universidad Nacional del Centro de la Provincia de Buenos Aires, Buenos Aires, Argentina, psureda@niem.exa.unicen.edu.ar

³Universidad Autónoma de Zacatecas, Zacatecas, México, jlopez@uaz.edu.mx

⁴Universidad Politécnica de Zacatecas, Zacatecas, México, emagallanes@upz.edu.mx

*Corresponding Author: denilson.silgado@uaz.edu.mx

Abstract

This article examines how university professors who teach mathematics conceptualize and use Generative Artificial Intelligence (GenAI) chatbots such as ChatGPT and Gemini, as expressed in their interview discourse. Drawing on the Theory of Conceptual Fields, the study adopts a qualitative, exploratory case-study design bounded to ten practicing professors from the academic units of Mathematics, Physics, Light and Matter Science and Technology (LUMAT), and Engineering at a public university in Mexico. Semi-structured interviews were segmented into episodes, from which theorems-in-action (TA) were inferred by one researcher, reviewed by a second researcher, and agreed upon by the research team; the resulting 68 TA were grouped into seven categories and judged as correct, incorrect, or not yet established in light of current literature on GenAI. The findings reveal the coexistence of contradictory theorems-in-action: some reflect schemes built through experience with search engines (GenAI provides correct results), while others are more appropriate for the use of these technologies (GenAI does not always respond correctly). Furthermore, although most professors report using GenAI for personal or professional tasks, its reported pedagogical integration into the classroom remains limited. These results suggest that professional development for mathematics teachers should go beyond instrumental training and explicitly address the theorems-in-action that guide the use of GenAI, particularly those concerning the verification of its outputs.

Article History

Received:

August 15, 2026

Revised:

September 24, 2026

Accepted:

October 10, 2026

Published Online:

October 12, 2026

Keywords:

Generative artificial
intelligence;
Theorems-in-action;
University mathematics
professors

1. Introduction

Generative artificial intelligence (GenAI) has moved rapidly from a specialized technology to an everyday resource in higher education. Chatbots such as ChatGPT, Gemini, Copilot, Claude, LLaMA, Grok, and DeepSeek now generate explanations, worked

solutions, and teaching materials on demand, and in doing so they reconfigure the didactic relationship between teacher, student, and knowledge (Silgado-Tuñón & López-Flores, [2025a](#), [2025b](#)). University mathematics is a particularly sensitive setting for this change, because GenAI is already reshaping teaching practices, students' modes of access to knowledge, and the ways in which non-routine problems are approached (Corica et al., [2024](#); Y. J. Lee, [2023](#); Wardat et al., [2023](#)). Its integration, however, also raises unresolved questions of ethics, equity, and teacher preparation for effective and responsible use (Silgado-Tuñón & López-Flores, [2025b](#)).

The educational potential of these tools is well documented. Studies report that ChatGPT can support mathematical problem solving, the adaptation of learning strategies, and personalized feedback (Ellis & Slade, [2023](#); Kang, [2024](#); Kwon et al., [2023](#)), and AI-based platforms have been used to adapt learning paths in mathematics (G. M. Lee et al., [2023](#)). The same body of research, however, documents persistent weaknesses in the mathematical coherence, epistemic validity, and theoretical adequacy of chatbot responses (Parra, Sureda, Corica, et al., [2024](#); Sureda et al., [2025](#)). In tasks that demand deep conceptual understanding, such as solving geometric problems, generative models frequently fail to justify results or intermediate steps (Alsharidah & Alkramiti, [2024](#); Corica et al., [2024](#), [2025](#)). GenAI therefore presents university mathematics teaching with both an opportunity and a risk, and which of the two prevails depends largely on how the tool is understood and used.

These weaknesses are not incidental; they follow from the way the underlying technology works. GenAI chatbots, and conversational artificial intelligence more generally, rely on large language models (LLMs) trained on massive volumes of text to capture the semantic, syntactic, and pragmatic regularities of natural language, which allows them to produce fluent responses adapted to the context of an interaction. Because users interact with them in natural language, they are highly accessible and appear to be ready-made assistants for students and professors alike. Their responses, however, are produced by probabilistically predicting the next token of a sequence on the basis of patterns learned from training data (H. Huang et al., [2024](#)). This does not mean that LLM outputs lack semantic content: the internal representations of these models can encode semantic relationships, and their outputs are often semantically coherent (Piantadosi & Hill, [2022](#)). The generation process is nevertheless not anchored in the verification of statements against an external mathematical or factual reference, so fluent and plausible responses may still be false, unjustified, or invented, a phenomenon known as hallucination (L. Huang et al., [2025](#)). For this reason, LLMs have been characterized as “stochastic parrots” (Bender et al., [2021](#); La Fontaine, [2024](#)), a metaphor that underlines the gap between linguistic fluency and warranted knowledge.

For mathematics education, the consequences of this gap are concrete. Analyses of chatbot responses to mathematical and didactic tasks report errors in mathematical coherence and conceptual validity, responses poorly adapted to the educational level, and unfamiliarity with specific didactic frameworks (e.g., Corica et al., [2024](#); Gao et al., [2023](#); Parra, Sureda, & Corica, [2024](#); Sureda et al., [2024](#)). In a discipline in which formalization,

validation, and conceptual understanding are central, the probabilistic nature of LLMs offers no guarantee that generated texts meet the epistemic demands of the discipline or the didactic criteria required of a formative resource. It is therefore necessary to examine not only the products that LLMs generate but also the ways in which teachers interpret, use, or discard them, taking into account the cognitive, epistemic, and ethical dimensions of GenAI as well as its instrumental ones (Selwyn et al., [2022](#); Sigman & Bilinkis, [2023](#)). Professors who are unaware of the scope of GenAI, or who interpret it narrowly, are unlikely to be able to support its pedagogical use.

A growing body of research has accordingly turned to teachers' beliefs, attitudes, and practices regarding GenAI and to the tensions and appropriations that emerge in its integration (Alsharidah & Alkramiti, [2024](#); Song et al., [2025](#)). In Mexico and Latin America, studies report generally positive attitudes among professors, who recognize the potential of GenAI to improve productivity and the quality of learning, alongside persistent barriers such as insufficient training, ethical doubts, and the absence of clear institutional regulations (Díaz, [2025](#); Ramírez & Casillas, [2024](#)). Similar concerns appear in other contexts, where plagiarism, technological dependence, and the erosion of critical and creative skills are emphasized (Castillo et al., [2025](#); Morán-Ortega et al., [2024](#)). Across this literature, the recommended stance is a critical and balanced one: GenAI should serve not as a substitute for cognitive effort but as a support tool that requires supervision, analysis, and ongoing teacher development in digital and ethical competencies (Socorro, [2024](#)).

Most of these studies, however, rely on questionnaires or perception scales and therefore describe what teachers declare about GenAI (their attitudes, expectations, and perceived barriers) rather than the implicit knowledge that organizes their actual use of the tool. Two questions consequently remain largely unexplored. First, little is known about the propositions that practicing university professors who teach mathematics hold to be true when they decide whether, how, and for what purpose to use GenAI. Second, it is unclear how their long experience with other digital technologies, particularly web search engines, shapes the way they approach a technology that generates rather than retrieves information. Answering these questions requires a framework capable of reaching beneath declared opinions to the knowledge that guides action.

The Theory of Conceptual Fields (Vergnaud, [1990](#), [2009](#)) offers such a framework. It models activity as organized by schemes, whose central components are operational invariants (OI), namely concepts-in-action and theorems-in-action; these invariants can be inferred from what subjects say and do in specific situations (Sureda & Otero, [2011](#); Vergnaud, [2007](#)). A theorem-in-action (TA) is a proposition that the subject holds to be true, often implicitly, when dealing with a class of situations. From this perspective, a mathematics professor's decision to use, or not to use, GenAI in teaching is guided by schemes whose theorems-in-action concern what GenAI is, what it is useful for, and how its outputs should be treated (Vergnaud, [1990](#), [2013](#)). The framework therefore makes it possible to identify not only whether professors value GenAI positively or negatively, but also which implicit propositions guide their use, whether these propositions are appropriate, and whether contradictory propositions coexist, a situation that the theory

regards as characteristic of schemes under construction. Because the data in this study consist of interviews, the theorems-in-action analyzed here are inferred from the professors' verbal accounts of their use of GenAI and not from direct observation of their classroom practice.

