Neutrosophic Sets and Systems, Vol. 101, 2026
University of New Mexico

# Generative Artificial Intelligence as a Formative Mediator in Mathematical Task Design: An Ontosemiotic–Neutrosophic Analysis of Didactical Suitability in Teacher Education

**Arián Vázquez Álvarez 1,***

1 Universidad Bolivariana del Ecuador, Guayaquil, Ecuador. E-mail: avazqueza@ube.edu.ec. ORCID: https://orcid.org/0009-0001-8605-491X

* Correspondence: avazqueza@ube.edu.ec

**Abstract:** The transformation of mathematics teaching in the digital society demands the integration of emerging technologies with solid pedagogical foundations. This study analyses the mediation of generative artificial intelligence (GAI) in the design of mathematical tasks from the Ontosemiotic Approach (OSA), which supplies the criteria and objects of didactical evaluation across six facets of suitability. GAI-mediated tasks were found to carry simultaneously documented contributions, documented risks, and outcomes that remain unresolved until a teacher validates them; a neutrosophic reading is introduced to preserve, rather than collapse, this evidential structure, annotating each facet with independent degrees of truth (T), indeterminacy (I), and falsity (F) and modelling the teacher's critical mediation as an operator of evidence refinement that can reduce indeterminacy without automatically increasing truth. In a qualitative, descriptive, longitudinal study with in-service teachers enrolled in a master's programme in Basic Education, tasks designed without technological mediation were compared with versions reformulated in dialogue with GAI, following a Case Collection and Analysis Protocol. The results show documented contributions in conceptual coherence, argumentation demand, diversity of semiotic representations, and contextualization, coexisting with documented risks of conceptual imprecision, cognitive overload, and homogenization, and with facet-specific questions that remain unresolved pending the teacher's professional judgement. The neutrosophic annotation reported here is qualitative; no numerical T-I-F values were elicited. GAI does not replace the teacher but strengthens their role as epistemological mediator and fosters critical didactic thinking.

**Keywords:** Generative artificial intelligence; mathematical task design; Ontosemiotic Approach; didactical suitability; teacher education; neutrosophy; indeterminacy; evidence refinement.

## 1. Introduction

Mathematics teaching in the contemporary context is immersed in a scenario of profound transformation, driven by technological advances and by new forms of production, circulation, and appropriation of knowledge. The irruption of emerging technologies, particularly generative artificial intelligence (GAI), not only introduces new tools into the classroom but structurally questions traditional ways of teaching and learning [1, 2]. The current challenge does not lie merely in incorporating these technologies into educational processes, but in understanding how they can be articulated from solid pedagogical and epistemological foundations that guarantee the quality of learning.

Historically, mathematics teaching has been crossed by tensions between conceptual understanding and procedural application: in many educational contexts, practice centred on the repetition of algorithms and the mechanical solving of exercises predominates. Although operationally functional, this approach limits the development of more complex cognitive competencies such as argumentation, modelling, problem interpretation, and the construction of mathematical meaning. Tasks, understood as fundamental units of didactic practice, are not simple statements: they are complex configurations articulating learning objectives, contents, representations, procedures, and forms of validation, and their design demands pedagogical and epistemological decisions that go far beyond the writing of a prompt for students.

In this context, GAI introduces a new dimension into didactic design. Tools such as ChatGPT can generate tasks, reformulate statements, propose examples, and analyse mathematical structures almost instantaneously, opening a field of possibilities for teachers who can now interact with systems offering multiple alternatives for enriching their planning [3]. Yet this very potential carries risks, particularly when the technology is used acritically or instrumentally [4, 5]. One of the principal risks is the illusion of automatic improvement: more elaborate or contextualized statements may give the impression that didactic quality has increased when the mathematical content remains superficial or incoherent.

The Ontosemiotic Approach to mathematical knowledge and instruction (OSA) offers a pertinent theoretical perspective for this problem [6]. It understands mathematics as a social practice in which meanings are constructed through the interaction of subjects, objects, and systems of representation, and its notion of didactical suitability — the degree to which a teaching process contributes to the construction of quality mathematical knowledge — provides the evaluative core of this study, organized into six facets that are introduced in Section 2.2. Within it, the epistemic facet answers the question of whether the mathematics being taught is "good mathematics" [7], and receives particular attention here because it is the facet most directly exposed to the risk described above: a generated task can look pedagogically rich while remaining mathematically imprecise.

