International Journal of Science Academic Research
Vol. 07, Issue 05, pp.811-817, May 2026
Available online at http://www.scienceijsar.com

ISSN: 2582-6425

Research Article

CONVERSATIONAL AI FOR PRIMARY EDUCATION: DESIGN OF A CULTURALLY-AWARE LLM ASSISTANT

***Alexandr Veaceslav Parahonco**

Laboratory "Programming Systems – IurieRogojin", Vladimir Andrunachievici Institute of Mathematics and Computer Science, Moldova State University, Chișinău, MD-2028, Moldova

Received 17th March 2026; **Accepted** 20th April 2026; **Published online** 15th May 2026

Abstract

This paper presents AI Guguță, a culturally aware conversational assistant designed for primary education in Moldova. The system combines hybrid large language model pipelines with Retrieval-Augmented Generation (RAG) to provide curriculum-grounded and age-appropriate tutoring for Romanian language and mathematics. Unlike conventional educational chatbots, AI Guguță integrates cultural adaptation through a pedagogical persona inspired by Moldovan children's literature together with collections of traditional Moldovan proverbs. The assistant further employs Socratic-style conversational strategies that encourage guided reasoning and critical thinking rather than direct answer generation. The paper describes the system architecture, RAG workflow, pedagogical design principles, and conversational interaction mechanisms underlying the assistant. The proposed framework demonstrates the feasibility of culturally grounded AI tutoring systems for primary education in low-resource linguistic and educational contexts.

Keywords: Conversational AI, Large Language Models; Retrieval-Augmented Generation, Primary Education, Culturally-Aware AI, Intelligent Tutoring Systems, Socratic Learning, Low-Resource Languages.

INTRODUCTION

Large language models (LLMs) are rapidly entering educational practice as tutoring, feedback, and content-generation tools. Recent empirical reviews, synthesizing over 88 studies from 2022 to 2025, demonstrate that intelligent tutoring systems (ITS) have become one of the most prominent applications of LLMs in education [1]. These systems represent a shift from traditional chatbots to sophisticated pedagogical agents capable of engaging learners in natural language dialogue, adapting to individual learning paces, and providing dynamic feedback [2, 3]. The integration of Retrieval-Augmented Generation (RAG) has further advanced these systems by grounding their responses in authoritative educational content, thereby mitigating hallucinations and ensuring alignment with specific curricula [4, 5]. However, growing adoption does not imply pedagogical readiness for young learners. Primary education, particularly at the 4th-grade level, demands a specialized approach that goes beyond merely providing correct answers. Off-the-shelf models often fail to produce grade-appropriate explanations, and recent studies show persistent gaps in curriculum alignment, contextual relevance, and cultural grounding [6]. For instance, research on "Classroom AI" highlights that off-the-shelf LLMs struggle to maintain grade-appropriate responses without specific developmental calibration, underscoring that age calibration is an independent research challenge, not merely a cosmetic adjustment [7]. Effective learning at this stage relies heavily on the development of critical thinking and problem-solving skills competencies that are often bypassed when an AI simply dispenses direct solutions [8]. A more profound limitation of current educational AI is the pronounced linguistic and cultural divide.

Corresponding Author:** ***Alexandr Veaceslav Parahonco
Laboratory "Programming Systems – IurieRogojin", Vladimir Andrunachievici Institute of Mathematics and Computer Science, Moldova State University, Chișinău, MD-2028, Moldova.

The vast majority of LLM development and research is concentrated on high-resource languages, leaving low-resource languages and their specific cultural contexts severely underserved [9]. Cultural adaptation in educational AI extends far beyond simple multilingualism or interface translation; it requires cultural sensitivity for inclusivity and the ability to utilize locally relevant examples [8]. Education is inherently a cultural transmission process, yet most global AI systems operate in a cultural vacuum, devoid of local context, traditions, and historical narratives that resonate with the student's lived experience. The absence of culturally-aware AI interventions for underrepresented communities not only hinders educational equity but also risks linguistic homogenization [11]. At the same time, policy and design guidance increasingly emphasize child-centred, teacher-guided, and locally grounded AI for learning [12, 13]. These trends motivate the need for an assistant designed not only for correctness, but also for developmental appropriateness and cultural relevance in primary school contexts. To address these intersecting challenges, this paper proposes AI Guguță: a culturally-aware, curriculum-grounded, age-calibrated conversational assistant for primary education in the Republic of Moldova.

RELATED WORK

The development of conversational AI for primary education sits at the intersection of multiple distinct research domains: intelligent tutoring systems, prompt-based educational agents, retrieval-augmented generation (RAG) applications, and the challenges of low-resource languages. This section reviews recent advancements across these areas and highlights the critical gaps that our proposed system, AI Guguță, seeks to address.