This article is part of a broader research project conducted with the same group of professors. A previous publication from the project (Silgado-Tuñón, Sureda, et al., 2025) offered a descriptive overview of the operational invariants identified in the interviews, organized into thematic categories. The present article extends that work in three respects: (a) it evaluates the validity of each theorem-in-action against an explicit reference standard derived from the literature on GenAI (correct, incorrect, or not yet established); (b) it reports, for each theorem-in-action, both its frequency of occurrence and its distribution across the professors in whose discourse it was inferred; and (c) it interprets the coexistence of contradictory theorems-in-action explicitly in terms of scheme construction, that is, as the transfer to GenAI of invariants developed through the use of search engines.

Accordingly, the objective of this study is to explore how practicing university professors who teach mathematics conceive of and report using GenAI in their teaching, on the basis of the theorems-in-action that can be inferred from their discourse within the framework of the Theory of Conceptual Fields. Three research questions guided the study:

RQ1. What theorems-in-action about GenAI, its use, and its usefulness for teaching can be inferred from the discourse of practicing university professors who teach mathematics?

RQ2. Which of these theorems-in-action are appropriate, inappropriate, or not yet established in light of current knowledge about GenAI?

RQ3. What do the co-occurrence and distribution of these theorems-in-action among the professors reveal about the construction of their schemes of use of GenAI, particularly regarding its integration into classroom teaching?

2. Method

2.1 Research Design

The research adopts a qualitative, descriptive-exploratory case-study design (Hernández et al., 1997; Stake, 1995; Yin, 2018). The case is bounded as the group of practicing professors who teach mathematics at one public university in Mexico during the period in which GenAI chatbots became widely available in higher education. Although the participants belong to different academic units, they share the same institutional context (curricular framework, regulations, and resources), and all of them teach mathematics courses; this makes the group a bounded system appropriate for an in-depth study of how teachers' schemes of use of an emerging technology are constituted. The case is instrumental (Stake, 1995): it is studied not for its own sake but to understand the theoretical phenomenon of scheme construction when facing a new technology. The unit of analysis is the theorem-in-action inferred from the professors' discourse, and the unit of segmentation is the episode (speaking turn), as described below. Consequently,

the study does not aim to draw generalizable conclusions applicable to all university professors; its contribution is analytical rather than statistical (Yin, 2018).

2.2 Research Context

Semi-structured interviews were conducted with ten practicing university professors from the academic units of Mathematics, Physics, Light and Matter Science and Technology (LUMAT), and Engineering at a public university in Mexico. All the professors interviewed teach mathematics in different years and degree programs at the university. The profiles of the practicing university professors are as follows: four with degrees in mathematics — professors coded P1, P4, P6, and P8; two with degrees in economics — professors P2 and P7; two engineers, one in agronomy (P3) and one in computer systems (P5); and two with degrees in physics — professors P9 and P10. All hold a doctoral degree, with careers spanning between 2 and 35 years. Participants were selected through purposive sampling among professors who taught mathematics courses at the university and agreed to participate voluntarily. The sample size was defined by the exploratory purpose of the study and by the information power of the sample, which is high when the aim is narrow, the participants have specific experience relevant to that aim, and the interview dialogue is rich (Malterud et al., 2016); statistical representativeness was not sought.

2.3 Data Collection Instrument

A semi-structured interview was used as the main instrument, lasting approximately 50 minutes per in-person session. The interview guide was structured around three main axes:

Axis I: Conceptualization of GenAI. Definition, construction, and exemplification of prompts.

Axis II: Teaching praxis. Personal, academic, and classroom use of the technology.

Axis III: Pedagogical outlook. Perceived impact on the teaching and learning of mathematics.

The interview guide was developed by the research team from the theoretical framework and the literature on teachers' use of GenAI. The interviews were conducted in Spanish and audio-recorded with the participants' consent; the excerpts presented in this article were translated into English by the authors. The main questions of each axis were posed to all participants, while follow-up probes were adaptive, formulated in response to each interviewee's answers to elicit examples of actual use and the reasons given for it. Examples of the main questions of each axis can be seen in the interviewer turns reproduced in the excerpts (e.g., Tables 1, 3, and 4).

2.4 Data Processing

For the processing of the qualitative data, the interviews were transcribed and diarized using GenAI-based tools. Specifically, Whisper was used for automatic speech recognition, ChatGPT for organizing the responses and conducting a preliminary analysis, and cloud services to store and process the information securely and efficiently. To safeguard confidentiality, participants were assigned codes (P1–P10), and names, institutional references, and other identifying information were removed from the

transcripts before any text was entered into ChatGPT. Audio files and transcripts were stored in access-restricted storage, accessible only to the research team. The accuracy of the automatic transcriptions and speaker labels was verified by the researchers, who compared each transcript with its audio recording and corrected recognition and diarization errors. ChatGPT was used only as an auxiliary tool to organize the responses by interview axis and to produce a preliminary summary; it was not used to infer or classify theorems-in-action. All inferences, categorizations, and validity judgments reported in this article were made by the researchers directly from the verified transcripts, and the preliminary ChatGPT outputs were not treated as data.

2.5 Ethical Considerations

Participation was voluntary. Before the interviews, all participants were informed of the purpose of the study, the audio recording, the use of AI-based tools for transcription and organization of the data, the anonymization procedure, and their right to withdraw at any time, and they gave their informed consent.

2.6 Data Analysis

Below is an illustrative example (Table 1) of the procedure used to construct the episodes and infer the theorems-in-action from the analysis of the interviews. The excerpt corresponds to professor P1 and shows how, from certain statements made by the interviewee, different theorems-in-action are inferred. The letter E indicates the corresponding episode, whose numbering changes each time the speaking turn shifts between interviewer and interviewee. The third column presents the TA inferred by the researcher from the professor's discourse. The excerpts from which the inference is drawn appear in italics within the interviewee's column.

In the example (Table 1), the inference of the theorem-in-action is shown. In episode E.11 the interviewer asks about the role the professor assigns to GenAI. In E.12 the interviewee states that GenAI can be used to check results or calculations. From this statement it is possible to infer TA.5: GenAI provides correct results; and that it has the authority of a book (TA.8), since the professor states that it can be used to check. Likewise, by stating that GenAI is useful for searching for information, it can be inferred that this professor regards GenAI as having properties similar to a search engine (TA.2). It is important to distinguish between what the professor explicitly states and what the researcher infers. P1 does not literally say that GenAI provides correct results; the inference rests on the fact that using a tool "to check results" presupposes attributing to it the status of a reliable reference against which one's own results can be contrasted. The following decision rules were applied: (a) a TA was inferred only when the statement, or the action described, would be incoherent without assuming the proposition; and (b) when a statement admitted more than one reasonable interpretation, the TA was retained only if it was corroborated in another episode of the same interview or if both researchers agreed on it after discussion; otherwise, it was discarded.

Table 1

Excerpt Corresponding to Interviewee P1

Interviewer	Interviewee (P1)	TA
E. 11: What role do you think artificial intelligence could play?	<i>E. 12: I think it can be used to check results or calculations (TA.5; TA.8). I also think it's useful as a way to search for information (TA.2). To replace [teachers]? I don't know. I think that if you ask artificial intelligence something, it will probably give you an answer.</i>	TA.5: GenAI provides correct results. TA.2: GenAI works like a search engine. TA.8: GenAI has the authority of a book or a professor.

Once the TA were inferred, they were classified into categories. The inference and validation of each theorem were carried out in three steps. First, the interviewer read the verified transcripts in full, segmented them into episodes, and inferred possible TA, recording for each one the episode and the excerpt supporting it. Second, a fellow researcher reviewed every inferred TA against its supporting excerpt, accepting, modifying, or rejecting it. Third, disagreements were discussed in consensus meetings of the research team, where each disputed TA was resolved by returning to the transcript and applying the decision rules described above; TA on which no agreement was reached were discarded. Following O'Connor and Joffe (2020), consensus coding was preferred to a statistical agreement coefficient, since the aim was interpretive agreement on inferences rather than the reproducibility of a fixed codebook. As a result, a final list of 68 TA was agreed upon among the authors of the study. An audit trail (transcripts, episode codes, inferred TA, and consensus decisions) was kept to ensure the traceability of the inferences (Lincoln & Guba, 1985). Figure 1 summarizes the complete analytical procedure.