The evaluative form of that question is not binary. Across the six facets, evidence about GAI-mediated tasks turned out to be simultaneously favourable (documented improvements), unfavourable (documented risks), and undecided (outcomes that depend on the teacher's mediation); Section 2.4 explains why collapsing these three states into a single net suitability judgement would discard exactly the information a teacher needs. The study therefore asks: how does GAI-mediation affect the didactical suitability of mathematical tasks designed by in-service teachers, across its six facets, and how should the non-automatic and evidentially mixed character of that effect be represented formally?

## 2. Theoretical Framework

### *2.1. Task design from the Ontosemiotic Approach*

The OSA proposes that mathematical objects are not isolated entities but constructions emerging from practices in which languages, concepts, properties, procedures, and arguments intervene [6]. Its notion of ontosemiotic configuration describes the internal structure of a task in terms of the objects and processes it mobilizes: an activity on proportionality involves not only operations but the understanding of the concept of proportion, the interpretation of context, the selection of representations, and the justification of results. In educational practice, however, a reduction of this complexity is frequently observed — tasks limited to indicating the procedure to follow, eliminating the need to interpret, decide, or argue. This algorithmic reduction of mathematical knowledge diminishes the diversity of objects at play and limits meaningful understanding [11].

### *2.2. Didactical suitability and its six facets*

The construct of didactical suitability organizes the evaluation of teaching quality into six facets — epistemic, cognitive, interactional, mediational, affective, and ecological — each answering a distinct evaluative question about the same task [12, 13]. The epistemic facet values whether the mathematical meanings proposed are correct, complete, coherent, and representative of disciplinary knowledge, and whether the tasks favour understanding and the articulation of different forms of representation [7, 14]. The cognitive, mediational, interactional, affective, and ecological facets extend that evaluation to the fit between the task and learners' prior knowledge, the pertinence of the resources mobilized, the opportunities for discussion the task opens, its capacity to engage rather than discourage, and its coherence with the curriculum and the institutional context, respectively (these are specified operationally in Table 2). A task can be formally correct and didactically weak if it does not promote comprehension or reflection in any of these dimensions. The competence to analyse and value suitability is itself an object of teacher education [15, 16], and prior research shows that technology alone does not guarantee educational improvement: used instrumentally, it can even reinforce traditional practices [17].

### *2.3. Generative artificial intelligence in mathematics education*

Unlike earlier educational technologies, GAI does not merely present information: it produces content, including mathematical tasks. From an ontosemiotic perspective, GAI can be understood as a system that produces semiotic configurations — combinations of language, concepts, and procedures — which do not by themselves guarantee didactical suitability. Recent studies have begun to evaluate the suitability of tasks reformulated with artificial intelligence [18] and to explore GAI in the education of in-service mathematics teachers [19], converging on a common caution: generated tasks may exhibit elaborate language and apparently meaningful contexts while lacking conceptual depth, posing questions that are answered by mechanically following implicit procedures. The quality of a task does not depend on its appearance but on the structure of knowledge it mobilizes.

### *2.4. Why a neutrosophic reading of didactical suitability*

The Ontosemiotic Approach supplies the criteria and objects of didactical evaluation — the six facets of Section 2.2 and the mathematical languages, concepts, procedures, and arguments they are evaluated against. It does not, by itself, specify how to represent the state of the evidence once that evaluation has been carried out. The findings reported in Section 4 show that, for every facet, the evidence gathered was not uniform: it included documented contributions, documented risks, and questions that could not be settled without the teacher's intervention. A neutrosophic annotation is introduced precisely to preserve this evidential status rather than to add a further layer of abstraction on top of the ontosemiotic analysis: the Ontosemiotic Approach determines what is evaluated, and the neutrosophic representation records the evidential status of the resulting judgement.

Concretely, a GAI-mediated task may display richer semiotic representations — support for a favourable judgement — while at the same time containing a conceptual imprecision — support for an unfavourable judgement — and its adequacy for a particular group of learners may remain unresolved until the teacher evaluates the classroom context — support for an indeterminate judgement. A single net suitability score computed from such a task would not indicate whether the score reflects genuine strength, a masked deficiency compensated by other strengths, or an unresolved judgement still awaiting teacher validation; those are three analytically different situations that call for three different professional responses, and a scalar collapses them into one. This coexistence, documented empirically in Section 4 and not a generic claim that educational processes are uncertain, is the reason a three-valued representation is introduced here.