Large language models in education and conversational agents

The integration of LLMs into educational technology has spurred a shift from traditional computer-assisted instruction toward sophisticated conversational AI tutoring systems. Recent empirical reviews, such as the DOAJ/Elsevier review of 88 studies from 2022 to 2025, demonstrate that intelligent tutoring systems (ITS) are among the most prominent applications of LLMs in education, providing benefits such as academic performance improvement and personalized learning [1, 10]. Furthermore, the transition from simple chatbots to LLM agents for education introduces architectural capabilities like memory, tool use, and multi-agent communication [14]. Conceptually we can represent the important features in Figure 1.

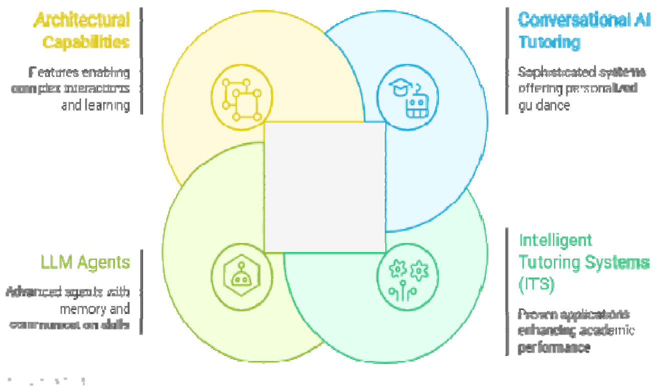

Fig. 1. The Evolution of AI in Education

Despite these advancements, the field remains fragmented, and a significant limitation persists: the majority of conversational AI systems are optimized for adult learners or secondary education, often lacking the pedagogical nuance required for primary school students [15]. For example, the "Path to Conversational AI Tutors" framework proposes maintaining knowledge tracing and affect-related components from traditional ITS while transitioning to generative dialogue, emphasizing the need to center student agency and diagnose reasoning [2]. AI Guguță directly addresses this by functioning as a pedagogically constrained student-support agent rather than a generic dialogue interface

Educational design for primary school learners

The effectiveness of conversational AI in primary education is heavily reliant on rigorous prompt engineering and developmental calibration. Research on "Classroom AI" highlights that off-the-shelf models often fail to produce grade-appropriate explanations, demonstrating that large language models as grade-specific teachers require a readability-aware pipeline [7]. In a study with 208 participants, such a pipeline achieved a 35.64 percentage point improvement in grade-level alignment while maintaining accuracy, proving that primary education requires distinct developmental calibration, not merely simpler prompts [7]. Pedagogical prompting further structures these interactions. The Socratic LLM framework, which created 6,846 dialogues covering 513 knowledge points in primary school mathematics, demonstrates that the instructional structure of a conversation can be engineered through data and design [16]. Similarly, connecting prompting with dialogic pedagogy and the Socratic method enhances student-AI interaction [17]. AI Guguță integrates these pedagogical designs to ensure age-appropriate, step-by-step reasoning rather than direct answer provision.

Retrieval-augmented generation systems in education

To address the inherent issues of hallucinations and factual inaccuracies in generative AI, researchers have increasingly turned to Retrieval-Augmented Generation (RAG) frameworks. Since its introduction in 2020 as a method to combine parametric memory with external non-parametric sources, RAG has become a specialized direction in educational AI, with over 51 studies by 2025 highlighting its benefits for dynamic knowledge updates and factual grounding [18, 19]. RAG in education is not just a technical technique but a method to rely on curriculum-aligned content, explaining concepts "by the book" rather than hallucinating [20]. For instance, the EduX-RAG system implemented a cross-lingual educational chatbot across 13 languages, achieving high accuracy in language detection and material citation [21]. While this validates the multilingual RAG architecture, it does not fully address cultural adaptation or grade-specific pedagogy. AI Guguță builds upon this by utilizing a dual-collection RAG system one for the cultural context (Moldavian proverbs) and another for the curriculum (course materials) providing a robust grounding layer specifically tailored for local primary schools. For instance, the EduX-RAG system implemented a cross-lingual educational chatbot across 13 languages, achieving high accuracy in language detection and material citation [21]. While this validates the multilingual RAG architecture, it does not fully address cultural adaptation or grade-specific pedagogy. AI Guguță builds upon this by utilizing a dual-collection RAG system one for the cultural context (Moldavian proverbs) and another for the curriculum (course materials) providing a robust grounding layer specifically tailored for local primary schools.

