Prototype Effects and Implicit Models in LLM Responses to Geometry Tasks

Magdalena Krátká, Michaela Tichá and Jiří Přibyl
Department of Mathematics, Jan Evangelista Purkyně University in Ústí nad Labem, Czechia

magdalena.kratka@ujep.cz
michaela.ticha@ujep.cz
jiri.pribyl@ujep.cz

Abstract: This paper investigates the role of implicit models in large language model (LLM) responses to geometry questions, with a particular focus on the phenomenon of prototype effects (Rosch, 1975). Building on Fischbein's account of tacit and implicit models in mathematical thinking (Fischbein, 1987; 1989), together with his later theory of figural concepts (Fischbein, 1993), we analyze how LLMs reproduce and reinforce such models in their explanations. The study draws on a large dataset of elementary geometry tasks in Czech language, as a case of less dominant language. The theoretical framework combines Fischbein's theory of figural concepts (Fischbein, 1993), emphasizing the dual conceptual–figurative nature of geometric objects, with research on prototype effects in geometry education. These effects manifest, for example, when angles are interpreted primarily as convex regions or when geometric figures are associated with dominant visual prototypes rather than formal definitions (Brunheira and da Ponte, 2019). We consider that LLMs can tend to systematically privilege such prototypical representations. Methodologically, we conduct a qualitative content analysis of LLM responses (ChatGPT and Claude) to geometry prompts. Unlike previous analyses focused on correctness, richness, or errors within Czech and English prompts' responses, this study codes for the presence, type, and didactical role of implicit models. Particular attention is paid to how these models align with or diverge from local curricular norms and classroom expectations. Results show recurring prototype-based representations, some of which may conflict with local curricular standards despite being mathematically valid. This creates a new layer of didactical tension: not merely between correct and incorrect answers, but between formally valid explanations and those aligned with classroom practice. We discuss this tension in relation to the didactical contract, where AI systems become co-constructors of mathematical meaning. The paper contributes to research on AI in education by showing that LLMs may reproduce and reinforce representations that could influence learners' implicit conceptual frameworks. This has significant implications for teaching: rather than treating AI outputs as neutral resources, educators must address their underlying implicit models and integrate them into reflective classroom practice.

Keywords: Artificial Intelligence in Education, Prototype Effect, Implicit Model, Didactical Contract, Geometry Education

1. Introduction and Theoretical Background

Geometry makes the tension between formal mathematical definitions and intuitive representations especially visible. This study examines how LLM-generated explanations may reproduce prototype effects and implicit models, not merely whether they are correct. The framework therefore combines cognitive didactics of geometry with research on LLMs and the didactical contract.

1.1 Geometrical Concepts Between Definitions and images

Geometrical objects are not encountered by learners as purely abstract definitions. They are also associated with diagrams, typical shapes, canonical orientations, symbolic notation, and everyday language. Fischbein's theory of figural concepts captures this duality: geometrical objects combine a conceptual component, governed by definitions and logical relations, with a figural component, linked to spatial, visual and gestalt-like properties. Successful reasoning requires conceptual control over the figural component, although in learning situations the visual or intuitive component may dominate (Fischbein, 1993).

This tension is also reflected in the van Hiele model of geometrical thinking. At the earliest levels, students tend to recognise figures holistically, on the basis of their overall appearance, orientation or visual similarity to familiar examples. Only later do they begin to analyse figures through properties, relations and definitions (van Hiele, 1986). Errors in geometry therefore need not be failures to recall definitions; they may reveal the representational models through which learners interpret an object.

This makes geometry a productive context for analysing LLM responses. An LLM explanation selects definitions, properties, examples, metaphors, symbols and visual descriptions, and may thereby reinforce particular representations of the concept. The present study consequently focuses on the geometrical representations that LLM responses implicitly privilege.

1.2 Implicit Models and Prototype Effects in Mathematical Thinking

The notion of implicit model is central to Fischbein's account of mathematical thinking. Fischbein describes implicit models as intuitive cognitive structures that operate below explicit reasoning and guide how an individual interprets and acts upon mathematical concepts (Fischbein, 1987). An implicit model is not identical with a diagram, word, symbol or example; it is an internal construct through which an abstract concept becomes mentally accessible, supported by visual, symbolic, linguistic, action-based or classificatory forms.

In geometry, such models are especially visible because geometrical concepts have both conceptual and figural aspects (Fischbein, 1993). The aim of geometry education is not to suppress the figural component, but to bring it under conceptual control. Difficulties arise when an intuitive model becomes dominant and the learner reasons from a typical representation rather than from defining properties.