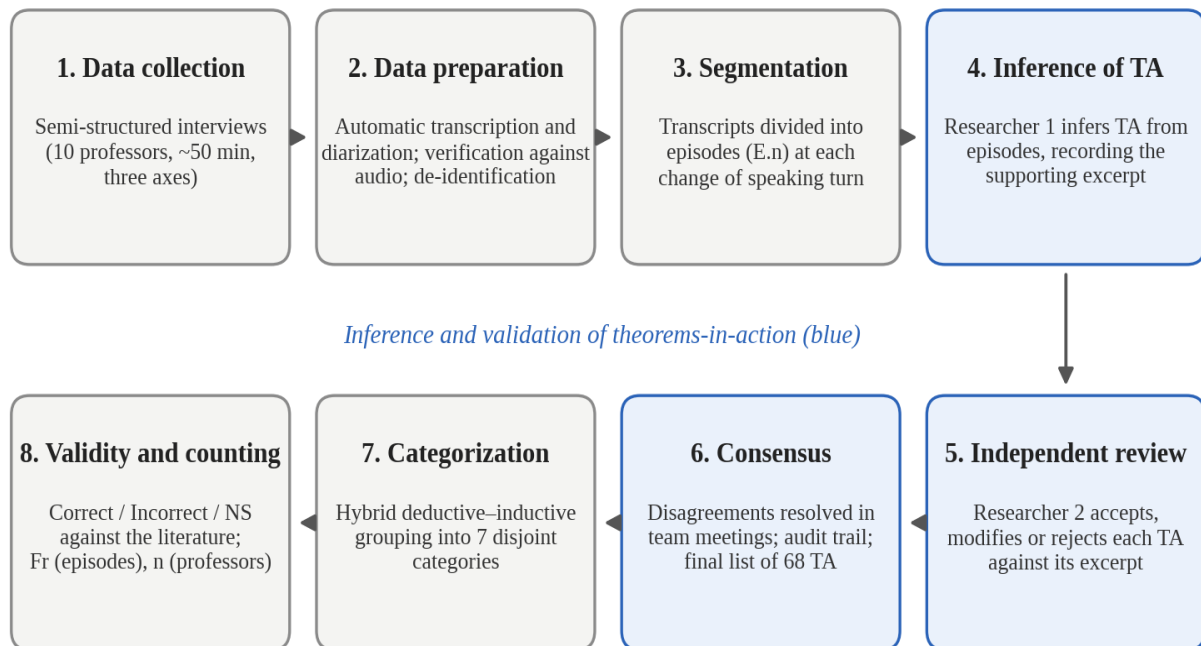
The categories were constructed through a hybrid deductive-inductive process (Fereday & Muir-Cochrane, 2006). Deductively, the three axes of the interview guide and the constructs of the Theory of Conceptual Fields (what the subject considers GenAI to be, the rules of action for using it, and the conditions under which its results are held to be valid) provided initial sensitizing areas. Inductively, the inferred TA were grouped according to the object to which each proposition refers, which led to seven final categories: (1) General Functionality of GenAI; (2) Use and Construction of Prompts; (3) Authority and Validation of Knowledge; (4) Academic and Educational Applications; (5) Ethics, Regulation, and Plagiarism; (6) Relationship with Teaching and Learning; and (7) Teacher Knowledge and Use of GenAI. Because a single statement may touch on several aspects (for example, a belief about the authority of GenAI may have implications for teaching), each TA was assigned according to the main object of the proposition asserted to be true, not according to its possible implications; when a TA seemed to fit two categories, the researchers discussed it until agreeing on its main object, and the decision was recorded in the audit trail. When one excerpt expressed two distinct propositions, two different TA were inferred, each assigned to its own category.

As a result of this procedure, the categories are disjoint and exhaustive by construction; that is, each TA was assigned to exactly one category, and every TA was

assigned to a category. Once this classification was completed, two indicators were computed for each TA: its frequency (Fr), defined as the number of episodes (interviewee speaking turns) in which the TA could be inferred, so that a TA was counted at most once per episode but could be counted in several episodes of the same interview; and the number of professors (n) in whose interviews the TA was inferred at least once. Reporting both indicators prevents repeated statements by a single professor from being mistaken for widespread ideas. These data were entered into a table created for this purpose; the frequencies of the theorems-in-action discussed in this article are shown in Figure 2, and their distribution across professors in Figure 3. In Table 2, the first column shows the inferred TA number, the second column the description of the TA, the third column the number of times each TA is repeated, and the fourth column indicates whether the theorem-in-action is appropriate, not appropriate, or whether the field of study does not know whether it is or not (NS — not specified). The validity judgment was made by the research team using the following reference standard: a TA was classified as Correct when the proposition is consistent with current technical and educational literature on how GenAI works and how it should be used (e.g., L. Huang et al., [2025](#); Lo, [2023](#); Mishra et al., [2023](#)); Incorrect when it contradicts that literature, or when it presents as sufficient a condition that the literature considers insufficient; and NS when research has not yet produced sufficient evidence to establish whether the proposition is true, as occurs with claims about the effect of GenAI on students' learning. Each judgment was first made individually by two researchers and then agreed upon by consensus. The judgment refers to the proposition inferred in context and not to isolated words. In this case, it is important to note that even when a theorem-in-action is functional for the subject, it is not necessarily appropriate within a scientific or socialized knowledge system. The analysis of its validity is carried out within the field of appropriate GenAI use, which makes it possible to determine whether the professor's use of the tool is satisfactory. For example, TA.2: GenAI works like a search engine is not an appropriate TA for the use of GenAI, since, unlike a search engine, which retrieves and ranks existing documents whose sources can be inspected, a chatbot generates a new synthesis probabilistically, which may contain errors or invented data that are statistically plausible; even when some chatbots incorporate web retrieval, the final response remains a generated synthesis (Shah & Bender, [2022](#)). Finally, the last column gives the name of the corresponding category. This example corresponds to the category General Functionality of GenAI, which groups ideas related to what GenAI is and what it can do.

Figure 1

Analytical Procedure for the Inference, Validation, and Quantification of Theorems-in-Action



Note. Steps 4–6 correspond to the inference and validation of the TA.

Table 2

Example of How the TA Were Categorized

TA	Description	Fr	Validity	Category
2	GenAI works like a search engine.	23	Incorrect	General Functionality of GenAI (What it is, and what it can do)
51	GenAI can provide immediate, non-judgmental access.	9		
6	GenAI is a generator of simple, correct code.	16	Incorrect	
7	GenAI is an image generator.	7		
19	GenAI can be used as a programming assistant.	6		
20	GenAI makes it possible to optimize complex processes.	18	Correct	

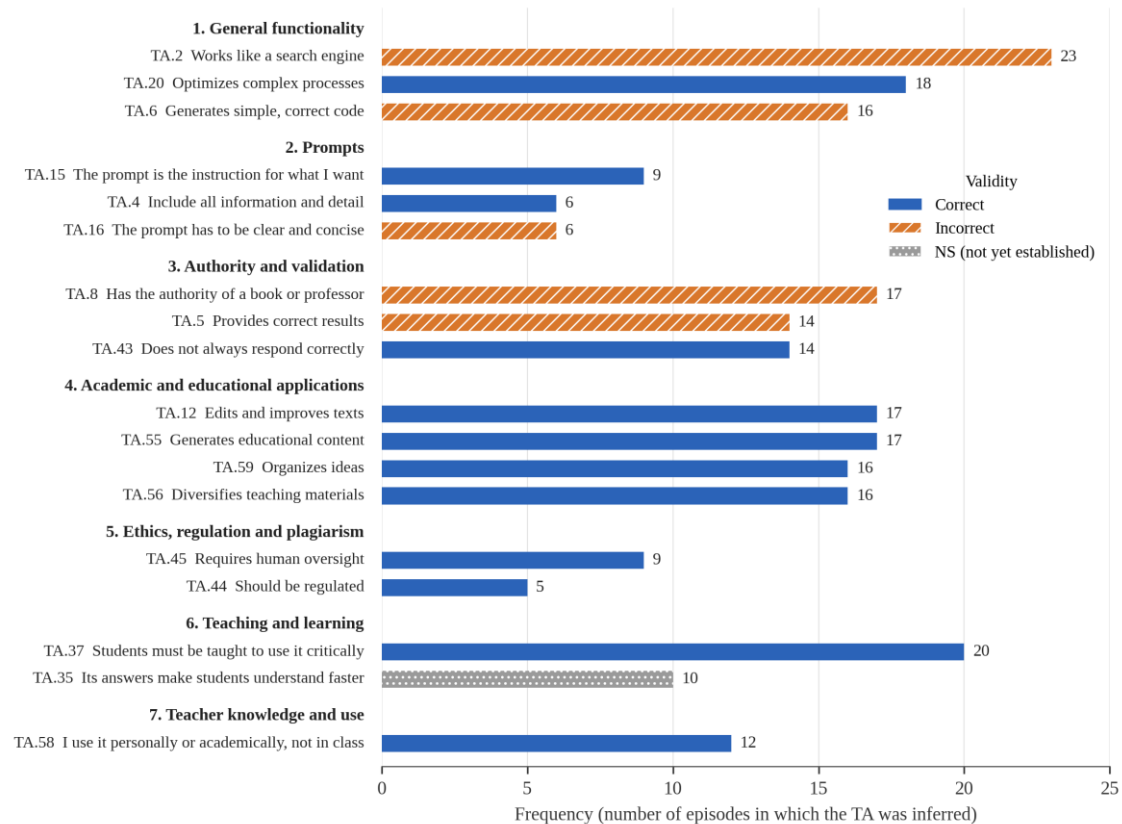
3. Results

3.1 Category 1: General Functionality of GenAI

For Category 1, on the General Functionality of GenAI, which groups the theorems-in-action about what GenAI is and what it can do, the most frequent TA are shown in Figure 2.