Culturally-aware AI and low-resource languages

The linguistic and cultural divide in AI development is a growing concern, particularly for underrepresented educational settings. While initiatives exist to support low-resource languages, they are rarely integrated into comprehensive educational tutoring systems with deep cultural adaptation. Cultural adaptation is a distinct design challenge requiring teacher involvement, as demonstrated by the CulturAIEd tool, which helped K-12 teachers adapt AI literacy activities to students' cultural contexts [22]. Furthermore, framing AI as a cultural mediator in classroom interactions provides a conceptual basis for culturally responsive pedagogy [11]. Practical deployments in low-resource environments further validate the potential of conversational tutors. The Rori system in Ghana, an AI-powered conversational math tutor deployed via WhatsApp for grades 3-9, achieved a significant effect size in low-bandwidth environments, proving the viability of chat-oriented tutor systems outside ideal infrastructural conditions [23]. Similarly, the MultiAiTutor framework demonstrated the potential of child-friendly multilingual tutoring systems for low-resource language contexts, achieving strong human preference evaluations compared to baseline approaches. [24]. AI Guguță extends this paradigm by specifically targeting the Moldavian variant of Romanian, acting as a cultural mediator through the persona of Guguță.

RESEARCH GAPS

Prior research has explored several important dimensions of educational conversational AI, including conversational tutoring, retrieval-grounded educational assistants, pedagogical prompting strategies, multilingual learning support, and culturally responsive educational systems. Studies on grade-specific tutoring emphasize the importance of developmental adaptation, while Retrieval-Augmented Generation (RAG) approaches improve factual grounding and alignment with curriculum materials. At the same time, research in culturally responsive pedagogy increasingly highlights the role of local linguistic and cultural context in shaping educational interaction. Despite these advances, the existing literature largely treats these directions independently. Current systems rarely combine age-appropriate conversational tutoring, curriculum-grounded retrieval mechanisms, and culturally contextualized interaction within a unified framework designed for real primary-school environments. This limitation appears particularly visible in low-resource linguistic contexts, where educational AI systems often rely on generic interaction patterns and globally oriented datasets. AI Guguță seeks to address this gap through the integration of conversational tutoring, curriculum-grounded RAG workflows, and culturally aware interaction strategies rooted in Moldovan educational and literary context. Rather than focusing on direct response generation, the assistant employs Socratic-style guidance intended to support reasoning, reflection, and gradual learner engagement.

ARCHITECTURAL AND PEDAGOGICAL DESIGN OF AI GUGUȚĂ

AI Guguță was designed as a hybrid educational assistant that combines large language models with a dual-collection Retrieval-Augmented Generation (RAG) architecture. The technical objective behind this design was not only to improve factual consistency, but also to maintain close alignment with the Moldovan primary-school curriculum. At the same time, the project deliberately moves beyond a purely engineering-oriented perspective. Cultural grounding became an equally important design principle during development. Rather than presenting the assistant as a generic educational chatbot, the system adopts the persona of Guguță, a well-known figure from Moldovan children's literature. The conversational environment is further enriched through the integration of carefully selected Moldovan proverbs and culturally familiar linguistic patterns. This combination was intended to create an interaction style that feels recognizable and emotionally accessible for young learners while still preserving educational rigor (see Fig. 2). From the implementation perspective, the platform relies on a relatively lightweight but flexible technology stack. Ollama is used for local deployment of language models, Streamlit provides the user interface layer, and Chroma DB supports the retrieval pipeline. Between the interface and the main language model, the system introduces an intermediate analytical layer responsible for preprocessing and restructuring user requests. This layer proved particularly important during experimentation. Instead of forwarding raw user input directly to the model, the system first evaluates the structure and context of the request. In some situations, additional contextual information is attached automatically. In others, the query is reformulated into a markdown-oriented structure that improves interpretability for downstream reasoning. Although technically simple, this stage noticeably improved response stability and reduced cases of pedagogically inappropriate generation (see Fig. 3).

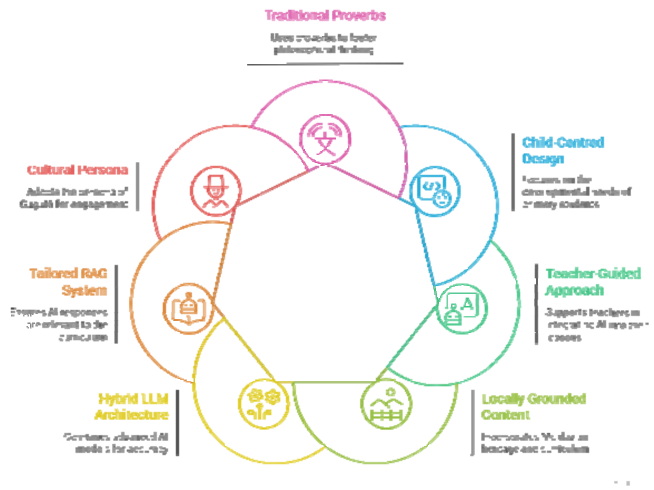

Fig. 2. AI Guguță as a culturally grounded educational assistant

Considerable attention was also devoted to the visual and interaction design of the platform. Since the assistant targets primary-school learners, the interface intentionally avoids the appearance of a conventional productivity application. Instead, the main page was designed to resemble the cover of a school textbook. The goal was subtle but important: the system should feel like part of the educational environment rather than an external technological instrument.