Prototype theory helps describe how this dominance becomes observable. Rosch (1975) showed that categories are often organised around typical or central members rather than only strict definitions. In geometry education, prototype effects appear when learners treat one familiar representative as standing for the whole class. Hershkowitz (1990) describes prototypical judgement as evaluation according to a prototype rather than formal criteria: either a mainly visual decision based on whether a figure looks like the expected case, or the transfer of non-critical attributes of the prototype to the whole class. The desired alternative is analytical judgement based on critical attributes, definitions and inclusion relations (Hershkowitz, 1990; van Hiele, 1986).

This distinction is important because implicit models are not necessarily problematic in themselves. They are part of how learners initially make sense of abstract concepts. Didactical problems arise when teaching practices, textbooks or explanations reinforce a restricted representation without later expanding it or bringing it under definitional control. This may happen for legitimate pedagogical reasons, for example when instruction begins with simpler or more typical cases in order not to overload learners. However, if such cases remain unmarked as partial representations, they may support the formation of an inadequate implicit model.

Implicit models may be supported not only by visual images but also by symbolic, linguistic and classificatory representations. A recurrent notation or canonical example may become a privileged symbolic form through which a concept is understood. Everyday language may make an explanation accessible, but it may also blur the boundary between informal experience and mathematical meaning if it replaces precise terminology (Planas and Setati-Phakeng, 2014; Barwell et al., 2016). Similarly, partitive classifications may become entrenched when categories are repeatedly presented as mutually exclusive even in contexts where hierarchical inclusion is mathematically relevant. These phenomena differ in form, but they share the same didactical risk: a limited representation begins to function as if it were the concept itself.

Implicit models may also be embedded in curricular traditions. Definitions, classifications and preferred representations used in school mathematics are not culturally neutral; they reflect local conventions about how mathematical concepts are introduced, named and related to one another. Earlier work on LLM-generated explanations has shown that such curricular and linguistic misalignments can occur both in classification tasks and in broader patterns of accuracy, richness and didactical usefulness (Krátká, Přibyl and Tichá, 2025; Přibyl, Krátká and Tichá, 2025; Tichá, Přibyl and Krátká, 2025). When LLMs respond in a less dominant language while relying on conceptual defaults from English-language training data, they may introduce representations that are mathematically plausible but didactically misaligned with local classroom expectations.

This framework allows LLM geometry explanations to be analysed beyond the correct/incorrect distinction. We do not claim that LLMs possess implicit models in the human cognitive sense. As the study does not examine learners directly, it does not test whether or how these representations are subsequently internalised by students. Rather, we examine whether their responses contain formulations that may reproduce, authorise or strengthen prototype-based representations for learners. Such formulations may arise when LLMs repeat didactical simplifications, privilege frequent representations or overgeneralise from typical cases. Prototype effects in LLM responses are therefore relevant not as evidence of human-like cognition, but as possible influences on students' implicit models.

1.3 LLMs as Actors in the Didactical Contract

The didactical contract, originally developed by Brousseau, refers to the implicit expectations that organise the relationship between teacher, learner and mathematical knowledge (Brousseau, 2002). In traditional classroom settings, students learn not only mathematical content, but also what counts as a valid explanation, a legitimate representation, or an acceptable method. With digital technologies, this contract is no longer shaped only by

the teacher, the textbook and classroom norms. Earlier research on technology in mathematics education has shown that technological tools can transform what students and teachers expect from mathematical work, including what counts as acceptable verification, reasoning or justification (Pierce, Stacey and Wander, 2010). LLMs intensify this transformation because they provide immediate, fluent and dialogic responses that can resemble tutoring or textbook explanation rather than ordinary information retrieval (Pepin, Buchholtz and Salinas-Hernández, 2025).

When students interact with LLMs, the chatbot may therefore become a quasi-teacher, tutor or textbook-like authority, even though it is not embedded in the local curriculum and does not share the teacher's pedagogical responsibility. This is particularly important in mathematics education, where recent studies emphasise both the potential of LLMs to support explanation, feedback and problem solving, and the need for critical teacher mediation because their outputs remain variable, context-sensitive and sometimes unreliable (Ergene and Ergene, 2025; Dilling and Herrmann, 2024; Korkmaz Guler et al., 2024). In geometry, this issue is especially acute because understanding depends on precise definitions, inclusion relations, symbolic notation and the coordination of conceptual and figural components.