Each facet f of didactical suitability is therefore annotated with the neutrosophic triplet of Equation (1), in which truth, indeterminacy, and falsity vary independently [8, 9]: no unit-sum constraint $T_f + I_f + F_f = 1$ is imposed, and $T_f$ and $F_f$ may be simultaneously high because the same GAI-mediated task can display a documented improvement and a documented deficiency at once.

$$A_f = \langle T_f, I_f, F_f \rangle,\ \ T_f, I_f, F_f \in [0,1] \qquad (1)$$

Most of the indeterminacy documented in this study is epistemic and context-dependent in nature: it reflects that the adequacy of a specific mediated task cannot yet be established until a teacher verifies its mathematical correctness, its fit to a particular group's prior knowledge, and its curricular placement — not an irreducible indeterminacy that no future evidence could resolve. This is the sense of I used throughout the paper; where neutrosophic paraconsistent logic distinguishes evidence that further inquiry can refine from indeterminacy of an ontological kind [10], the present study concerns the former.

Because the central finding is that this indeterminacy is not permanent but is routinely resolved by teachers in practice, the study also needs a way to represent that resolution formally. Following the evidence-refinement vocabulary of neutrosophic paraconsistent logic [10], the teacher's critical mediation is modelled as an operator M on annotations, Equation (2), whose defining effect is to reduce the indeterminacy component of a facet's annotation:

$$M(A_f) = A'_f = \langle T'_f, I'_f, F'_f \rangle,\ \ I'_f \le I_f \qquad (2)$$

Section 2.5 relates this operator to the concrete professional actions that instantiate it; Section 3.4 specifies how the same idea is applied at the level of an individual task case; Section 4.3 reports how the annotation behaves on the empirical material of this study.

### *2.5. Teacher mediation as evidence refinement*

Operator M is not a generic label for “the teacher intervenes”: it stands for a set of observable professional operations that recur across the findings reported in Section 4 — mathematical validation of generated content, epistemological vigilance toward claims that sound plausible but are imprecise, adjustment of task complexity and length, selection or correction of representations, contextual adaptation to a particular group, verification of curricular coherence, orchestration of argumentation and classroom interaction, and outright rejection or correction of generated content when it is inadequate. Each of these operations is a way of moving a previously undecided piece of evidence into either a validated contribution or an identified and controlled limitation.

Two clarifications matter for how Equation (2) should be read. First, mediation does not automatically raise $T_f$: its defining property is that $I_f$ decreases (or, in the limit, stays the same), not that the outcome is always favourable. Second, no monotonicity is imposed on $T_f$ or $F_f$ individually

— the definition does not require $T_f' \geq T_f$ nor $F_f' \leq F_f$ — because teacher validation may reveal a previously unnoticed error as readily as it may confirm a strength; a mediation that converts unresolved evidence into a documented risk is as much an instance of M as one that converts it into a validated contribution. This is a theoretically deliberate choice: it keeps the model from building in the assumption that professional scrutiny is always benign to the tool's output. The property $I_f' \leq I_f$ is accordingly presented as a hypothesised refinement property of critical mediation, consistent with what the qualitative material of this study shows, rather than as a statistically demonstrated result; its systematic verification across cohorts is future work (Section 4.5).

## 3. Materials and Methods

### *3.1. Design and participants*

The study follows a qualitative, descriptive, longitudinal design oriented to understanding transformations of didactic thinking in in-service teachers. The participants are teachers in continuous formation enrolled in a master's programme in Basic Education, in the academic space of Didactics of Mathematics, at Universidad Bolivariana del Ecuador; the sample was purposive, selecting teachers willing to analyse their practice critically and to experiment with GAI tools. The Ecuadorian curricular context, organized in sublevels requiring coherent learning progressions, frames the study [20]. The exact number of participants, the number of initial task designs and of GAI-mediated versions analysed, and the specific inclusion criteria applied are not reported in the present account of the study; Section 4.6 returns to this as a limitation of the current write-up rather than of the study design itself.