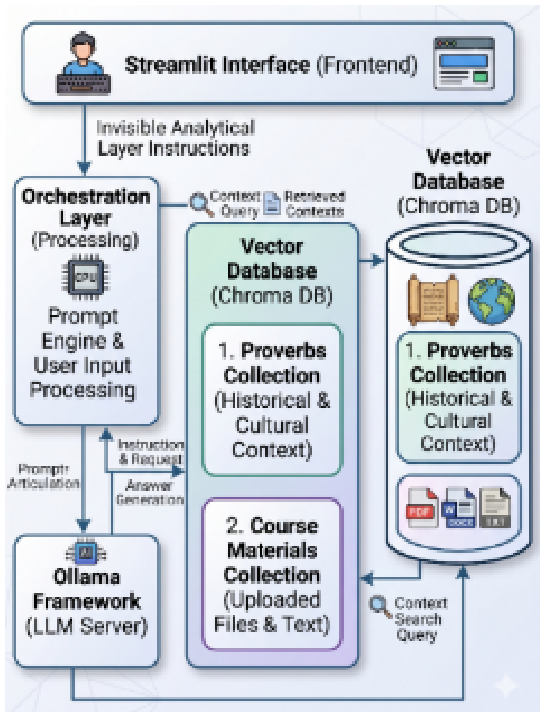

Fig. 3. Hybrid conversational and retrieval architecture of AI Guguță

The visual identity in the Figure 4 of the assistant follows the same principle. Cultural motifs inspired by Moldovan educational and literary traditions are reflected throughout the interface and remain consistent across different subjects and interaction modes. While these decisions may appear secondary from a technical standpoint, they significantly influence the emotional tone of interaction.

Fig. 4. Textbook-inspired user interface for primary-school interaction

A central component of the system is the pedagogical persona referred to as "Professor Guguță" (see Fig. 5). The assistant was designed to behave as a patient, attentive, and supportive tutor, particularly during mathematical problem solving. This behavior is implemented through a persistent system prompt together with a collection of carefully prepared few-shot interaction examples.

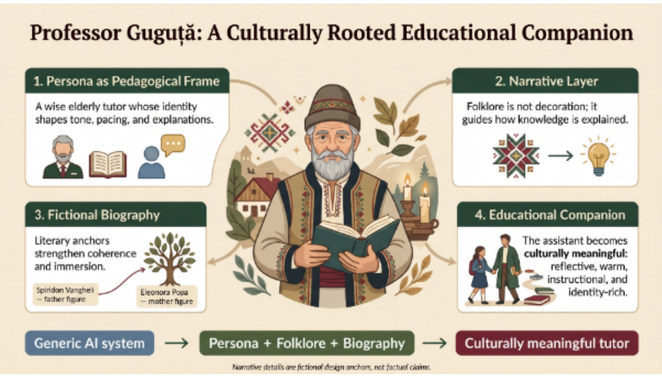

Fig. 5. Conceptual foundations of the Professor Guguță pedagogical persona

The persona itself emerged from the combination of two distinct influences. The first derives from the teaching philosophy of Prof. Boris IvanoviciCinic, whose pedagogical style combined academic rigor with accessible explanation and conversational warmth. The second originates from the literary image of Guguță, a culturally familiar character associated with kindness, wisdom, and simplicity. By merging these influences, the system attempts to balance intellectual guidance with emotional accessibility. Professor Guguță is therefore framed not simply as a virtual assistant, but as an experienced teacher capable of integrating folklore, reflective questions, and everyday examples into the learning process (see Fig. 5). This narrative layer is not merely decorative. It directly shapes the pacing of explanations, conversational tone, and scaffolding strategies employed during interaction. To strengthen interaction coherence, the assistant also incorporates elements of lightweight fictional biography. For instance, the system references SpiridonVangheli as the symbolic "father" of the character and Eleonora Popa as the symbolic "mother." These references are not presented as factual statements; rather, they function as narrative anchors that reinforce continuity within the conversational environment. This approach supports one of the broader conceptual goals of the project: transforming the assistant from a generic AI application into a culturally meaningful educational companion. In practice, the system attempts to combine tutoring, reflection, encouragement, and cultural familiarity within a single interaction space. The platform additionally includes a dedicated file-processing module responsible for handling both document-based and image-based input. This component was introduced to make interaction more natural for younger learners, who often work directly with printed textbooks and worksheets. Image processing became especially important during early experimentation. Rewriting mathematical exercises manually proved inconvenient for many pupils, particularly when tasks involved structured notation or longer textual fragments. Allowing students to simply photograph an exercise reduced interaction friction considerably and made the system easier to approach. After image submission, the visual content is transformed into a structured representation suitable for downstream reasoning by the language model. The assistant can then generate explanations, clarification questions, or guided step-by-step support depending on the educational context. Beyond image interaction, the system also supports alternative forms of exercise retrieval (see Fig. 6). Students may refer to activities using page numbers, exercise identifiers, or unit references. To support this behavior, the RAG database indexes educational tasks together with associated metadata, including page number, unit title, and exercise identifiers.