Technical research on AI geometry solving points in the same direction. Benchmarks such as GeoEval show that geometry remains a demanding domain for LLMs and multimodal models, especially in harder tasks and tasks requiring backward reasoning (Zhang et al., 2024). Systems that achieve stronger performance in geometry, such as AlphaGeometry and AlphaGeometry2, do not operate as ordinary chatbots; they combine neural language-model components with symbolic deduction, search mechanisms or synthetic geometry data (Trinh et al., 2024; Chervonyi et al., 2025). Similarly, recent reinforcement-learning approaches such as GeometryZero treat auxiliary construction and robust geometric reasoning as central challenges rather than solved capabilities of standard LLMs (Wang et al., 2025). For the present study, this technical literature is relevant because it suggests that fluent LLM explanations of geometry should not be interpreted as evidence of definition-based or deductive control; they may instead reflect statistically privileged textual and representational patterns.

Previous studies have shown that LLM-generated geometry explanations may contain factual errors, conceptual inconsistencies or culturally mismatched definitions (Dilling and Herrmann, 2024; Korkmaz Guler et al., 2024; Krátká, Přibyl and Tichá, 2025; Přibyl, Krátká and Tichá, 2025). Research on LLMs in education also points to a broader problem: models trained on large-scale language data may reproduce dominant conceptual patterns and may not reliably adapt them to local curricular contexts (Yan et al., 2024; Pepin, Buchholtz and Salinas-Hernández, 2025). However, the present study addresses a subtler issue. Even when LLM responses are not plainly wrong, they may still guide learners toward restricted representations of geometrical concepts. A response can be formally acceptable and at the same time didactically problematic if it privileges a static visual prototype, a canonical symbolic example, a partitive classification, or a curricular convention that is misaligned with local classroom expectations.

This shifts the focus from error detection to the analysis of representational effects. The central question is not only whether the answer is correct, but what kind of conceptual framework the answer authorises. If a chatbot repeatedly describes figures through everyday language, standard images, familiar symbols or non-inclusive classifications, students may internalise these representational defaults as part of mathematical meaning. In this sense, LLMs do not merely transmit knowledge; they may participate in the construction of geometrical concepts by stabilising some representations and marginalising others.

The theoretical framework of this study therefore connects classical cognitive didactics of geometry with current research on AI in mathematics education. Fischbein's theory explains why geometrical concepts are vulnerable to intuitive and figural representations (Fischbein, 1987; Fischbein, 1993); prototype theory explains why some representatives become privileged over others (Rosch, 1975; Hershkowitz, 1990); research on geometry learning explains how such prototypes affect classification and reasoning (van Hiele, 1986; Brunheira and da Ponte, 2019); and the didactical contract explains why AI-generated outputs may acquire authority in students' learning (Brousseau, 2002; Pierce, Stacey and Wander, 2010). Building on these perspectives, the present study analyses LLM responses not only as potentially correct or incorrect answers, but as textual environments that may support, reproduce or challenge prototype-based reasoning in geometry.

2. Methodology

2.1 Research Design and Data Collection

This study employs qualitative content analysis combined with quantitative to examine how LLMs reproduce implicit models and prototype effects in elementary geometry. The analysis presented here focuses on two widely used models — ChatGPT and Claude — responding to geometry prompts in Czech.

The prompt set consisted of 46 Czech-language questions covering elementary planar and spatial geometry. The plane-geometry prompts addressed quadrilaterals, triangles, regular polygons, circles, disks, ellipses, symmetry, relative positions of figures, and divisions of figures into congruent or equal-area parts. The spatial-geometry prompts concerned basic solids such as cubes, cuboids, prisms, pyramids, cones, cylinders, spheres, regular polyhedra, frustums and solids of revolution.

The prompts were not designed as a standard achievement test. Rather, they were constructed to elicit explanations in contexts where prototype-based reasoning and restricted representations are likely to become visible. They included questions about properties of individual figures, comparisons between related objects, classification and inclusion relations, positional configurations, and boundary or non-prototypical cases. All prompts were phrased as short, natural Czech questions resembling the way a student might ask for help from a chatbot. The full list of prompts is available in the online supplementary material: https://osf.io/tupj2/.

To approximate the conditions under which a typical student would interact with these systems, all prompts were posed in Czech and submitted to the freely accessible, non-paid tiers of each model — that is, the default models a non-subscribing user is served once any allowance of premium messages is exhausted. The queries were issued on 26 May 2026 through the OpenRouter API, using the model identifiers openai/gpt-5.4-mini for ChatGPT and anthropic/claude-4.6-sonnet for Claude. Each of the 46 prompts was submitted once to each model, yielding a corpus of 92 Czech-language responses. Posing the questions in Czech and using the free-tier models was a deliberate ecological choice: it reproduces the representational defaults a Czech student would actually encounter, rather than those of a premium configuration optimised for expert users.