Figure 2

Most Frequent Theorems-in-Action by Category, Frequency, and Validity



Note. Fr = number of episodes in which the TA was inferred. Incorrect TA are shown with hatching and NS TA with dots so that validity is not conveyed by color alone.

As Figure 2 shows, TA.2 was the most frequently inferred theorem in the interviews, with a frequency of 23 episodes, inferred in the interviews of seven professors. This frequency is also the highest across all categories. This theorem can be seen, for example, in the interview with professor P1, in episode E.12, already presented in Table 1, where the professor states that GenAI is “useful as a way to search for information”.

TA.2 was identified in the interviews of professors P1, P2, P4, P5, P7, P9, and P10. This indicates that seven of the ten professors have adopted GenAI as a search engine. This matters because the two technologies work differently. A search engine retrieves and ranks existing documents; its results are not inherently correct, since they may be inaccurate, biased, or commercially motivated, but they point to identifiable sources whose authorship, provenance, and quality can be evaluated by the user (Wineburg & McGrew, 2019). A GenAI chatbot, in contrast, produces a new, probabilistically generated synthesis that frequently lacks traceable sources and may contain plausible but false content (L. Huang et al., 2025; Shah & Bender, 2022). Both therefore require verification, but of different kinds: evaluating sources in the case of search engines, and checking the generated content itself in the case of GenAI. Treating GenAI as a search engine may thus lead professors to transfer to it verification habits, such as trusting a well-ranked result, that are insufficient for generated responses, which is particularly problematic when it

comes to the teaching of knowledge. TA.6, also with a high frequency of 16, refers to the veracity of GenAI's responses, in this case for code generation. Finally, TA.20, relating to the optimization of complex processes, has a frequency of 18, indicating its presence across several interviews.

3.2 Category 2: Use and Construction of Prompts

For Category 2, on the use and construction of prompts, which groups theorems-in-action about how to interact with GenAI, the most frequent TA are shown in Figure 2.

In this category, the three most frequent TA are TA.16, TA.4, and TA.15, the first being incorrect with respect to GenAI, and the other two correct. TA.16 rests on an idea of GenAI more as a search engine, which corresponds to TA.2 mentioned above. The classification of TA.16 as incorrect requires clarification. Clarity and concision are recognized as useful principles in prompt design (Lo, 2023); what is considered incorrect is not these qualities in themselves, but the operational invariant inferred from the professors' discourse, according to which a short, clear instruction is sufficient to obtain an adequate response, as occurs when entering keywords into a search engine. The literature on prompting emphasizes that clarity must be accompanied by context, an explicit specification of the task and of the expected output, and iterative refinement of the interaction (Lo, 2023; White et al., 2023). In the excerpt below, P2 describes a one-line instruction without any context about the purpose, audience, or criteria for improvement, and presents it as the usual way of interacting with the tool. This theorem can be inferred, for example, in the interview with professor P2, in the following excerpt (Table 3).

Table 3

Excerpt From the Interview With Professor P2 Regarding TA.16

Interviewer	Interviewee (P2)	TA
E. 28: And when is it for a review, like text editing?	E. 29: Ah, yes, for example, also. In that case, the instruction I give it is: 'Can you read and improve my writing?', and then I give it the text and it helps me with that.	TA.16: The prompt has to be clear and concise.

On the other hand, a TA that can be useful when entering a prompt into GenAI is TA.4, since it is more consistent with the appropriate use of a prompt. However, it is important to note that a better prompt does not guarantee that GenAI will provide a correct response. It may be more complete, more detailed, with a tone or writing style more appropriate to what is required, but not correct. This theorem is found in the interviews with professors P1, P3, P4, P9, and P10 — that is, it could be inferred in only half of the interviews. An example of this inference is shown in the interview with professor P10, in E.14 (Table 4).

Table 4

Excerpt From the Interview With Professor P10 Regarding TA.4

Interviewer	Interviewee (P10)	TA
E. 13: What elements do you consider when interacting with an artificial intelligence to obtain better results?	<i>E. 14: When I consult an artificial intelligence, I give it contextual information: for example, I clarify that I am a researcher or that I have a certain specialty. I think it's important to train it a little, because if it isn't given enough context, the results may not be accurate or useful. In other words, you have to know how to ask it in order to get something of value.</i>	TA.4: In a prompt I include all the information and explain in detail what I want.

Finally, TA.15, relating to the definition of a prompt, could be inferred a total of 9 times (P1, P2, P3, P4, P8, P9, and P10), which is a good indicator of what professors consider a prompt to be.

3.3 Category 3: Authority and Validation of Knowledge

For Category 3, called Authority and Validation of Knowledge, relating to trust, validity, and verification of the responses provided by the various GenAI tools, the most frequent TA are shown in Figure 2.

These three theorems-in-action are fundamental to understanding the differences between a search engine like Google and GenAI. That TA.5 and TA.8, with frequencies of 14 and 17 respectively, are the most frequent shows how the schemes associated with search engines, built up over a long period of use, continue to be activated when using GenAI, even when they are not appropriate. The literature shows that the responses provided by GenAI do not undergo the editorial and expert validation of books (TA.8), nor are they necessarily correct (TA.5) (L. Huang et al., 2025). TA.5 was inferred in episodes such as E.12 of the interview with professor P1 (Table 1), in which the professor states that GenAI “can be used to check results or calculations”.

This theorem-in-action is inferred in the interviews of eight of the ten professors interviewed (P1, P2, P3, P4, P5, P8, P9, and P10), which underscores the problem for the teaching of knowledge. The same pattern is repeated with TA.8, inferred in seven of the interviews (P1, P2, P3, P4, P8, P9, and P10). The presence of these two TA corresponds to the activation of schemes linked to the use of search engines like Google. An example of TA.8 is shown in E.60 of the interview with P8 (Table 5).

Table 5

Excerpt From the Interview With Professor P8 Regarding TA.8

Interviewer	Interviewee (P8)	TA
E. 59: Do you think artificial intelligence can have a positive impact on the contextual understanding of mathematics?	<i>E. 60: Yes, definitely. As I said before, it's like having a professor available 24 hours a day, 365 days a year. The fact that a student can be reflecting on a concept and formulate a question whose answer will come without judgment is very useful. (...)</i>	TA.8: GenAI has the authority of a book or a professor.

An appropriate TA for the good use of GenAI is TA.43, linked to the lack of veracity of these chatbots' responses. Its high frequency (14) indicates that professors realize that GenAI does not always give correct answers. This theorem is inferred in the discourse of professors P2, P4, P5, P7, P9, and P10. However, since TA.5 and TA.8 could also be inferred in the interviews of most of these same professors, it is possible to assume that, during the use of GenAI, different schemes coexist — some stemming from the use of search engines, and others that are being built as they use this new technology. In the Theory of Conceptual Fields, when a subject faces a new class of situations, schemes developed for familiar situations are first mobilized and only progressively adapted; during this process, operational invariants that are inappropriate for the new situations may coexist with others under construction, and the subject may activate one or the other depending on the situation (Vergnaud, 1990, 2009). This kind of coexistence has been documented in the construction of new mathematical knowledge (Sureda & Otero, 2013, 2014). TA.43 can be inferred in episode 46 of the interview with professor P7 (Table 6).