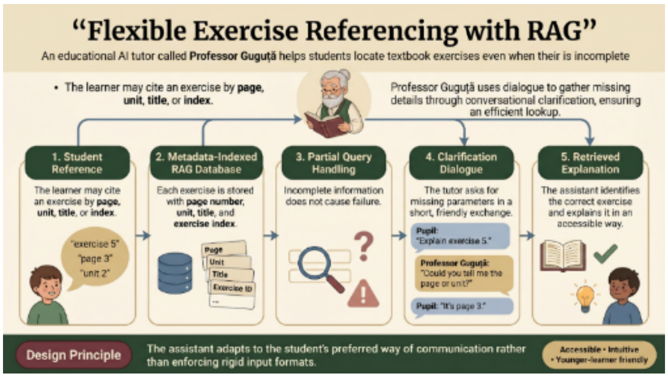

Fig. 6. Retrieval and clarification workflow for textbook exercises

This structured organization allows the assistant to retrieve relevant material even when learner requests remain incomplete. Rather than rejecting ambiguous input, the system attempts to identify missing information through short clarification dialogues.

A typical interaction may proceed as follows:

Pupil: Explain exercise 5.

Professor Guguță: Could you tell me the page or unit number, my dear?

Pupil: It's page 3.

Professor Guguță: Now it is clear. The exercise refers to ...

This interaction pattern reflects one of the broader pedagogical assumptions underlying the project. Instead of forcing pupils to adapt to rigid command structures, the assistant attempts to adapt to the communicative habits of younger learners. In this sense, conversational flexibility becomes part of the educational design itself rather than merely an interface feature.

CONVERSATIONAL ARCHITECTURE AND INTERACTION WORKFLOW

Although conversational AI is widely discussed in educational research, effective educational dialogue does not emerge from a language model automatically. In practice, the quality of interaction depends on how conversational behavior is organized at the architectural level. Routing strategies, intermediate processing steps, retrieval workflows, and pedagogical control mechanisms all influence how the assistant communicates with learners. In AI Guguță, conversational interaction is coordinated through a modular pipeline that combines several language models with specialized educational workflows. Instead of relying on a single model for every operation, the assistant distributes tasks across lightweight and large-scale components depending on the complexity of the request. The main conversational layer is handled by the aya-expanse: 8b model. Embedding generation and retrieval operations rely on embeddinggemma: 300m, while smaller models such as qwen3:1.7b and gemma3:1b perform auxiliary tasks including lightweight classification, validation, and multimodal preprocessing. This organization helps balance computational efficiency with conversational flexibility. During experimentation, smaller models proved sufficient for routine analytical operations, allowing the primary conversational model to focus on reasoning, explanation, and dialogue generation. The interaction workflow itself is divided into three main stages: task identification, preliminary task analysis, and response preparation (see Fig. 7). Together, these stages form the core conversational architecture of the assistant.

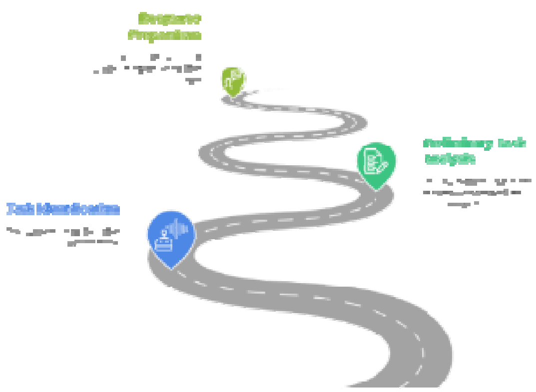

Fig. 7. Conversational workflow of the assistant

Task identification and routing

The first stage focuses on identifying the type of educational request submitted by the learner. To keep the process lightweight, the system combines simple NLP operations with structured validation mechanisms rather than relying entirely on large-scale inference. Initially, the user query undergoes tokenization and stemming through the SnowballStemmer algorithm. The resulting tokens are compared against predefined educational task patterns in order to estimate the most likely interaction category.

For routing purposes, requests are grouped into four broad categories: web search tasks, textbook-related interactions, file-based requests, and a general fallback category for open conversational queries. This lightweight classification strategy reduces unnecessary computation while still allowing the assistant to dynamically select an appropriate processing path.
Additional verification is performed when uploaded files are detected or when the classification stage produces several possible interpretations. In these situations, the assistant activates the BookTask validation component in order to determine whether the request refers to a textbook-related activity. The BookTask mechanism relies on Ollama structured outputs together with Pydantic-based schema validation. Instead of generating unrestricted text, the model produces constrained responses such as *True* or *False*. This approach makes routing decisions more stable and predictable. For these lightweight validation tasks, the system uses the compact gemma3:1b model. In practice, this proved sufficient for classification operations while keeping latency relatively low.