2.2 Coding Scheme

Responses were coded in Atlas.ti using a theoretically grounded scheme developed prior to analysis. Drawing on Fischbein's theory of implicit models and figural concepts (Fischbein, 1989, 1993) and on research on prototype effects in geometry education the scheme distinguishes six categories described in Table 1. Coding was conducted independently by two researchers; disagreements were resolved through discussion until consensus was reached.

2.3 Analytical Procedure

Coded data were exported from Atlas.ti and processed in Python (pandas, scipy). For inferential analysis, segments were aggregated to the level of the individual question — the counts of each category were summed across all segments within a model's response to a given prompt. The counts were computed for all 46 questions for each model, including zeros for questions in which a given phenomenon did not occur.

Table 1: Coding categories

Coding Category	Brief Characterization	Illustrative Example
Intuitive language	Use of informal, everyday or metaphorical language instead; may reinforce pseudo-concepts and non-formal modes of expression.	Using *roh* (corner/horn) instead of the geometrical term *vrchol* (vertex), or saying that two lines *kříží se* (cross each other) rather than *protínají se* (intersect).
Visual dominance	Description or representation of a geometrical object mainly through its standard visual appearance, frequent orientation, or static mental image.	Presenting a parallelogram as necessarily "*slanted*" or a square as normally standing on a horizontal base.

Coding Category	Brief Characterization	Illustrative Example
Symbolic prototype	Exclusive or recurrent use of one canonical symbolic example, notation, or formula as a stand-in for an entire category; fixation on a standard formal representation	Always using △ABC for a triangle or replacing the Pythagorean Theorem with a symbolic expression a2+b2=c2.
Prototype overgeneralization	Transfer of properties of a prototypical case to the whole class; non-critical attributes are treated as defining.	Stating that the altitude of a triangle always lies inside the triangle, based on the prototype of an acute triangle.
Classification conflict	Treatment of inclusive hierarchical relations as if they were disjoint categories; failure to recognise that a subtype belongs to a broader class, or vice versa.	Claiming that a cube is not a prism because it is a cube.
Curricular model	Application of an implicit model rooted in a specific curricular tradition in a mismatched context, especially the use of Anglo-American definitions in Czech-language responses.	Defining a trapezoid as a quadrilateral with at least one pair of parallel sides — the Anglo-American inclusive definition — rather than exactly one pair, as required by the Czech curriculum.

3. Results

3.1 Overall Frequency of Implicit Model Types

Coding the 92 Czech-language responses yielded 368 substantive coded segments (excluding Repeat): 145 for ChatGPT and 223 for Claude. Because a segment may carry more than one code, these correspond to 392 individual code instances. Figure 1 shows the overall distribution. Intuitive language was by far the most frequent (131 instances), followed by Visual dominance (113). The remaining four categories were considerably rarer, with Curricular model the least frequent (21).