Table 6

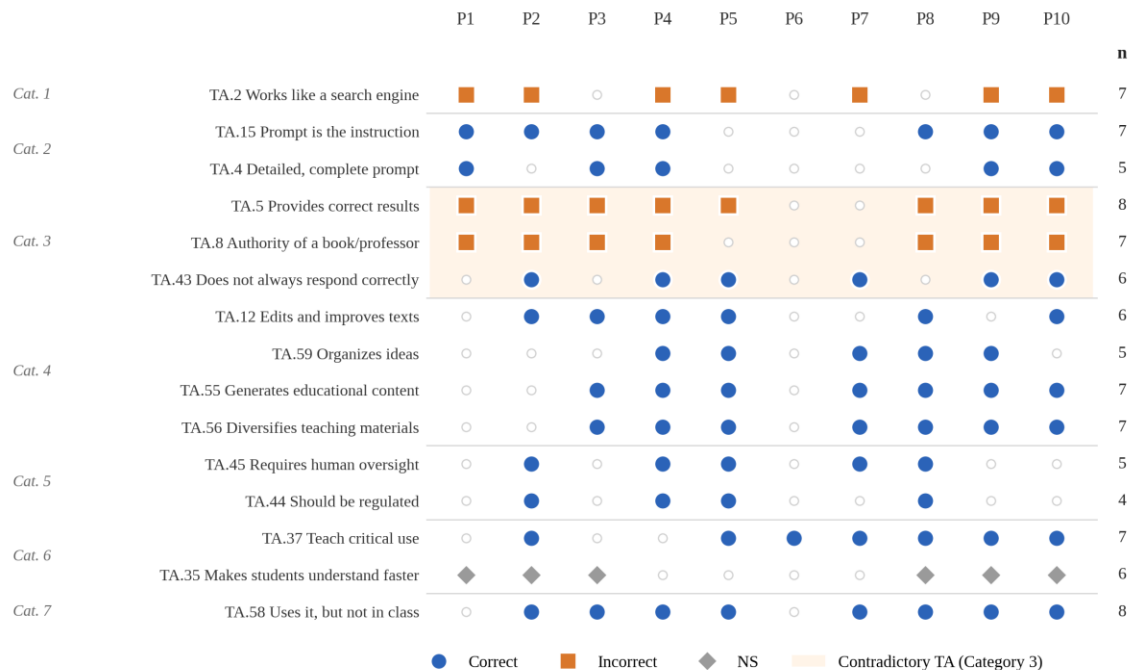
Excerpt From the Interview With Professor P7 Regarding TA.43

Interviewer	Interviewee (P7)	TA
E. 45: To what extent do you think generative artificial intelligence can transform the way students acquire mathematical knowledge?	E. 46: Also, artificial intelligence can provide incorrect information. I worry that students might assume certain content is true without checking its veracity. (...)	TA.43: GenAI does not always respond correctly.

Figure 3 shows the distribution of the most frequent theorems-in-action across the ten professors and makes this coexistence visible at the individual level. TA.5 (GenAI provides correct results) and TA.43 (GenAI does not always respond correctly) were both inferred in the interviews of P2, P4, P5, P9, and P10, and TA.8 and TA.43 in those of P2, P4, P9, and P10. In other words, it is not that some professors hold appropriate invariants and others inappropriate ones: the same professors mobilize contradictory theorems-in-action depending on the situation they are describing, which is consistent with schemes that are still under construction. The figure also shows that TA.58 (use outside the classroom) and TA.37 (the need to teach critical use) are simultaneously present in the discourse of six professors (P2, P5, P7, P8, P9, and P10), a tension discussed below.

Figure 3

Distribution of the Most Frequent Theorems-in-Action Across the Ten Professors



Note. A filled marker indicates that the TA was inferred at least once in that professor's interview; n = number of professors. The shaded band highlights the contradictory TA of Category 3.

3.4 Category 4: Academic and Educational Applications

For Category 4: Academic and Educational Applications, relating to the uses professors make of GenAI in both their personal and professional activity, the most frequent TA are shown in Figure 2.

These theorems-in-action can be grouped into two clusters: TA.12 and TA.59 refer to the use these professors make of GenAI in their professional research tasks, while the other two (TA.55 and TA.56) are more closely linked to the possibilities these tools offer when organizing their classes. As an example, TA.12 appears in E.87 of the interview with P3 (Table 7).

Table 7

Excerpt From the Interview With Professor P3 Regarding TA.12

Interviewer	Interviewee (P3)	TA
E. 86: Ok, yes, in that sense, for example, you say in your research that you use artificial intelligence, well, to investigate matters, but for example, beyond that use, does it help you, for example, in writing articles or that kind of thing?	E. 87: Yes. Translations? No, I use it mainly for paraphrasing, to paraphrase, because unfortunately we are now heavily audited, (...) so I try to use this kind of tool to help me paraphrase my own documents, so it's paradoxical, isn't it?	TA.12: GenAI is a good tool for editing and improving texts.

The interviews of six of the professors (P2, P3, P4, P5, P8, and P10) also show the presence of TA.12, which suggests it is one of the most frequent uses of this tool. TA.59, relating to the use of GenAI as a tool for organizing ideas, is also fairly frequent, 16 times in total, in the interviews with professors P4, P5, P7, P8, and P9. An excerpt from which it was inferred is E.31, from the interview with professor P4 (Table 8).

Table 8

Excerpt From the Interview With Professor P4 Regarding TA.59

Interviewer	Interviewee (P4)	TA
E. 30: And in mathematics, for example, is there anything specific?	E. 31: Yes, for example, when I have to give a talk on a new topic, I like to see how ChatGPT organizes it. I ask it to give me ideas on how to structure the talk, and I like how it proposes something like: 'First an introduction, then these topics, then this...' And then I say: 'Ah, how cool!' (...)	TA.59: GenAI is a good tool for organizing ideas.

Regarding the other two theorems-in-action (TA.55 and TA.56), relating to the use of GenAI as a tool for preparing classes, both for the creation of teaching materials (TA.55) and for diversifying exercises (TA.56), their high frequency in the discourse of seven of the professors (P3, P4, P5, P7, P8, P9, and P10) shows their usefulness for these tasks. Notably, both TA are inferred from the interviews of the same professors. TA.55 is inferred, for example, in E.12 of professor P9 (Table 9).

Table 9

Excerpt From the Interview With Professor P9 Regarding TA.55

Interviewer	Interviewee (P9)	TA
E. 11: Have you used generative artificial intelligence tools? Could you give an example of how you use them?	<i>E. 12: Yes, definitely. I have used several artificial intelligence platforms, such as ChatGPT, Bing Copilot, Claude, Perplexity, Pi, and Gemini, among others. One strategy I use is to pose a mathematical or physics problem to these platforms. I ask them to give me different strategies for solving that problem, and then I compare the answers each one gives. This process is useful for me, since it lets me have different perspectives and prepare my classes better, drawing on varied approaches that I can present to the students.</i>	TA.55: GenAI can be used to generate new educational content, such as tutorials and exercises.

Table 10 shows an example of TA.56 in the interview with professor P8.

Table 10

Excerpt From the Interview With Professor P8 Regarding TA.56

Interviewer	Interviewee (P8)	TA
E. 13: Are you familiar with artificial intelligence and generative artificial intelligence?	<i>E. 14: (...) For example, in the design of teaching materials, it lets me quickly generate clear, well-structured presentations, with good spelling and visual design, since I now have a visually impaired student. In my case, working in mathematics, being able to integrate LaTeX code quickly and efficiently has been a great advantage (...).</i>	TA.56: GenAI can contribute to the diversification of teaching materials.

3.5 Category 5: Ethics, Regulation, and Plagiarism

For Category 5, on Ethics, Regulation, and Plagiarism, referring to responsibility, the appropriate use of GenAI, and the norms that should govern its use, the most frequent TA are TA.44, relating to the desire for regulatory norms to exist, and TA.45, which advocates human oversight as an ethical requirement. Both frequencies are shown in Figure 2.

TA.45, which refers to human oversight as an ethical requirement for the use of GenAI, is found in the discourse of five of the professors (P2, P4, P5, P7, and P8). As an example, this TA appears in E.112 of the interview with P2 (Table 11).

Table 11

Excerpt From the Interview With Professor P2 Regarding TA.45 and TA.44

Interviewer	Interviewee (P2)	TA
E. 111: What do you think about regulations for the use of artificial intelligence?	<i>E. 112: I think regulations are necessary, especially in areas like law, where an AI could be involved in important decisions. Although AI can speed up processes, human judgment must be present to ensure that decisions are appropriate. In that sense, regulations are useful for governing the use of the technology.</i>	TA.45: The use of GenAI requires human oversight as an ethical requirement. TA.44: There should be regulations governing the use of GenAI.

As the use of GenAI becomes widespread in classrooms, professors P2, P4, P5, and P8 show concern about the responsible use of GenAI, emphasizing that it is important for regulations to be put in place governing its use (TA.44). Within this case, only four of the ten professors expressed this concern. Although this finding cannot be generalized, it is consistent with studies in Mexico and Latin America that report the absence of clear institutional regulations and teachers' ethical doubts about GenAI (Díaz, 2025; Ramírez & Casillas, 2024), and suggests that regulatory issues are not yet a central element of these professors' schemes of use. TA.44 was inferred from the same excerpt, E.112 of the interview with P2 (Table 11). The same episode allowed both TA.45 and TA.44 to be inferred because it contains two distinct propositions: that human judgment must oversee decisions, and that regulations are necessary.