Preliminary task analysis

Before the primary conversational model generates a final response, the assistant performs a set of intermediate processing operations. The exact workflow depends on the educational task identified during the previous stage. At this point, the assistant may perform web search operations, retrieval procedures, file interpretation, or contextual restructuring (see Fig. 8). To support these behaviors, the system relies on a collection of specialized prompts adapted to different educational situations. Textbook-related interaction represents one of the most important educational workflows in the system. To support this functionality, the complete Grade 4 Romanian Language and Literature textbook was transformed into a structured retrieval database. ChromaDB stores embeddings together with metadata describing page numbers, units, and exercise identifiers.

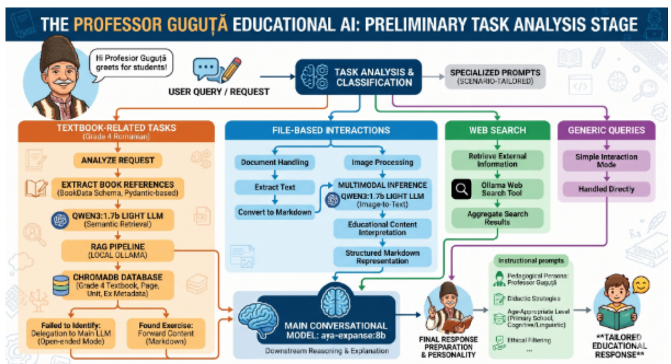

Fig. 8. Intermediate conversational processing and task analysis

When a learner submits a request, the assistant first attempts to identify references to textbook structure, including page numbers, unit titles, or exercise indices. For this purpose, the project introduces a dedicated BookData schema implemented through Pydantic validation. Even when learners provide incomplete information, the assistant still attempts to infer the intended exercise through semantic retrieval operations. A lightweight qwen3:1.7b model is used during this stage in order to reduce response latency while preserving retrieval quality. Once the exercise is identified, the retrieved content is forwarded to the primary conversational model for explanation and dialogue generation. If retrieval confidence remains low, the assistant shifts into a more open conversational mode and continues the interaction directly through the main LLM.

Multimodal file interaction

The platform additionally supports both document-based and image-based educational interaction. This functionality was introduced because younger learners frequently work with printed materials rather than digital exercises. Document processing follows a relatively simple workflow. Extracted textual content is converted into markdown format and then forwarded to the main language model for reasoning and response generation. Image-based interaction requires an additional multimodal processing step. In this scenario, the qwen3:1.7b model performs image interpretation in order to identify the educational content shown in the uploaded image.

After interpretation, the extracted information is transformed into a structured markdown representation before being delivered to the primary conversational model. This intermediate representation improved response stability during experimentation and helped maintain more consistent explanations. The educational motivation behind this feature is practical rather than purely technical. Rewriting textbook exercises manually can be inconvenient for younger pupils, especially when mathematical notation or longer textual fragments are involved. Allowing learners to simply photograph an exercise makes interaction faster and considerably more accessible.

Web interaction and educational synthesis

For broader educational discussions and informational questions, the assistant can retrieve external web content through the Ollama web search tool. Retrieved materials are aggregated and then synthesized by the primary conversational model into a unified educational response. This functionality allows the assistant to move beyond textbook-centered interaction while still preserving conversational continuity and pedagogical tone. Generic conversational requests represent the simplest interaction mode and are processed directly by the primary language model without additional preprocessing stages. Regardless of the original interaction path, all workflows eventually converge at the final response preparation stage.

Pedagogical response preparation

During the final stage, the assistant applies a set of instructional prompts that reinforce the pedagogical identity of Professor Guguță. These prompts regulate conversational tone, explanation strategies, ethical filtering constraints, and age-appropriate communication rules. As a result, the generated responses are adapted not only to the educational content of the task, but also to the linguistic and cognitive characteristics of primary-school learners. In practice, the assistant attempts to maintain a balance between explanation, guidance, encouragement, and conversational accessibility.