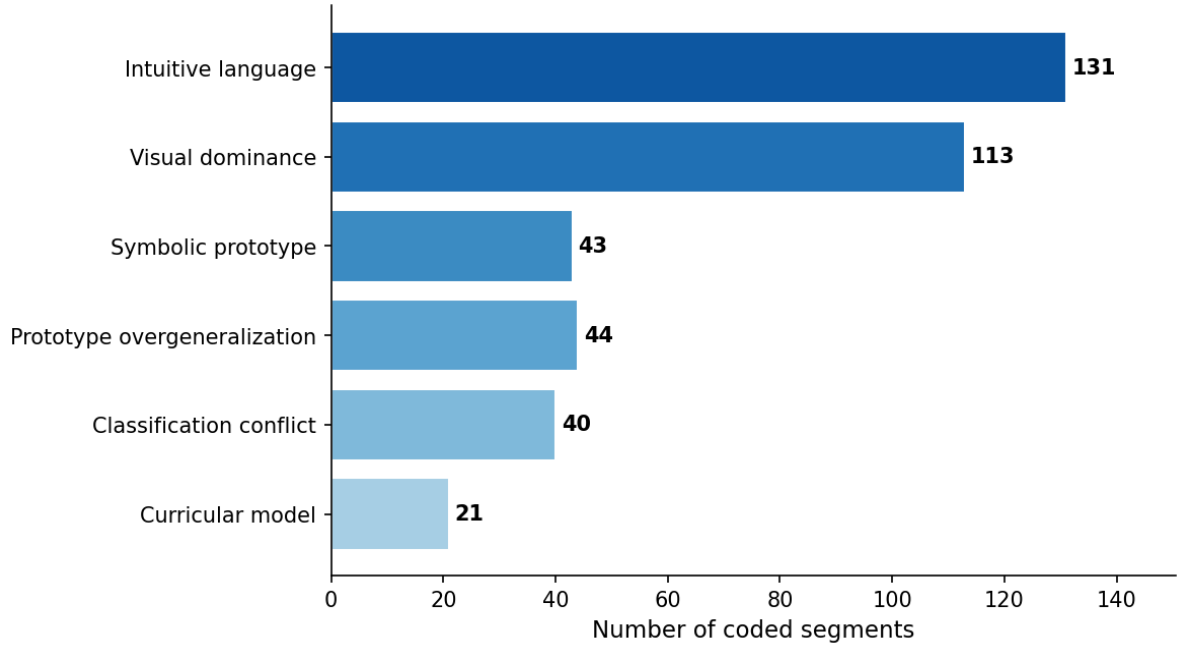


Figure 1: Overall frequency of implicit model types across both models (Czech responses, main analysis, Repeat excluded).

Disaggregating by model (Figure 2), Claude produced more instances than ChatGPT in every category except Classification conflict, where the two were equal (20 each). The largest absolute gaps appeared in Visual dominance (Claude 70 vs. GPT 43), Intuitive language (77 vs. 54), and Symbolic prototype (34 vs. 9).

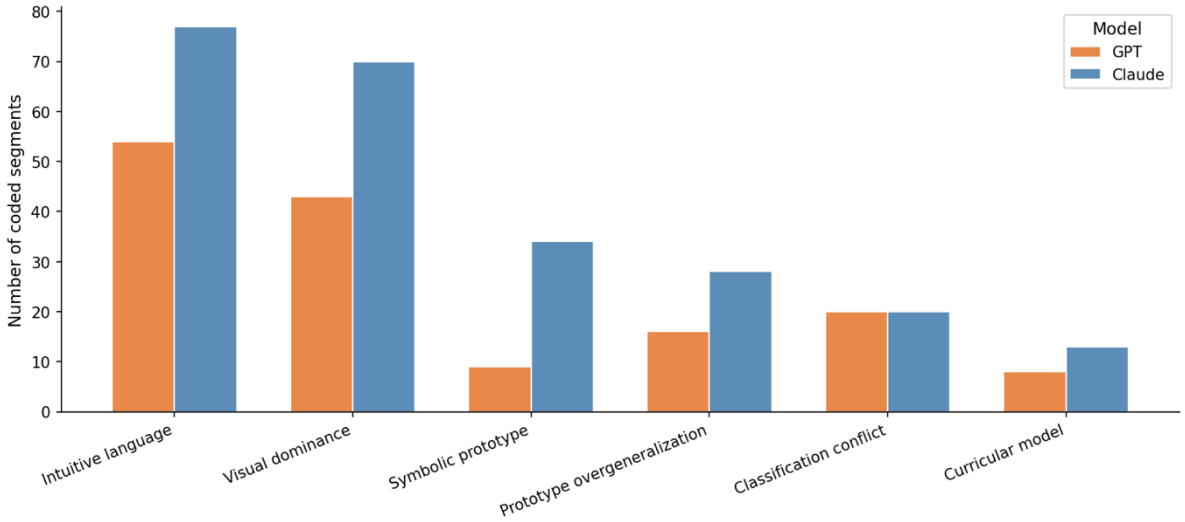


Figure 2: Frequency of implicit model types by model (absolute code instances).

At the question level, Claude produced a mean of 5.26 coded occurrences per question (median = 5), compared with 3.26 for ChatGPT (median = 3). This descriptive difference is consistent with the higher overall number of code instances observed for Claude across most categories.

It should be noted, however, that this difference may be partly attributable to longer responses; response length was not controlled for in the present analysis.

3.2 Questions Most Susceptible to Implicit Model Activation

Implicit model activation was not uniform across prompts (Figure 3). The questions eliciting the most occurrences concerned the distinction between convex and concave quadrilaterals, the division of a cube into two parts, the mutual positions of two squares or two disks, and the properties of regular hexagons and polyhedra. These items share a common feature: they require reasoning about boundary cases, inclusion relations, or abstract generalisation — precisely the contexts in which intuitive shortcuts and canonical visual images are most readily activated. The heat-map also shows category specialisation by topic: Symbolic prototype clusters on the angle-sum and two-disks questions, whereas Visual dominance dominates the convex/concave and positional-configuration questions.