3.6 Category 6: Relationship With Teaching and Learning

For Category 6, called Relationship with Teaching and Learning, on the effect of GenAI use on educational spaces linked to the teaching and learning process, the most frequent TA are shown in Figure 2.

In this category, the two most frequent TA are TA.35 and TA.37. TA.37 is one of the most frequent theorems-in-action in the entire study (with a frequency of 20) and rests on the idea that students must use GenAI critically, which corresponds with TA.45, mentioned above, regarding the ethical and responsible use of GenAI. This theorem can be inferred, for example, in the interview with professor P2, in the following excerpt (Table 12).

Table 12

Excerpt From the Interview With Professor P2 Regarding TA.37

Interviewer	Interviewee (P2)	TA
E. 109: How do you identify and address the limitations and potential biases of generative artificial intelligence in the context of teaching mathematics?	<i>E. 110: The limitations are clear, because AI doesn't always respond correctly. That's why I think it's important for students to develop critical thinking. They shouldn't accept everything artificial intelligence says, but rather question it. Although AI will improve over time, students and professors will always need to play an active role, verifying the information.</i>	TA.37: Students must be taught to use GenAI critically.

This TA is also inferred in seven of the interviews (P2, P5, P6, P7, P8, P9, and P10), showing professors' concern about the unsupervised use of GenAI, since many of them state that GenAI does not always respond correctly (TA.43). TA.35, on the idea that students understand faster when they use GenAI, is inferred in the interviews of professors P1, P2, P3, P8, P9, and P10. This reflects a fairly widespread idea in favor of using GenAI in the learning of knowledge that has not yet been validated by research in the field. An example of TA.35 is found in E.58 of the interview with P9 (Table 13).

Table 13

Excerpt From the Interview With Professor P9 Regarding TA.35

Interviewer	Interviewee (P9)	TA
E. 57: To what extent do you think generative artificial intelligence can transform the way students acquire mathematical knowledge?	<i>E. 58: I believe that generative artificial intelligence can significantly facilitate learning, since it allows students to understand certain concepts more clearly and reduce the time needed to assimilate complex ideas. In this sense, it becomes a valuable tool for reinforcing knowledge.</i>	TA.35: GenAI's answers allow them to understand faster.

3.7 Category 7: Teacher Knowledge and Use of GenAI

For Category 7, on Teacher Knowledge and Use of GenAI, which groups theorems-in-action about professors' attitudes, experiences, and difficulties, the most frequent TA was TA.58, as shown in Figure 2.

TA.58 is inferred in eight of the interviews (P2, P3, P4, P5, P7, P8, P9, and P10), showing that most of the professors interviewed use GenAI personally or professionally, but not in the classroom. This theorem can be inferred, for example, in the interview with professor P4, in the following excerpt (Table 14).

Table 14

Excerpt From the Interview With Professor P4 Regarding TA.58

Interviewer	Interviewee (P4)	TA
E. 62: And during class, how do you decide on the use of technology?	<i>E. 63: Within class, I don't think I use it directly. I mean, I do everything more beforehand (...) But in class as such, I haven't used technology.</i>	TA.58: I use GenAI personally or academically, but not in class.

The strong presence of TA.58 in the professors' discourse indicates that, in this case, the use of GenAI is concentrated in lesson preparation and professional tasks rather than in classroom activity. The interviews do not allow the reasons for this to be established with certainty: in the excerpt from P4 (Table 14), the professor describes preparing everything beforehand, but does not explicitly mention difficulties in designing tasks with GenAI. Possible explanations, such as the difficulty of designing classroom tasks that involve GenAI, the lack of institutional guidelines, or concerns about the reliability of its responses (TA.43), are consistent with previous studies reporting a lack of training and

of institutional regulation (Díaz, 2025; Ramírez & Casillas, 2024), but should be examined in future research. In any case, this pattern creates a tension with TA.37: it is difficult to teach students to use GenAI critically if the tool is not incorporated as an object of work in classroom activity.

4. Discussion

Regarding RQ1, the professors' discourse is organized mainly around pragmatic invariants: GenAI is conceived as a tool that searches for information, optimizes processes, generates code, edits texts, and produces teaching materials (TA.2, TA.20, TA.6, TA.12, TA.55, and TA.56). This predominantly instrumental conception agrees with studies in Mexico and Latin America that report a positive attitude of teachers towards GenAI based on its potential to increase productivity (Díaz, 2025; Ramírez & Casillas, 2024; Socorro, 2024), and with international studies describing teachers' interest in using it to prepare their lessons (Alsharidah & Alkramiti, 2024; Song et al., 2025). The contribution of the present analysis lies not in confirming these attitudes but in making visible what teachers take to be true about the tool when they use it, which is where the most relevant problems appear.

With respect to RQ2, the most frequent inappropriate theorems-in-action (TA.2, TA.5, TA.8, and TA.6) share a common origin: they attribute to GenAI properties of the digital tools that professors have used for decades, especially web search engines. This is to be expected in the Theory of Conceptual Fields: when a subject faces a new class of situations, the first schemes mobilized are those that have proved effective in neighboring situations, and their operational invariants are revised only progressively, as the subject encounters their limits (Vergnaud, 1990, 2009). In the case of GenAI, however, the transferred invariants are particularly unsuitable. Search engines point to identifiable sources whose credibility can be evaluated (Wineburg & McGrew, 2019), whereas chatbots produce a fluent synthesis without traceable sources and with a propensity to hallucinate (L. Huang et al., 2025; Shah & Bender, 2022). The risk is greater in mathematics. The use described by P1 in Table 1, employing GenAI "to check results or calculations", relies precisely on a capability that research has shown to be unreliable: studies that analyze chatbot responses to mathematical tasks report calculation errors, invalid justifications, and inadequate treatment of geometric and conceptual problems (Corica et al., 2024, 2025; Frieder et al., 2023; Kwon et al., 2023; Parra, Sureda, Corica, et al., 2024; Sureda et al., 2025). In other words, the use of GenAI as a verifier of mathematical results rests on a theorem-in-action (TA.5) that the evidence contradicts, and the authority attributed to the tool (TA.8) may lead professors, and eventually their students, to lower their level of scrutiny precisely where it is most needed.

Regarding RQ3, one of the most informative findings is that appropriate and inappropriate theorems-in-action do not distinguish between groups of professors but coexist within the same professors. As Figure 3 shows, P2, P4, P5, P9, and P10 expressed both that GenAI provides correct results (TA.5) and that it does not always respond correctly (TA.43). This coexistence should not be read as mere inconsistency. In the

Theory of Conceptual Fields, the conceptualization of a new domain is a long process during which invariants of different scope and validity are available to the subject and are activated according to the characteristics of each situation (Sureda & Otero, 2011; Vergnaud, 2009). Similar patterns have been documented in the construction of mathematical concepts, where schemes combine appropriate and inappropriate invariants before they stabilize (Sureda & Otero, 2013, 2014). Its practical consequence is that declarative knowledge that GenAI makes mistakes (TA.43) does not guarantee that verification is part of the scheme of use. Professors who acknowledge the fallibility of chatbots when asked about it in general terms may still act, in specific situations such as checking a calculation or looking up information, as if the answers were reliable. Training that merely warns about hallucinations is therefore insufficient; professors need situations that confront them with the limits of the invariants they actually mobilize.

The theorems-in-action about prompts point in the same direction. Only half of the professors expressed the idea that a prompt should include context and detailed information (TA.4), while TA.16 reflects a search-query logic in which a brief instruction is considered sufficient. The literature on prompt design conceives the prompt as an iterative, contextualized, and explicit interaction rather than as a keyword query (Lo, 2023; White et al., 2023). Nevertheless, as noted in the results, a well-designed prompt improves the relevance and form of a response but does not guarantee its mathematical correctness. Prompting competence and verification competence must therefore be developed together, and both can be regarded as part of the technological pedagogical content knowledge that GenAI now demands of mathematics teachers (Mishra et al., 2023).

The results also reveal a tension between personal and pedagogical use. Most professors use GenAI for their own academic and lesson-preparation tasks (TA.58, TA.12, TA.55, and TA.56) and state that students should be taught to use it critically (TA.37); in six of them both TA.58 and TA.37 coexist (Figure 3), which means that the critical use they consider necessary is not being addressed in the classroom, where it could be taught. This gap between favorable beliefs and classroom integration is well documented in research on educational technology, where teachers' pedagogical beliefs and knowledge, and not only access to technology, have been identified as decisive barriers to integration (Ertmer, 2005). At the same time, the theorem-in-action that GenAI helps students understand faster (TA.35), classified as NS, calls for caution. Recent experimental evidence in secondary mathematics shows that unrestricted access to a GPT-4-based tutor improved students' performance while it was available but reduced their learning once it was withdrawn, whereas a version with pedagogical safeguards largely mitigated this effect (Bastani et al., 2025). Faster access to explanations should not, therefore, be confused with deeper learning.