Conclusion and Future directions

This study presented AI Guguță, a culturally aware conversational assistant designed for primary education in the Republic of Moldova. The work began from an observable limitation in current educational AI research. Although large language models, RAG, and conversational tutoring systems have developed rapidly in recent years, these approaches rarely combine age-appropriate pedagogy, curriculum grounding, and local cultural adaptation within a single framework for younger learners. AI Guguță was proposed as an attempt to connect these dimensions inside one educational environment. The review of related work showed that most existing systems focus primarily on one aspect of educational interaction at a time. Some approaches emphasize factual grounding through RAG pipelines, while others focus on conversational tutoring strategies, prompting techniques, or multilingual support. Research on culturally responsive educational AI remains comparatively limited, particularly in low-resource linguistic contexts. In this respect, AI Guguță brings together several research directions through a hybrid architecture that combines local language model orchestration, structured retrieval from educational materials, multimodal interaction, and a pedagogically constrained conversational layer shaped by the persona of Professor Guguță. One of the central observations of this work is that cultural grounding should not be reduced to a decorative interface element. In the proposed system, the use of the Guguță persona, Moldovan literary references, and traditional proverbs forms part of the pedagogical design itself. This cultural layer influences the tone of interaction, explanation style, and conversational pacing used during tutoring. As a result, the assistant moves beyond the role of a generic question-answering system and becomes closer to a culturally familiar educational companion for primary-school learners.

The proposed architecture also suggests that effective educational dialogue depends less on a single powerful model and more on how conversational workflows are organized. By distributing tasks across lightweight classification models, embedding models, retrieval components, and a primary conversational model, the system balances computational efficiency with instructional quality. This organization supports multiple forms of educational interaction, including textbook exercise retrieval, clarification dialogue, image-based input, document interpretation, web-assisted educational queries, and open conversational support. Overall, the study argues that conversational AI for primary education should remain simultaneously pedagogical, explainable, culturally situated, and operationally flexible. AI Guguță represents one possible step in this direction by showing how curriculum grounding, Socratic dialogue, multimodal interaction, and cultural identity may be combined within a single educational assistant for young learners. Several limitations remain. The current study primarily focused on architectural design and interaction workflows rather than large-scale classroom evaluation. Future work should therefore include systematic testing with pupils and teachers, measurement of learning outcomes, and comparative evaluation against non-culturally adapted tutoring systems. Additional directions include extending support to other school subjects, improving personalization across different learner profiles, and strengthening resources for low-resource linguistic settings. Such work would help evaluate not only the technical effectiveness of the assistant, but also its longer-term educational and cultural value.

Acknowledgments: This work was carried out within the framework of Project AI Guguță (Code: 25.80012.0807.21TC). The author gratefully acknowledges the support provided by this project for the development of the research presented in this article.