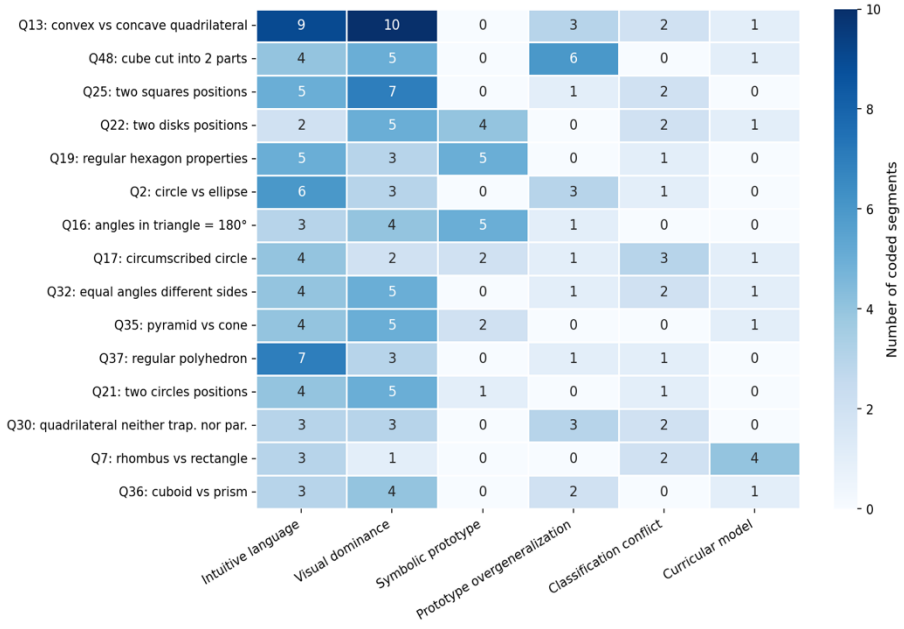


Figure 3: The fifteen questions with the highest total implicit model occurrences, broken down by category (both models combined).

4. Discussion and Conclusions

The results confirm that LLM-generated explanations in geometry may reproduce features associated with implicit models and prototype effects even when they do not contain explicit factual errors. This finding is consistent with Fischbein's view of geometrical concepts as figural concepts, in which conceptual and visual components must be coordinated (Fischbein, 1993). In the analysed responses, this coordination was often weakened by formulations that privileged intuitive language, standard visual images or canonical symbolic forms. The two most frequent categories, Intuitive language and Visual dominance, therefore point to a central didactical risk: LLMs may make geometrical concepts accessible to students, but at the same time stabilise restricted representations of these concepts.

The category-level results also illuminate how this happens. The fact that implicit model activation was not uniform across prompts, but concentrated on questions involving convex–concave distinctions, division of solids, mutual positions of figures, and properties of regular polygons and polyhedra, supports the theoretical expectation that prototype effects surface most strongly in boundary cases, inclusion relations and abstract generalisations. These are the contexts where, as Rosch (1975) and Brunheira and da Ponte (2019) argue, reasoning from a central category member rather than from formal definitions is most likely to mislead. The topical specialisation visible in the data — Symbolic prototype clustering on angle-sum and two-disk questions, Visual dominance on convex/concave and positional configurations — indicates that the models reproduce specific didactical simplifications rather than a single undifferentiated bias.

This interpretation is also supported by the technical literature on AI geometry solving discussed in Section 1.3. Standard chatbots should not be understood as systems with robust symbolic or deductive control over geometrical reasoning. Unlike specialised geometry-solving systems that combine language-model components with symbolic deduction, search mechanisms or synthetic geometry data (Trinh et al., 2024; Chervonyi et al., 2025), ordinary LLMs generate fluent explanations through statistically privileged textual and representational patterns. This distinction is also consistent with benchmark studies showing that geometry remains difficult for LLMs, especially in tasks requiring structured reasoning and backward deduction (Zhang et al., 2024; Wang et al., 2025). The implicit models identified in this study can therefore be interpreted not as evidence of human-like geometrical cognition, but as didactically relevant traces of the representations that LLMs tend to privilege in geometry explanations.

The findings also extend the discussion of the didactical contract (Brousseau, 2002). When a chatbot gives immediate and fluent responses, it can enter this contract as a quasi-authoritative source. The present study shows that this authority concerns not only correctness, but also the implicit models authorised by the explanation. Through the LLM's answers, students can repeatedly encounter geometry through informal wording, visually dominant examples or prototypical cases. These forms may become part of their understanding of the concept. In this sense, LLMs may participate in the construction of mathematical meaning, as suggested by recent research on technology-mediated didactical contracts (Pierce, Stacey and Wander, 2010) and on LLMs in mathematics education (Pepin, Buchholtz, and Salinas-Hernández, 2025).