### *3.2. Procedure*

The process was organized in phases: an initial diagnosis of conceptions about technology in mathematics teaching, which evidenced a tendency to treat technology as a complementary resource centred on content presentation; a theoretical grounding phase on the OSA and didactical suitability; a phase in which teachers designed tasks without technological mediation, establishing the comparative baseline (closed exercises, procedural emphasis, scarce argumentation and representational diversity predominated); and the core mediation phase, in which teachers interacted with GAI tools, principally ChatGPT, to reformulate their initial tasks through dialogue — formulating requests, comparing alternatives, identifying errors, and adjusting generated proposals. The quality of prompts emerged as a key element, evidencing that using GAI requires specific competencies to formulate didactically pertinent demands [3]. The specific GAI system version used, the duration or time points of the longitudinal process, and the detailed prompt-interaction protocol followed by participants are likewise not specified in the present account.

### *3.3. Case Collection and Analysis Protocol (CCAP)*

The analysis was structured through a Case Collection and Analysis Protocol (CCAP) that systematically compares initial and mediated versions, incorporating ontosemiotic analysis of the objects and processes mobilized and the teachers' reflection on the process; content analysis followed standard procedures [21]. Table 1 presents the analytical sequence, which runs from the initial task, through the GAI interaction and the resulting mediated task, to the ontosemiotic analysis of both versions and the teacher's own reflection on the process.

Table 1: Analytical sequence of the Case Collection and Analysis Protocol (CCAP).

| Moment of analysis | Purpose | Evidence considered |
|---|---|---|
| Initial version of the task | Recognize the didactic configuration produced by the teacher without support. | Statement, objective, content, requested procedures, evaluation criteria. |
| Interaction with GAI | Identify the type of demand formulated to the tool and its orientation. | Prompts, reformulations, improvement requests, explicit criteria. |
| Mediated version | Assess the transformations introduced in the task. | Changes in context, language, representation, argumentation, sequence. |
| Ontosemiotic analysis | Characterize the mathematical objects and processes in both versions. | Languages, concepts, properties, procedures, arguments, situations. |
| Teacher reflection | Understand the professional valuation of the experience. | Narratives, comments, justified decisions, recognized difficulties. |

Source: methodological synthesis based on content analysis [21] and the ontosemiotic tools of [12]. Section 3.4 classifies this same evidence neutrosophically.

### *3.4. Neutrosophic annotation of the CCAP evidence*

The neutrosophic layer is not assigned independently after the ontosemiotic analysis of Section 3.3: it classifies the same recording units that analysis produces. The unit of neutrosophic analysis is the pair case × facet (c, f): for each task case c examined under the CCAP and each of the six didactical-suitability facets f of Section 2.2, the ontosemiotic evidence recorded for that case and that facet is read as supporting evidence (contributing to $T_{cf}$), as unresolved or context-dependent evidence (contributing to $I_{cf}$), or as counter-evidence or a documented risk (contributing to $F_{cf}$), giving the case-indexed annotation of Equation (3):

$$A_{cf} = \langle T_{cf}, I_{cf}, F_{cf} \rangle, \; c \in Cases, \quad f \in Facets \quad (3)$$

The present study does not report a numerical elicitation of T, I, and F for individual cases: the classification described above was applied qualitatively, and Table 4 aggregates it at the level of the six facets rather than reporting per-case triplets. A future quantitative implementation could estimate each coordinate from independent evidence rubrics scored per case — for instance, $T_{cf}$ as an aggregation of indicators supporting a positive contribution, $I_{cf}$ as an aggregation of indicators left unresolved, and $F_{cf}$ as an aggregation of indicators supporting a limitation or risk — evaluated independently of one another so that conflicting evidence is preserved rather than averaged away; no such numerical values were computed in the present study, and none are reported below. The ontosemiotic coding and the neutrosophic classification were performed by the author on the study's corpus of cases; no independent inter-coder agreement check is reported for either layer, a limitation returned to in Section 4.6.

## 4. Results and Discussion

### *4.1. Transformations in the tasks*

The comparison between initial and mediated versions shows substantive, consistently documented transformations in the ontosemiotic configuration of the tasks. First, a diversification of the languages used: where symbolic or numerical registers predominated, the reformulated tasks incorporate verbal, graphical, tabular, and schematic representations, establishing more complex relations among them and strengthening the construction of meaning. Second, an increase in the demand for argumentation: mediated tasks include questions oriented to explaining, justifying, and comparing procedures, displacing the focus from obtaining results toward constructing mathematical reasoning. Third, improved coherence between learning objectives and proposed activities, and the incorporation of more elaborate contexts — with the caveat that the presence of a context does not by itself guarantee quality: when it is not articulated with the mathematical content, it becomes a superficial element. Table 2 systematizes the analysis criteria per facet and the risks to control; Table 3 summarizes the observed transformations.

Table 2: Criteria for the analysis of didactical suitability in GAI-mediated tasks, and risks to control.

| **Suitability facet** | **Analysis criterion in GAI-mediated tasks** | **Risk to control** |
|---|---|---|
| Epistemic | Representativeness of concepts, procedures, properties, and arguments. | Algorithmic reduction or mathematically imprecise content. |
| Cognitive | Fit between task demand, prior knowledge, and real possibilities. | Overload from excessive instructions or abrupt jumps in complexity. |
| Mediational | Pertinence of resources, registers, time, and technological supports. | Ornamental use of technology without impact on mathematical meaning. |
| Interactional | Opportunities for discussion, explanation, validation, feedback. | Individualized activity that limits mathematical debate. |
| Affective | Capacity to generate interest, confidence, and disposition. | Artificial contexts or statements producing anxiety or demotivation. |
| Ecological | Coherence with curriculum, school level, institutional context. | Decontextualized proposals removed from real classroom conditions. |

Source: author's elaboration from the suitability criteria of the OSA [11, 12, 15].

Table 3: Ontosemiotic transformations observed in the mathematical tasks.

| **Mathematical object/process** | **Tendency in initial tasks** | **Contribution observed after GAI mediation** |
|---|---|---|
| Languages | Predominance of isolated numerical or symbolic expressions. | Greater combination of verbal language, tables, graphs, and expressions. |
| Concepts | Implicit definitions, weakly connected to the activity. | Greater explicitation of the meaning of the content. |
| Procedures | Emphasis on executing operations or applying rules. | Inclusion of strategy comparison and verification of results. |

| | | |
|---|---|---|
| Arguments | Little demand for justification or explanation. | Questions oriented to explaining, validating, and communicating reasoning. |
| Contexts | Generic situations or decontextualized exercises. | Scenarios more related to school and everyday problems. |
| Interaction | Activity centred on the individual answer. | Greater openness to discussion, socialization, and feedback. |

Source: author's comparative analysis of tasks without and with GAI mediation [6, 12].

The improvements, however, are not automatic. In some cases GAI generates open questions without a genuine argumentative demand, tasks that are excessively complex or extensive for real classrooms, or conceptual imprecisions that require critical revision [4, 18]. The mediated versions also tend to propose more structured sequences, with differentiated moments of exploration, development, and closure — at the risk of excessive length — and the use of similar prompt patterns can homogenize the resulting activities, limiting the diversity of learning experiences. These limits reinforce the study's central thesis: the technology puts professional knowledge into play rather than replacing it.

### *4.2. The mediation model*

The findings are synthesized in an interpretive model organized around five moments: the teacher's didactic intention, the formulation of the request to the tool, the generated production, the ontosemiotic evaluation, and the teacher's professional decision and redesign (Figure 1). Each moment is an instance of mediation in which the teacher intervenes to orient, adjust, and validate the design process; the quality of the task depends on the interaction between the technology and pedagogical judgement [1, 13, 15].

The last two moments of Figure 1 — ontosemiotic evaluation and professional decision — are where operator M of Section 2.5 becomes visible empirically: it is at these moments that a teacher performs the validation, adjustment, or correction that moves a facet's annotation from an unresolved state toward either a documented contribution or a documented risk. The five-moment model and operator M describe the same process at two levels of granularity: the model organizes it procedurally, and M represents its effect on the evidential status of each facet.

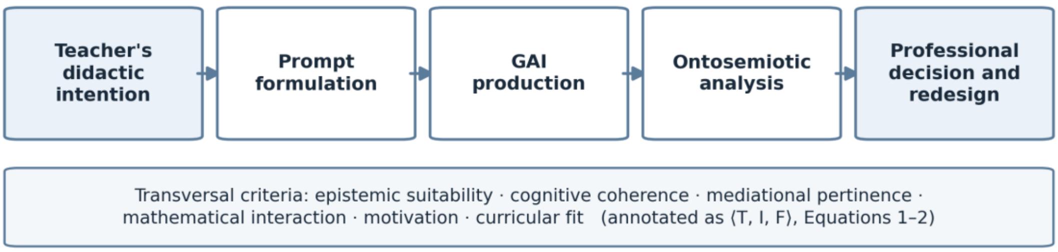


Figure 1: Interpretive model for the design of mathematical tasks mediated by GAI: five moments of mediation under the teacher's didactic control, evaluated against transversal suitability criteria annotated as $\langle T, I, F\rangle$. Source: author's elaboration; cf. [1, 13, 15].

*4.3. Neutrosophic annotation of the facets: three properties*

Table 4 applies the annotation of Section 3.4 to the six facets of didactical suitability, using the empirical material of Tables 2 and 3 and of Section 4.1: for each facet it records the OSA criterion evaluated, the evidence supporting T, the evidence supporting F, the source of the remaining I, and the teacher-mediation operation identified in this study that can refine that I. Three properties of the resulting model follow directly from the table and are illustrated with material already reported above, not with new data.

Coexistence. Positive and negative evidence coexist within the same facet rather than cancelling out. In the epistemic facet, the explicitation of concepts documented in Table 3 (support for T) coexists with the conceptual imprecision documented in Table 2 (support for F) within the same set of mediated tasks; a scalar average of the two would report a middling score that corresponds to neither situation actually observed.

Independent indeterminacy. I is not the arithmetic complement of T and F, and is not reducible to “we did not measure this”: it represents a question that remains open pending contextual or professional validation. In the mediational facet, whether a diversified resource actually serves mathematical meaning or is merely ornamental (Table 2) cannot be settled from the task statement alone; the same is true of whether an elaborated context in the affective facet (Table 3) will in fact generate interest for a specific group, which depends on that group and not on the task in isolation.

Evidence refinement. Indeterminacy is not static: it is reduced by the mediating actions of Section 2.5. In the cognitive facet, whether a more structured sequence fits a given group's pace (a source of I in Table 4) is resolved when the teacher adjusts the task's complexity and length to that group; in the ecological facet, whether a scenario fits the institutional context (Table 4) is resolved when the teacher verifies curricular coherence. In both cases the resolution can go either way — toward a validated contribution or toward an identified risk — consistent with Section 2.5's refusal to impose a favourable direction on M.

The practical import of separating T, I, and F is that a scalar judgement such as “this GAI-mediated task is suitable, score 0.75” would conceal which of these three situations produced the number: strong positive evidence undiluted by concerns, strong positive evidence masking a documented mathematical defect that happened to be outweighed in the average, or a genuinely unresolved judgement that a teacher has not yet had the opportunity to validate. Each of the three calls for a different professional response — adoption, correction, or further scrutiny — and only the triplet keeps the three visible.

Table 4: Qualitative neutrosophic evidence annotation of the six facets of didactical suitability in GAI-mediated task design, linking each OSA criterion to its supporting, contrary, and unresolved evidence and to the teacher-mediation operation that can refine the latter (Equations 1-2).

| **Facet** | **OSA evaluation criterion** | **Evidence supporting T** | **Evidence supporting F** | **Source of I / unresolved question** | **Teacher mediation M that can refine I** |
|---|---|---|---|---|---|
| Epistemic | Representativeness of concepts, procedures, properties, and | Explicitation of concepts; richer semiotic configurations | Conceptual imprecision in generated content (Table 2). | Whether the proposed meanings are faithful to the | Mathematical validation of generated content. |

| | | | | | |
|---|---|---|---|---|---|
| | arguments (Table 2). | (Table 3). | | discipline in this specific task. | |
| Cognitive | Fit between task demand, prior knowledge, and real possibilities (Table 2). | More structured sequences of exploration, development, and closure (Section 4.1). | Overload; excessive length or complexity (Table 2). | Fit to the real group's prior knowledge and pace. | Adjustment of task complexity and length. |
| Mediational | Pertinence of resources, registers, time, and technological supports (Table 2). | Diversified registers and supports (Table 3). | Ornamental technology without semantic impact (Table 2). | Whether the added resources actually serve mathematical meaning. | Selection or correction of representations. |
| Interactional | Opportunities for discussion, explanation, validation, feedback (Table 2). | Questions inviting explanation and comparison (Table 3). | Individualized activity limiting debate (Table 2). | Whether the questions elicit genuine argumentation in classroom interaction. | Orchestration of argumentation and interaction. |
| Affective | Capacity to generate interest, confidence, and disposition (Table 2). | More elaborate contexts, closer to school life (Table 3). | Artificial contexts; anxiety or demotivation (Table 2). | Whether contexts actually generate interest and confidence for this group. | Contextual adaptation and validation with the real group. |
| Ecological | Coherence with curriculum, school level, institutional context (Table 2). | Scenarios closer to school and everyday problems (Table 3). | Decontextualized proposals; homogenization (Section 4.1). | Whether the task fits the curriculum and the institutional context. | Verification of curricular coherence. |

Source: author's qualitative neutrosophic annotation of the study's empirical material (Tables 2-3 and the narrative results of Section 4.1), under the classification protocol of Section 3.4 and Equations (1)-(2); no numerical T-I-F values were elicited.