REFERENCES

1. Létourneau, A., Dörfler, T., Pelikan, E., Martin, F., Bieg, M. and Cress, U., "A systematic review of AI-driven intelligent tutoring systems in K-12 education", *npj Science of Learning*, 2025. [Online]. Available: https://doi.org/10.1038/s41539-025-00320-7. [Accessed May 3, 2026].
2. Vanacore, K., Baker, R.S., Closser, A.H. and Roschelle, J., "The path to conversational AI tutors: Integrating tutoring best practices and targeted technologies to produce scalable AI agents", *arXiv preprint*, 2026. [Online]. Available: https://doi.org/10.48550/arXiv.2602.19303. [AccessedMay 3, 2026].
3. Pereira, A.J., Gomes, A.S. and Primo, T.T., "Evaluation of human-artificial hybrid tutoring system for mediation of engagement in e-learning", *Revista Brasileira de Informáticana Educação*, 33. 216-246. 2025. [Online]. Available: https://doi.org/10.5753/rbie.2025.4988. [Accessed May 3, 2026].
4. Li, Z., Wang, Z., Wang, W., Hung, K., Xie, H. and Wang, F.L., "Retrieval-augmented generation for educational application: A systematic survey", *Computers and Education: Artificial Intelligence*, 8. Article 100417. 2025. [Online]. Available: https://doi.org/10.1016/j.caeai.2025.100417. [Accessed May 3, 2026].
5. Swacha, J. and Gracel, M., "Retrieval-Augmented Generation (RAG) chatbots for education: A survey of applications", *Applied Sciences*, 15 (8). Article 4234. 2025. [Online]. Available: https://doi.org/10.3390/app15084234. [AccessedMay 3, 2026].
6. Oh, J., Whang, S.E., Evans, J. and Wang, J., "Classroom AI: Large language models as grade-specific teachers", *npj Artificial Intelligence*, 2. Article 28. 2026. [Online]. Available: https://doi.org/10.1038/s44387-026-00081-7. [AccessedMay 3, 2026].
7. Digital Divide Data, "Low-resource languages in AI: Closing the global gap", 2026. [Online]. Available: https://www.digitaldividedata.com/blog/low-resource-languages-in-ai. [AccessedMay 3, 2026].
8. Mullah, N.S. and Zainon, W.M.N.W., "AI and low-resource languages: Bridging the linguistic divide", in *Reshaping Language and Cognition in Education through AI*, IGI Global Scientific Publishing, 77-128. 2026. [Online]. Available: https://doi.org/10.4018/979-8-3373-4397-6.ch003. [AccessedMay 3, 2026].
9. Center for Low-Resource Languages and Cultures, "Centering low-resource languages and cultures in the age of large language models", in *NeurIPS 2025 Workshop*, 2025. [Online]. Available: https://openreview.net/group?id=NeurIPS.cc/2025/Workshop_Mexico_City/CLRLC-LLMs. [AccessedMay 3, 2026].
10. Huang, J., Saleh, S. and Liu, Y., "LLM agents for education: Advances and applications", in *Findings of the Association for Computational Linguistics: EMNLP 2025*, 2025. [Online]. Available: https://arxiv.org/abs/2503.11733. [AccessedMay 3, 2026].
11. Aji, A.F., Winata, G.I., Koto, F., Cahyawijaya, S. and others, "Centering low-resource languages and cultures in the age of large language models", *arXiv preprint*, 2026. [Online]. Available: https://doi.org/10.48550/arXiv.2601.15337. [AccessedMay 3, 2026].
12. UNESCO, *Education for Sustainable Development: A Roadmap*, UNESCO Publishing, Paris, 2023. [Online]. Available: https://media.unesco.org/sites/default/files/webform/ed3002/396190eng.pdf. [AccessedMay 3, 2026].
13. Skyba, Y.A., "Skills of the 21st century: Adaptation of educational programs of the national higher education institutions", *ProblemyOsvity*, 2 (99). 72-88. 2023. [Online]. Available: https://lib.iitta.gov.ua/id/eprint/741908/. [Accessed May 3, 2026].
14. Pawar, S., Mydlarz, C., Kshirsagar, M. and others, "Survey of cultural awareness in language models: Text and beyond", *arXiv preprint*, 2024. [Online]. Available: https://arxiv.org/abs/2411.00860. [Accessed May 3, 2026].
15. Quan, S., Tu-Shea, X., Ding, Y., Du, Y., Zheng, Q. and Gerdich, L.E., "Conversational AI in children's home literacy learning: Effectiveness, advantages, challenges, and family perception", *Computers and Education: Artificial Intelligence*. Article 100549. 2026. [Online]. Available: https://doi.org/10.1016/j.caeai.2026.100549. [Accessed May 3, 2026].
16. Ding, Y., Hu, H., Zhou, J., Chen, Q., Jiang, B. and He, L., "Boosting large language models with Socratic method for conversational mathematics teaching", in *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, ACM, 3730-3735. 2024. [Online]. Available: https://doi.org/10.1145/3627673.3679881. [AccessedMay 3, 2026].
17. Beale, R., "Dialogic pedagogy for large language models: Aligning conversational AI with proven theories of learning", *arXiv preprint*, 2025. [Online]. Available: https://arxiv.org/abs/2506.19484. [AccessedMay 3, 2026].
18. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.T., Rocktäschel, T., Riedel, S. and Kiela, D., "Retrieval-augmented generation for knowledge-intensive NLP tasks", *arXiv preprint*, 2020. [Online]. Available: https://arxiv.org/abs/2005.11401. [AccessedMay 3, 2026].
19. Maity, S., Deroy, A. and Sarkar, S., "Leveraging in-context learning and retrieval-augmented generation for automatic question generation in educational domains", *arXiv preprint*, 2025. [Online]. Available: https://arxiv.org/abs/2501.17397. [AccessedMay 3, 2026].
20. Oh, J., Whang, S.E., Evans, J. and Wang, J., "Classroom AI: Large language models as grade-specific teachers", *npj Artificial Intelligence*, 2. Article 28. 2026. [Online]. Available: https://doi.org/10.1038/s44387-026-00081-7. [AccessedMay 3, 2026].
21. Medina, J., Gyo-Jung, S., Salminen, J., Aldous, K.K. and Jansen, B.J., "EduX-RAG: Retrieval augmented generation framework for cross-lingual educational chatbots", in *2025 3rd International Conference on Foundation and Large Language Models (FLLM)*, IEEE, 465-473. 2025. [Online]. Available: https://doi.org/10.1109/FLLM67465.2025.11390969. [AccessedMay 3, 2026].
22. Wang, J., Xiao, R., Hou, X., Li, H., Tseng, Y.J., Stamper, J. and Koedinger, K., "LLMs to support K-12 teachers in culturally relevant pedagogy: An AI literacy example", *arXiv preprint*, 2025. [Online]. Available: https://arxiv.org/abs/2505.08083. [Accessed May 3, 2026].
23. Henkel, O., Horne-Robinson, H., Kozhakhmetova, N. and Lee, A., "Effective and scalable math support: Evidence on the impact of an AI-tutor on math achievement in Ghana", *arXiv preprint*, 2024. [Online]. Available: https://arxiv.org/abs/2402.09809. [AccessedMay 3, 2026].
24. Gao, X., Zhang, H. and Chen, N.F., "MultiAiTutor: Child-friendly educational multilingual speech generation tutor with LLMs", *arXiv preprint*, 2025. [Online]. Available: https://arxiv.org/abs/2508.08715. [AccessedMay 3, 2026].

****