Finally, concern for regulation (TA.44) and for human oversight (TA.45) was present in only half of the professors or fewer. International guidance on GenAI in education insists that its use in teaching requires institutional policies, the preservation of human agency, and the development of teachers' competencies (Miao & Holmes, 2023), and

regional studies identify the absence of clear regulations as one of the main barriers perceived by teachers (Díaz, [2025](#); Ramírez & Casillas, [2024](#)). This suggests that ethical and regulatory questions are not yet part of the schemes that organize these professors' everyday use of the tool. Taken together, and extending the descriptive overview previously reported for this project (Silgado-Tuñón, Sureda, et al., [2025](#)), the findings indicate that the appropriation of GenAI by these professors is at an early stage of conceptualization, characterized by the transfer of invariants from earlier technologies, the coexistence of contradictory propositions, and a use that is still more personal than didactic.

5. Conclusion

The analysis of interviews with ten university professors who teach mathematics shows a limited appropriation of GenAI tools, as reported in their discourse. There is a presence of theorems-in-action more typical of experience with search engines (such as TA.5: GenAI provides correct results) alongside others more appropriate for the use of these technologies (TA.43: GenAI does not always respond correctly). The coexistence of these, and other mutually contradictory theorems-in-action, when using GenAI is typical of situations in which people build schemes that allow them to deal with new situations (Vergnaud, [1990](#)). Furthermore, the fact that eight of the ten professors reported using GenAI personally or professionally but not in class (TA.58), while the interviews provide no evidence of systematic classroom use by the other two, suggests a still partial and incomplete construction of schemes of use. That these results are consistent with those of other studies (Alsharidah & Alkramiti, [2024](#); Ramírez & Casillas, [2024](#); Song et al., [2025](#)) underscores the fact that the appropriation of GenAI is a complex conceptualization that requires systematic training to enable teachers to move from personal use to didactic use of GenAI.

From a theoretical perspective, this study extends the Theory of Conceptual Fields, traditionally applied to the conceptualization of mathematical and scientific content (Sureda & Otero, [2013](#); Vergnaud, [1990](#)), to the analysis of how teachers conceptualize a technological tool. The results show that the notion of theorem-in-action captures implicit propositions that perception questionnaires do not reveal, such as the transfer of invariants from search engines to GenAI, and that the coexistence of contradictory theorems-in-action can serve as an indicator of schemes under construction and, therefore, of the stage of appropriation of a new technology.

In practical terms, the results suggest that professional development programs for university mathematics teachers should not be limited to instrumental training in GenAI tools, but should explicitly address the theorems-in-action that guide their use: in particular, the difference between retrieving and generating information, the need to verify generated content, the role of context in prompt design, and the design of classroom tasks in which students use GenAI critically. Frameworks such as TPACK (Mishra & Koehler, [2006](#)), revisited for the age of GenAI (Mishra et al., [2023](#)), can support the design of such programs by articulating technological, pedagogical, and mathematical

knowledge, including the potential of GenAI for inclusive teaching (Silgado-Tuñón, López-Flores, et al., 2025), as illustrated by P8's use of GenAI to prepare accessible materials for a visually impaired student (Table 10).

This study has limitations. First, it is a single-institution case with a small purposive sample of ten professors, so its findings are analytical and cannot be statistically generalized to other university contexts. Second, the theorems-in-action were inferred from self-reported discourse in interviews and not from the observation of classroom practice; what professors say about their use of GenAI may therefore differ from what they actually do. Third, frequencies depend on the course of each interview and on the follow-up probes posed, so they should be interpreted as indicators of salience in discourse rather than as measures of prevalence. Finally, given the rapid evolution of GenAI tools, some theorems-in-action may change as professors gain experience with them. The sample size was not determined by a formal saturation criterion, and the distribution by professor could only be reported for the theorems-in-action discussed in detail.

Future research should (a) combine interviews with classroom observation and with the analysis of the prompts and tasks that professors actually design, in order to infer theorems-in-action from action and not only from discourse; (b) follow the same professors longitudinally to examine how their schemes evolve; (c) extend the study to other institutions and countries to compare the invariants identified; and (d) design and evaluate professional development interventions aimed at transforming inappropriate theorems-in-action such as TA.2, TA.5, and TA.8.

Statements and Declarations

Funding:

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing Interests:

The authors declare that they have no financial or non-financial competing interests relevant to this study.

Ethics Approval:

Participation was voluntary, and informed consent was obtained before the interviews. Participants consented to audio recording and to the use of AI-based tools for processing their anonymized data. Identifying information was removed before transcripts were entered into ChatGPT, and audio files and transcripts were accessible only to the research team.

Informed Consent:

All participants gave informed consent, including consent to the audio recording and to the use of AI-based tools for processing their anonymized data.

Data Availability:

The anonymized transcripts supporting the findings of this study are available from the corresponding author upon reasonable request.

Author Contributions:

The research team developed the interview guide from the theoretical framework and relevant literature. The interviewer conducted the initial transcript analysis and inferred candidate theorems-in-action. A second researcher reviewed the inferences against the supporting excerpts. The team resolved disagreements through discussion, developed the final categorization, and agreed on the validity judgments.

Declaration of Generative AI Use:

Whisper was used for automatic transcription and ChatGPT (OpenAI) for organizing the interview responses, as described in the Method section. All outputs were reviewed and verified by the authors, who made all analytical inferences and take full responsibility for the content of the article.

Acknowledgements:

The authors thank the ten university professors who voluntarily participated in the interviews and shared their experiences and reflections on using generative artificial intelligence in mathematics teaching.

6. References

- Alsharidah, M., & Alkramiti, A. (2024). Teachers' perceptions towards using artificial intelligence in mathematics education. *Amazonia Investiga*, 13(84), 43–60. <https://doi.org/10.34069/AI/2024.84.12.3>
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, 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), Article e2422633122. <https://doi.org/10.1073/pnas.2422633122>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
- Castillo, D. I., Mendoza, S. M., & Gómez, A. J. (2025). Percepción y uso de la inteligencia artificial de los estudiantes de la licenciatura en matemáticas. *Estudios y Perspectivas Revista Científica y Académica*, 5(1), 3530–3549. <https://doi.org/10.61384/r.c.a.v5i1.1067>
- Corica, A., Parra, V., Figueroa, P., Schiaffino, S., & Godoy, D. (2025). Análisis y comparación de respuestas aportadas por IA generativa y por profesores de matemática a un problema de geometría euclidiana. *Educación Matemática*, 37(1), 127–160. <https://doi.org/10.24844/EM3701.05>
- Corica, A., Parra, V., Sureda, P., Schiaffino, S., & Godoy, D. (2024). Fractal de Koch: Análisis de respuestas de IA generativa y un profesor de matemática. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*, (37), 89–99. <https://doi.org/10.24215/18509959.37.e8>

- Díaz, J. P. (2025). Más allá de los algoritmos: Desafíos y percepciones docentes sobre la inteligencia artificial generativa en la enseñanza virtual. *Revista de Investigación en Tecnologías de la Información*, 13(29), 141–153. <https://doi.org/10.36825/RITI.13.29.012>
- Ellis, A. R., & Slade, E. (2023). A new era of learning: Considerations for ChatGPT as a tool to enhance statistics and data science education. *Journal of Statistics and Data Science Education*, 31(2), 128–133. <https://doi.org/10.1080/26939169.2023.2223609>
- Ertmer, P. A. (2005). Teacher pedagogical beliefs: The final frontier in our quest for technology integration? *Educational Technology Research and Development*, 53(4), 25–39. <https://doi.org/10.1007/BF02504683>
- Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, 5(1), 80–92. <https://doi.org/10.1177/160940690600500107>
- Frieder, S., Pinchetti, L., Chevalier, A., Griffiths, R.-R., Salvatori, T., Lukasiewicz, T., Petersen, P. C., & Berner, J. (2023). Mathematical capabilities of ChatGPT. *Advances in Neural Information Processing Systems*, 36, 27699–27744
- Gao, J., Pi, R., Zhang, J., Ye, J., Zhong, W., Wang, Y., Hong, L., Han, J., Xu, H., Li, Z., & Kong, L. (2023). G-LLaVA: Solving geometric problem with multi-modal large language model (arXiv:2312.11370). arXiv. <https://doi.org/10.48550/arXiv.2312.11370>
- Hernández, R., Fernández, C., & Baptista, P. (1997). *Metodología de la investigación* (3rd ed.). Panamericana Formas e Impresos.
- Huang, H., Shen, X., Wang, S., Liu, D., & Wang, H. (2024). Verbalized probabilistic graphical modeling with large language models (arXiv:2406.05516). arXiv. <https://doi.org/10.48550/arXiv.2406.05516>
- 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), Article 42. <https://doi.org/10.1145/3703155>
- Kang, Y. J. (2024). A study on the didactical application of ChatGPT for mathematical word problem solving. *Communications of Mathematical Education*, 38(1), 49–67. <https://doi.org/10.7468/jksmee.2024.38.1.49>
- Kwon, O. N., Oh, S. J., Yoon, J. E., Lee, K. Y., Shin, B. C., & Jeong, W. (2023). Analyzing mathematical performances of ChatGPT: Focusing on the solution of national assessment of educational achievement and the college scholastic ability test. *Communications of Mathematical Education*, 37(2), 233–256. <https://doi.org/10.7468/jksmee.2023.37.2.233>
- La Fontaine, G. (2024). Sobre los estocásticos: Una mirada a los modelos grandes de lenguaje. *Lógoi. Revista de Filosofía*, (45), 75–87. <https://doi.org/10.62876/lr.vi45.6480>