*4.4. Toward a worked case*

The strongest demonstration that the neutrosophic annotation is part of the analysis rather than a formalism added on top of it would be a single worked case tracing one initial task through its GAI-mediated reformulation, the resulting ontosemiotic change, the supporting evidence (T), the documented risk (F), the unresolved question (I), the teacher's mediating action, and the resulting professional judgement or redesign. The present manuscript does not include the verbatim task statements needed to reconstruct such a case without inventing them. Rather than substitute an invented example, this section stops at the demonstration of principle already given: Table 4 shows, for every facet, which category of evidence — T, I, or F — each documented finding of Sections 4.1–4.2 belongs to. A single-case narrative tracing one task end to end remains a natural extension for a fuller methodological report of this study.

### *4.5. Implications for teacher education*

The interaction with GAI favoured processes of awareness about teaching practice itself: comparing initial and reformulated versions allowed participants to identify patterns in their design habits — the predominance of closed exercises, the absence of argumentation, the limited diversity of representations. This recognition marks a transition from intuitive to reflective practice, in which reflection on experience is constitutive of professional knowledge [22]. Digital teaching competence is accordingly redefined: beyond technical skill, it is a form of professional judgement integrating disciplinary, didactic, and technological knowledge — the capacity to identify conceptual imprecision, value the pertinence of a representation, adjust complexity, and guarantee coherence across the design [15, 16]. Ethical dimensions — reliability of generated content, transparency of use, data protection — must be part of that formation [1].

### *4.6. Limitations of the study*

This is a qualitative study with a purposive sample of in-service teachers in one master's programme; its findings characterize processes, not population effects. As noted in Sections 3.1–3.2, the present account does not report the exact number of participants and tasks analysed, the GAI system version, or the study's duration; a fuller methodological report of this study would need to make these explicit. The ontosemiotic coding of tasks and the neutrosophic annotation of Table 4 were performed by the author on the study's corpus of cases without a reported inter-coder agreement check (Section 3.4), and the annotation is qualitative throughout: no numerical values of T, I, or F were elicited for individual cases or facets. The natural next steps are the numerical elicitation of case-level and facet-level annotations from panels of teacher educators using independent evidence rubrics, the empirical verification of the indeterminacy-reduction property $I_f' \leq I_f$ across cohorts and coders, and replication in other institutional contexts.

## 5. Conclusions

The incorporation of GAI into teacher education in mathematics transcends technological innovation: it operates on the epistemological and didactic transformation of pedagogical practice. When oriented by solid theoretical foundations, the mediation of GAI contributes to the didactical suitability of mathematical tasks across its six facets — improving conceptual coherence, the articulation of representations, the incorporation of meaningful contexts, and the demand for argumentation — and, more importantly, it triggers processes of reflection through which teachers revise their own conceptions, identify inconsistencies, and ground their decisions in didactic and epistemological criteria.

The study's three theoretical frameworks divide the analytical labour rather than competing with one another. The Ontosemiotic Approach determines what is evaluated — the objects, processes, and six facets of Section 2.2. The neutrosophic annotation represents the evidential state of that evaluation, keeping documented contributions, documented risks, and unresolved questions visible as separate coordinates instead of collapsing them into a single verdict. Teacher mediation, modelled as the operator M of Section 2.5, transforms that evidential state through validation, correction, and redesign, without either guarantee of success built into the model. In this precise sense, the authority of the teacher is not displaced by the technology: it is the condition of its didactic value. The present implementation of this framework is qualitative; Section 4.6 indicates where a quantitative or more fully traceable implementation would need additional material that the current account does not yet provide.

**Author contributions:** the author is responsible for the conceptualization, methodology, investigation, formal analysis, and writing of the manuscript.

**Conflicts of interest:** The author declares no conflict of interest.

## References


1. 1. W. Holmes, M. Bialik, and C. Fadel. Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Center for Curriculum Redesign, Boston, 2022.
2. 2. R. Luckin. AI for School Teachers. CRC Press, Boca Raton, 2023. https://doi.org/10.1201/9781003193173
3. 3. E. R. Mollick and L. Mollick. Assigning AI: seven approaches for students, with prompts. SSRN, 2023. https://doi.org/10.2139/ssrn.4475995
4. 4. M. Ramírez Bogantes and K. Mejía Luna. Inteligencia artificial generativa y educación matemática: tensiones entre el uso instrumental y el aprendizaje complejo. ASCE Magazine, 5(2) (2026), 1605-1627. https://doi.org/10.70577/asce.v5i2.844
5. 5. M. D. Pochulu. Inteligencia artificial generativa y su impacto en la educación matemática del siglo XXI. Journal Sibemir, 1(1) (2025). https://doi.org/10.63706/jsibemir.v1i1.10
6. 6. J. D. Godino, C. Batanero, and V. Font. Fundamentos de la enseñanza y el aprendizaje de las matemáticas para maestros. Universidad de Granada, Granada, 2003.
7. 7. V. Font, N. Planas, and J. D. Godino. Modelo para el análisis didáctico en educación matemática. Infancia y Aprendizaje, 33(1) (2010), 89-105. https://doi.org/10.1174/021037010790317243
8. 8. F. Smarandache. Neutrosophy: Neutrosophic Probability, Set, and Logic. American Research Press, Rehoboth, 1998.
9. 9. M. Leyva-Vázquez and F. Smarandache. Neutrosofía: nuevos avances en el tratamiento de la incertidumbre. Pons Publishing House, Brussels, 2018.
10. 10. M. Leyva-Vázquez and F. Smarandache. Neutrosophic paraconsistent logic: evidence degrees, ontological indeterminacy, and scientific evidence synthesis. Neutrosophic Computing and Machine Learning, 43 (2026), 211-221.
11. 11. J. D. Godino, A. Contreras, and V. Font. Análisis de procesos de instrucción basado en el enfoque ontológico-semiótico de la cognición matemática. Recherches en Didactique des Mathématiques, 26(1) (2006), 39-88.
12. 12. J. D. Godino, D. Bencomo, V. Font, and M. Wilhelmi. Análisis y valoración de la idoneidad didáctica de los procesos de estudio de las matemáticas. Paradigma, 27(2) (2007), 221-252.
13. 13. J. D. Godino. Indicadores de idoneidad didáctica de procesos de enseñanza y aprendizaje de las matemáticas. In: XIII Conferencia Interamericana de Educación Matemática (CIAEM), Recife, 2011.
14. 14. J. D. Godino. Enfoque ontosemiótico en educación matemática: fundamentos, herramientas y aplicaciones. McGraw Hill - Aula Magna, Madrid, 2024. https://hdl.handle.net/10481/93596
15. 15. A. Breda and V. M. Lima. Estudio de caso sobre el análisis didáctico realizado en un trabajo final de un máster para profesores de matemáticas en servicio. REDIMAT, 5(1) (2016), 74-103.

16. 16. P. Beltrán-Pellicer and B. Giacomone. Desarrollando la competencia de análisis y valoración de la idoneidad didáctica en profesores de matemáticas. Unión - Revista Iberoamericana de Educación Matemática, 14(52) (2018), 65-82.
17. 17. F. G. Montenegro. Caracterización de las tareas de matemática que se proponen en una formación continua de profesores cuando las TIC se establecen como recurso prioritario. Tesis doctoral, Universidad Católica de Santa Fe, 2022.
18. 18. M. D. Pochulu and V. Font. Idoneidad didáctica de tareas de matemáticas reformuladas con inteligencia artificial. Paradigma, 46(1) (2025), e2025027. https://doi.org/10.37618/PARADIGMA.1011-2251.2025.e2025027.id1621
19. 19. P. Sureda, A. R. Corica, and V. Parra. Inteligencia artificial generativa en la formación de profesores de matemática en servicio. UNIÓN: Revista Iberoamericana de Educación Matemática, 19(69) (2023), 1-13.
20. 20. Ministerio de Educación del Ecuador. Currículo de los niveles de educación obligatoria. Quito, 2016.
21. 21. L. Bardin. Análisis de contenido. Akal, Madrid, 2011.
22. 22. A. Bolívar, J. Domingo, and M. Fernández. La investigación biográfico-narrativa en educación: enfoque y metodología. La Muralla, Madrid, 2001.