The inclusion of two different LLMs strengthens the interpretation that prototype effects and implicit models are not isolated artefacts of a single system. Across the corpus, both models activated similar types of representational phenomena, especially intuitive language and visual dominance, while less frequent categories such as symbolic prototypes, prototype overgeneralisation, classification conflicts and curricular models also appeared. The relevance of this finding lies not in differences between the models, but in the recurrence of the same didactical phenomena across independently generated responses. This suggests that prototype-based and otherwise restricted representations may be a more general feature of LLM-generated geometry explanations, rather than an exceptional property of one particular chatbot.

4.1 Limitations

This study has several limitations. It analysed Czech-language responses from two publicly accessible LLMs to examine whether implicit-model phenomena occur across different systems. Each prompt was submitted once, although LLM outputs may vary across sessions. The analysis did not control for response length or content richness, so code frequencies cannot be interpreted as normalised measures of prototype-effect strength. The study was also limited to text-based prompts, and the identification of implicit models involved interpretative judgement. Future work will extend the analysis to English responses, repeated prompt submissions and multimodal geometry tasks. Most importantly, the study examined LLM outputs rather than learners'

interactions with them; it therefore cannot establish whether the identified representations actually influence students' conceptual understanding. This question requires studies involving learners directly.

4.2 Recommendations for Practice

For practice, the results suggest that AI-generated explanations should be treated as objects of mathematical and didactical reflection rather than as neutral learning resources. Because LLM responses are often fluent and authoritative while still privileging restricted representations, teachers should not outsource geometrical explanation to the model. Instead, chatbot responses can be used to examine which representation of a concept is being privileged, whether an explanation relies on a typical visual image or on a formal definition, and whether a stated property holds in boundary or non-prototypical cases. Particular attention should be paid to the question types identified in this study as especially susceptible to implicit-model activation: boundary cases, inclusion relations, alternative orientations, and curricular definitions where Anglo-American and Czech conventions may diverge. By explicitly surfacing and discussing the implicit models embedded in AI-generated responses, teachers can turn a potential source of misconception into an opportunity for developing analytical, definition-based reasoning. Such activities can help students move from prototypical judgement towards analytical judgement in the sense of van Hiele (1986) and Hershkowitz (1990).

Ethics Declaration

This study did not involve human participants or personal data and therefore did not require ethical approval.

AI Declaration

This study examines the output behaviour of LLMs, which also served as the primary data sources for the empirical analysis. Responses were collected from two publicly accessible AI systems, ChatGPT and Claude, using the model identifiers openai/gpt-5.4-mini and anthropic/claude-4.6-sonnet via the OpenRouter API. The generated Czech-language responses were subsequently coded, analysed and interpreted by the authors.

During manuscript preparation, the authors used ChatGPT and Notebook LM to support selected aspects of English-language phrasing, editorial refinement and structural organisation. All analytical decisions, coding procedures, interpretation of results and formulation of research conclusions were conducted exclusively by the human authors. Any AI-assisted text was carefully reviewed, edited and validated prior to inclusion.

References

Barwell, R., Clarkson, P., Halai, A., Kazima, M., Moschkovich, J., Planas, N., Setati-Phakeng, M., Valero, P. and Villavicencio Ubillús, M. (eds) (2016) *Mathematics Education and Language Diversity: The 21st ICMI Study*. Cham: Springer. doi: 10.1007/978-3-319-14511-2.

Brousseau, G. (2002) *Theory of Didactical Situations in Mathematics: Didactique des mathématiques, 1970–1990*. Edited and translated by N. Balacheff, M. Cooper, R. Sutherland and V. Warfield. Dordrecht: Kluwer Academic Publishers. doi: 10.1007/0-306-47211-2.

Brunheira, L. and da Ponte, J.P. (2019) 'From the classification of quadrilaterals to the classification of prisms: An experiment with prospective teachers', *The Journal of Mathematical Behavior*, 53, pp. 65–80. doi: 10.1016/j.jmathb.2018.06.004.

Chervonyi, Y., Trinh, T.H., Olšák, M., Yang, X., Nguyen, H.H., Menegali, M., Jung, J., Kim, J., Verma, V., Le, Q.V. and Luong, T. (2025) 'Gold-medalist performance in solving olympiad geometry with AlphaGeometry2', *Journal of Machine Learning Research*, 26, pp. 1–39.

Dilling, F. and Herrmann, M. (2024) 'Using large language models to support pre-service teachers' mathematical reasoning—An exploratory study on ChatGPT as an instrument for creating mathematical proofs in geometry', *Frontiers in Artificial Intelligence*, 7, 1460337, pp. 1–14. doi: 10.3389/frai.2024.1460337.

Ergene, O. and Ergene, B.C. (2025) 'AI ChatBots' solutions to mathematical problems in interactive e-textbooks: Affordances and constraints from the eyes of students and teachers', *Education and Information Technologies*, 30, pp. 509–545. doi: 10.1007/s10639-024-13121-z.

Fischbein, E. (1987) *Intuition in Science and Mathematics: An Educational Approach*. Dordrecht: D. Reidel.

Fischbein, E. (1989) 'Tacit models and mathematical reasoning', *For the Learning of Mathematics*, 9(2), pp. 9–14. http://www.jstor.org/stable/40247948

Fischbein, E. (1993) 'The theory of figural concepts', *Educational Studies in Mathematics*, 24, pp. 139–162. doi: 10.1007/BF01273689.

Hershkowitz, R. (1990) 'Psychological aspects of learning geometry', in Nesher, P. and Kilpatrick, J. (eds) *Mathematics and Cognition: A Research Synthesis by the International Group for the Psychology of Mathematics Education*. Cambridge: Cambridge University Press, pp. 70–95. doi: 10.1017/CBO9781139013499.006.

Korkmaz Guler, N., Dertli, Z.G., Boran, E. and Yildiz, B. (2024) 'An artificial intelligence application in mathematics education: Evaluating ChatGPT academic achievement in a mathematics exam', *Pedagogical Research*, 9(2), em0188. doi: 10.29333/pr/14145.

Krátká, M., Přibyl, J. and Tichá, M. (2025) 'AI in the Classroom: Didactical Misalignments in Geometry Between Czech and Anglo-Saxon Contexts', accepted for *European Conference on e-Learning (ECEL 2025)*.

Pepin, B., Buchholtz, N. and Salinas-Hernández, U. (2025) 'A scoping survey of ChatGPT in mathematics education', *Digital Experiences in Mathematics Education*, 11, pp. 9–41. doi: 10.1007/s40751-025-00172-1.

Pierce, R., Stacey, K. and Wander, R. (2010) 'Examining the didactic contract when handheld technology is permitted in the mathematics classroom', *ZDM Mathematics Education*, 42, pp. 683–695. doi: 10.1007/s11858-010-0271-8.

Planas, N. and Setati-Phakeng, M. (2014) 'On the process of gaining language as a resource in mathematics education', *ZDM – Mathematics Education*, 46, pp. 883–893. doi: 10.1007/s11858-014-0610-2.

Přibyl, J., Krátká, M. and Tichá, M. (2025) 'When AI teaches geometry wrong: Systematic errors in LLM-generated explanations and their educational risks', accepted for *International Conference of Education, Research and Innovation (ICERI 2025)*. doi: 10.21125/iceri.2025.1357.

Rosch, E. (1975) 'Cognitive representations of semantic categories', *Journal of Experimental Psychology: General*, 104(3), pp. 192–233. doi: 10.1037/0096-3445.104.3.192.

Tichá, M., Přibyl, J. and Krátká, M. (2025) 'English-Czech Output Bias in LLMs: A Geometry-Based Case Study', accepted for *International Conference on AI Research (ICAIR 2025)*. doi: 10.34190/icair.5.1.4326

Trinh, T.H., Wu, Y., Le, Q.V., He, H. and Luong, T. (2024) 'Solving olympiad geometry without human demonstrations', *Nature*, 625, pp. 476–482. doi: 10.1038/s41586-023-06747-5.

van Hiele, P.M. (1986) *Structure and Insight: A Theory of Mathematics Education*. Orlando: Academic Press.

Wang, Y., Wang, Y., Wang, D., Peng, Z., Guo, Q., Tao, D. and Wang, J. (2025) 'GeometryZero: Improving geometry solving for LLM with group contrastive policy optimization', *arXiv*. doi: 10.48550/arXiv.2506.07160.

Yan, L., Sha, L., Zhao, L., Li, Y., Martínez-Maldonado, R., Chen, G., Li, X., Jin, Y. and Gašević, D. (2024) 'Practical and ethical challenges of large language models in education: A systematic scoping review', *British Journal of Educational Technology*, 55(1), pp. 90–112. doi: 10.1111/bjet.13370.

Zhang, J., Li, Z.-Z., Zhang, M.-L., Yin, F., Liu, C.-L. and Moshfeghi, Y. (2024) 'GeoEval: Benchmark for evaluating LLMs and multi-modal models on geometry problem-solving', in *Findings of the Association for Computational Linguistics: ACL 2024*. Bangkok, Thailand: Association for Computational Linguistics, pp. 1258–1276. doi: 10.18653/v1/2024.findings-acl.73.