- Lee, G. M., Lee, Y. J., & Kim, H. J. (2023). Analysis of adaptive learning in Korea's AI mathematics learning platforms. *Journal of the Korean School Mathematics Society*, 26(3), 245–268. <https://doi.org/10.30807/ksms.2023.26.3.004>
- Lee, Y. J. (2023). An analysis of pre-service teachers' mathematics lesson design using ChatGPT. *Communications of Mathematical Education*, 37(3), 497–516. <https://doi.org/10.7468/jksmee.2023.37.3.497>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.
- Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship*, 49(4), Article 102720. <https://doi.org/10.1016/j.acalib.2023.102720>
- Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, 26(13), 1753–1760. <https://doi.org/10.1177/1049732315617444>
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and generative AI. *Journal of Digital Learning in Teacher Education*, 39(4), 235–251. <https://doi.org/10.1080/21532974.2023.2247480>
- Morán-Ortega, S.-A., Ruiz-Tirado, S.-G., Simental-López, L.-M., & Tirado-López, A.-B. (2024). Barreras de la inteligencia artificial generativa en estudiantes de educación superior: Percepción docente. *Revista de Investigación en Tecnologías de la Información*, 12(25), 26–37. <https://doi.org/10.36825/RITI.12.25.003>
- O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, 19. <https://doi.org/10.1177/1609406919899220>
- Parra, V., Sureda, P., & Corica, A. (2024). Teoría de situaciones didácticas e inteligencia artificial: Diseño de propuestas para enseñar las nociones de muestra y población en educación secundaria. *Uno: Revista de Didáctica de las Matemáticas*, (104), 43–50.
- Parra, V., Sureda, P., Corica, A., Schiaffino, S., & Godoy, D. (2024). Can generative AI solve geometry problems? Strengths and weaknesses of LLMs for geometric reasoning in Spanish. *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(5), 65–74.
- Piantadosi, S. T., & Hill, F. (2022). *Meaning without reference in large language models* (arXiv:2208.02957). arXiv. <https://doi.org/10.48550/arXiv.2208.02957>
- Ramírez, A., & Casillas, M. A. (2024). Percepciones docentes sobre la inteligencia artificial generativa: El caso mexicano. *Revista Paraguaya de Educación a Distancia (REPED)*, 5(2), 44–55. <https://doi.org/10.56152/reped2024-dossierIA1-art4>

- Selwyn, N., Rivera-Vargas, P., Passeron, E., & Miño-Puigcercos, R. (2022). ¿Por qué no todo es (ni debe ser) digital? Interrogantes para pensar sobre digitalización, datificación e inteligencia artificial en educación. In P. Rivera-Vargas, R. Miño-Puigcercos, & E. Passeron (Eds.), *Educación con sentido transformador en la universidad* (pp. 137–147). Octaedro.
- Shah, C., & Bender, E. M. (2022). Situating search. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval (CHIIR '22)* (pp. 221–232). Association for Computing Machinery.
<https://doi.org/10.1145/3498366.3505816>
- Sigman, M., & Bilinkis, S. (2023). *Artificial: La nueva inteligencia y el contorno de lo humano*. Debate.
- Silgado-Tuñón, D. A., & López-Flores, J. I. (2025a). Inteligencia artificial generativa en la educación superior: Una revisión sistemática. *Unión – Revista Iberoamericana de Educación Matemática*, 21(73).
<https://union.fespm.es/index.php/UNION/article/view/1709>
- Silgado-Tuñón, D. A., & López-Flores, J. I. (2025b). Inteligencia artificial generativa en el aula: ¿Aliada o amenaza para la enseñanza de las matemáticas? *Revista Electrónica Tecnologías Emergentes en la Educación*, 2(1), 53–66.
<https://doi.org/10.71713/retee.v2i1.3512>
- Silgado-Tuñón, D. A., López-Flores, J. I., & García, C. C. (2025). IA, discapacidad y docencia inclusiva: Una reflexión desde el DUA y el TPACK. *Revista Electrónica Tecnologías Emergentes en la Educación*, 2(2), 72–90.
<https://doi.org/10.71713/retee.v2i2.3728>
- Silgado-Tuñón, D. A., Sureda, P., López-Flores, J. I., & Magallanes, E. (2025). Exploring university mathematics professors' perceptions and use of GenAI: A conceptual fields approach. *Amazonia Investiga*, 14(86), 250–263.
<https://doi.org/10.34069/AI/2025.86.02.19>
- Socorro, J. (2024). Actitudes del profesorado ante el uso y manejo de la inteligencia artificial generativa (IAG) de modo eficiente. *Revista Científica de Salud y Desarrollo Humano*, 5(3), 1183–1213. <https://doi.org/10.61368/r.s.d.h.v5i3.325>
- Song, X., Mak, J., & Chen, H. (2025). Teachers' and learners' perceptions about implementation of AI tools in elementary mathematics classes. *SAGE Open*, 15(2).
<https://doi.org/10.1177/21582440251334545>
- Stake, R. E. (1995). *The art of case study research*. SAGE.
- Sureda, P., & Otero, M. R. (2011). Nociones fundamentales de la teoría de los campos conceptuales. *Revista Electrónica de Investigación en Educación en Ciencias*, 6(1), 124–138.
- Sureda, P., & Otero, M. R. (2013). Un estudio sobre el proceso de conceptualización de la función exponencial. *Educación Matemática*, 25(2), 89–118. <http://www.revista-educacion-matematica.com/descargas/Vol25-2.pdf>

- Sureda, P., & Otero, M. R. (2014). La conceptualización de la función exponencial y los sistemas de representación. In *La teoría de los campos conceptuales y la conceptualización en el aula de matemática y física* (pp. 33–45). Dunken.
- Sureda, P., Corica, A., Parra, V., Godoy, D., & Schiaffino, S. (2024). La evaluación en educación matemática: Aportes de chatbots y futuros profesores de matemática. *EduTec. Revista Electrónica de Tecnología Educativa*, (89), 64–83. <https://doi.org/10.21556/edutec.2024.89.3243>
- Sureda, P., Parra, V., Corica, A. R., Godoy, D., & Schiaffino, S. (2025). On the role of generative AI in fractals teaching: Solutions and class proposals designed by chatbots and mathematics teachers. *International Journal of Education in Mathematics, Science and Technology*, 13(5), 1298–1316. <https://doi.org/10.46328/ijemst.5713>
- Vergnaud, G. (1990). La théorie des champs conceptuels. *Recherches en Didactique des Mathématiques*, 10(2–3), 133–170.
- Vergnaud, G. (2007). ¿En qué sentido la teoría de los campos conceptuales puede ayudarnos para facilitar aprendizaje significativo? *Investigações em Ensino de Ciências*, 12(2), 285–302.
- Vergnaud, G. (2009). The theory of conceptual fields. *Human Development*, 52(2), 83–94. <https://doi.org/10.1159/000202727>
- Vergnaud, G. (2013). Pourquoi la théorie des champs conceptuels? *Journal for the Study of Education and Development*, 36(2), 131–161.
- Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *EURASIA Journal of Mathematics, Science and Technology Education*, 19(7). <https://doi.org/10.29333/ejmste/13272>
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT (arXiv:2302.11382). arXiv. <https://doi.org/10.48550/arXiv.2302.11382>
- Wineburg, S., & McGrew, S. (2019). Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information. *Teachers College Record*, 121(11), 1–40. <https://doi.org/10.1177/016146811912101102>
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE.