

Adaptive Neuro-Symbolic Learning Companions for Explainable Cognitive Scaffolding in Educational AI

Patrick O. Akinwumi

pakinwu@clemson.edu

Clemson University

Meihua Qian

Clemson University

Oyinkansola A. Babatope

Bournemouth University

Immaculate Asimole

University of Ottawa

Research Article

Keywords: Neuro-symbolic artificial intelligence, Adaptive learning companions, Intelligent tutoring systems, Cognitive scaffolding, Explainable AI in education, Zone of Proximal Development (ZPD), Educational conversational agents, Human-centered artificial intelligence (HCAI)

Posted Date: June 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9932715/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Adaptive Neuro-Symbolic Learning Companions for Explainable Cognitive Scaffolding in Educational AI

Patrick O. Akinwumi¹, Meihua Qian¹, Oyinkansola A. Babatope², Immaculate Asimole³

¹ Clemson University

² Bournemouth University

³ University of Ottawa

Author Note

Patrick O. Akinwumi, College of Education, Clemson University, Clemson, United States, ORCID: 0009-0003-9777-7755

Meihua Qian, College of Education, Clemson University, Clemson, United States, ORCID: 0000-0002-9383-7939

Oyinkansola A. Babatope, Business School, Bournemouth University, England, United Kingdom, ORCID: 0009-0006-1670-942X

Immaculate Asimole, Department of Law, University of Ottawa, Canada

Correspondence for this paper should be addressed to Patrick Akinwumi at College of Education, Clemson University, Clemson, SC 29634, USA. Contact: pakinwu@clemson.edu

Abstract

The increasing adoption of artificial intelligence (AI) in education has accelerated the development of intelligent tutoring systems capable of providing adaptive instructional support. However, many existing systems remain constrained by limited personalization, weak reasoning transparency, and insufficient integration of cognitive-developmental learning principles. Most approaches rely on either neural conversational intelligence or symbolic pedagogical reasoning independently, restricting their ability to provide explainable and developmentally aligned educational support. This study proposes and empirically evaluates an adaptive neuro-symbolic learning companion framework designed to support explainable cognitive scaffolding through integrated learner modeling and pedagogical reasoning.

Using the Khan Academy Tutoring Accuracy Dataset, a neuro-symbolic analytical framework was developed that combines transformer-based conversational analysis, symbolic tutoring inference, emotional state detection, scaffolding evaluation, and explainability assessment. The framework operationalizes scaffolding strategies grounded in Vygotsky's Zone of Proximal Development (ZPD), including adaptive hint escalation, conceptual reinforcement, reflective prompting, and meta-cognitive support. Statistical analyses, including independent samples t-tests and ANOVA, were conducted to examine relationships between scaffolding strategies and learning effectiveness.

Results indicate that existing tutoring interactions are predominantly procedural and direct-answer oriented, with only six explicit meta-cognitive prompts identified across 2,022 tutoring interactions. Although conversational coherence remained high (BERTScore F1 = 0.815), reflective scaffolding was largely absent. The proposed framework successfully identified interpretable pedagogical behaviors and demonstrated measurable explainability through symbolic reasoning structures. The study contributes a novel neuro-symbolic approach to adaptive learner

modeling, explainable educational AI, and cognitively aligned tutoring systems, advancing the development of transparent and personalized AI learning companions.

Keywords: Neuro-symbolic artificial intelligence; Adaptive learning companions; Intelligent tutoring systems; Cognitive scaffolding; Explainable AI in education; Zone of Proximal Development (ZPD); Educational conversational agents; Human-centered artificial intelligence (HCAI)

1. Introduction

Artificial intelligence (AI) has advanced rapidly, enabling AI agents to engage in human-like dialogue and problem solving, fueling interest in their potential as learning companions or intelligent tutors. These companions aim not only to deliver information but also to scaffold human cognitive development, mirroring the role of skilled human tutors. This vision builds on Vygotsky’s concept of the *zone of proximal development* (ZPD), where learners progress most effectively with guidance from a more knowledgeable other who provides structured support that is gradually withdrawn as competence grows (Sætra, 2025). Traditionally, such scaffolding has been offered by teachers or peers, but emerging AI systems seek to assume this role.

Recent large language models (LLMs) demonstrate strong capabilities in understanding and generating human-like text. However, purely neural (connectionist) architectures face limitations in reasoning and explaining their decisions. This has spurred interest in *neuro-symbolic* AI, which integrates neural learning with symbolic knowledge representation and reasoning (Jaldi et al., 2025). By uniting statistical learning with explicit reasoning, neuro-symbolic learning companions hold the promise of being both adaptive, learning from data and experience, and explainable, aligning more closely with human cognitive processes (Hu et al., 2025; Jaldi et al., 2025).

This paper examines AI learning companions as tools for supporting cognitive development. We outline the theoretical foundations underlying adaptive neuro-symbolic learning companions, review key contributions from prior work, including practical implementations such as intelligent tutoring systems and adaptive learning platforms, and analyze their limitations. Building on these insights, we identify research gaps and describe how our proposed adaptive neuro-symbolic learning companion addresses them with novel contributions. The paper is

organized as follows: (1) theoretical foundations of scaffolding and cognitive development in educational contexts; (2) advances in AI-driven learning companions, with emphasis on neuro-symbolic frameworks and real-world applications; (3) limitations of current approaches and open research questions; and (4) our proposed approach for overcoming these challenges.

2. Literature Review

2.1 Theoretical Foundations of Cognitive Development and Scaffolding

Contemporary AI learning companions draw heavily on theories of human cognitive development, particularly *Vygotsky's socio-cultural theory*, which frames learning as a fundamentally social process. Cognitive development, according to Vygotsky, occurs through interaction with more knowledgeable others (MKOs) within the learner's *zone of proximal development* (ZPD) (Sætra, 2025). The ZPD is the gap between what a learner can achieve independently and what they can accomplish with skilled guidance (Sætra, 2025; Vygotsky & Cole, 1978).

Learning in the ZPD is facilitated by *scaffolding*, in which the MKO provides structured support, such as hints, prompts, feedback, or task decomposition, enabling the learner to perform just beyond their current ability. This support is gradually withdrawn as competence increases. Traditionally, the MKO has been a human teacher, parent, or peer. However, Vygotsky noted that guidance need not be human or physically present; mediation can take diverse forms (Daniels, 2016; Lindblom & Ziemke, 2003). This opens the possibility for AI systems to serve as surrogate MKOs.

If sufficiently advanced, an AI tutor could operate within the learner's ZPD by delivering timely, adaptive hints, explanations, and feedback (Sætra, 2025). Recent work by Sætra (2025) examines this potential, showing how specialized AI can support learning as a tutor in the ZPD,

even for highly skilled “champion” learners who lack human mentors beyond their level. By applying Vygotskian principles, AI can function as a “more competent other” that accelerates and enriches intellectual development.

2.2 Scaffolding Functions

To design AI that genuinely scaffolds learning, it is essential to understand the functions of effective scaffolding. Wood et al. (1976) identified six core functions: *recruitment* (engaging the learner’s interest), *direction maintenance* (sustaining focus and motivation), *frustration control* (managing stress and discouragement), *task simplification* (reducing complexity or degrees of freedom), *marking critical features* (highlighting key aspects or errors), and *demonstration* (modeling solutions). These describe how skilled tutors support learners.

Importantly, such functions can be replicated by AI tutors, even if the AI lacks genuine human understanding (Sætra, 2025). An AI can foster learning by executing effective tutoring strategies, such as offering hints, encouragement, or worked examples, without “knowing” the subject in the human sense (Fulbright, 2020). This underscores that what matters for scaffolding is the quality and appropriateness of observable support behaviors, not human-like consciousness. To achieve this effectively, however, the AI must maintain accurate models of the learner’s current knowledge and emotional state, enabling it to determine *when* and *how* to scaffold.

2.3 Adaptive Learning and Zone of Proximal Development

In human tutoring, effective scaffolding requires continuous adaptation to the learner’s progress, a hallmark of one-on-one instruction. The ZPD framework implies that learning experiences must be personalized: if tasks are too easy, they fail to challenge; if too difficult, they risk frustration. Adaptive learning systems aim to operationalize this principle by dynamically adjusting difficulty and support.

Early examples include cognitive tutors based on ACT-R models and other intelligent tutoring systems (ITS), which selected problems or offered hints according to estimates of a student's skill mastery. These systems were typically rule-based or model-based, relying on symbolic representations of skills and misconceptions. Research showed that such individualized support produced greater learning gains than uniform instruction. However, these approaches required extensive expert-authored rules and were ill-suited for open-ended tasks such as natural language dialogue.

Modern AI, particularly approaches that integrate neural and symbolic methods, seeks to overcome these limitations, enabling more flexible, scalable, and context-aware adaptation within the learner's ZPD.

2.4 Advances in AI Learning Companions: Neuro-Symbolic Approaches

Recent years have witnessed major advances in AI-powered learning companions, ranging from sophisticated intelligent tutoring systems to large-scale adaptive learning platforms. One emerging trend is the adoption of neuro-symbolic AI in educational technologies—hybrid approaches that combine neural networks, which are effective for pattern recognition and natural language understanding, with symbolic reasoning systems that support knowledge representation, logical inference, and interpretability (Jaldi et al., 2024). This integration aims to capitalize on the complementary strengths of both paradigms: neural models, including deep learning systems and large language models (LLMs), are highly effective at processing complex, unstructured data such as student language input and learning from examples, whereas symbolic models are better suited for representing explicit domain knowledge, curriculum structures, and student knowledge graphs, as well as enabling logical reasoning and planning (Huynh et al., 2023; Jaldi et al., 2024).

A prominent example of neuro-symbolic AI in education is the Never-Ending Open Learning Adaptive Framework (NEOLAF), introduced by Tong et al. (2024). NEOLAF is a neuro-symbolic cognitive architecture that combines teachable, self-learning AI agents powered by large language models (LLMs) with symbolic reasoning components. These agents operate within a Self-Improving Adaptive Instructional System (SIAIS), where they collaborate to provide personalized support for both learners and educators. In this framework, different AI agents may assume specialized instructional roles, for example, one agent may identify and diagnose student misconceptions, while another provides hints, scaffolding, or motivational feedback, all coordinated through a symbolic layer that manages the broader instructional process (Tong & Lee, 2023; Xu et al., 2024).

The architecture emphasizes several important capabilities, including explainability, continual and incremental learning, collaborative problem solving, human-in-the-loop adaptability, and long-term self-improvement (Tong et al., 2023). Because the symbolic layer can represent reasoning explicitly, the system is able to provide interpretable, human-readable explanations for its decisions and recommendations. At the same time, the neural components enable the agents to learn from experience and adapt to unfamiliar problems. Early experiments involving NEOLAF-based agents solving mathematical tasks demonstrated stronger learning performance than either purely neural or purely symbolic systems alone. These findings suggest that neuro-symbolic learning companions can achieve a balance between adaptability, transparency, and scalability that earlier rule-based or connectionist tutoring systems struggled to provide.

A parallel development can be seen in industry-driven adaptive learning systems such as Squirrel AI, an AI-powered tutoring platform widely used across China. Squirrel AI delivers

personalized instruction in subjects such as mathematics and English and has served millions of learners through its network of learning centers. Its recent advancement, the Large Adaptive Model (LAM), integrates large-scale neural foundation models with retrieval-augmented generation (RAG), educational knowledge graphs, and intelligent agent frameworks (Xu et al., 2024). This hybrid architecture enables the system to model complex relationships among concepts, learning objectives, and student competencies in order to personalize instruction at a granular level.

For instance, when a learner struggles with a topic such as factoring quadratic equations, the system can use its symbolic knowledge graph to identify prerequisite concepts, such as integer factors or foundational algebraic operations, while the neural components dynamically generate targeted explanations, hints, or practice exercises to address those gaps. By combining neural adaptability with structured symbolic representations of knowledge, the platform can continuously refine individualized learning pathways in a manner similar to human tutoring. The large-scale implementation of Squirrel AI demonstrates the practical effectiveness of adaptive AI systems that adjust instructional content, feedback, and difficulty in real time according to a learner model, thereby supporting instruction within the student's Zone of Proximal Development (ZPD) (Xu et al., 2024).

In the field of conversational intelligent tutoring systems, the emergence of large language models (LLMs) has revitalized earlier projects such as AutoTutor, an ITS designed to support natural language dialogue with learners (Hu et al., 2025). Earlier versions of AutoTutor relied on techniques such as latent semantic analysis and predefined dialogue scripts to interpret student responses. While innovative for their time, these approaches were limited in their ability to understand complex or unexpected learner input. The integration of transformer-based LLMs has

significantly expanded these capabilities, enabling next-generation tutoring systems to engage in more sophisticated, context-aware conversations with students.

For example, emerging Socratic tutoring systems built on GPT-based architectures can conduct extended educational dialogues by asking probing questions, providing scaffolded hints, and dynamically adapting pedagogical strategies during instruction. These AI-driven learning companions benefit from the strengths of neural language models, particularly their fluency in natural language generation and broad contextual understanding. However, researchers have also recognized the limitations of purely neural systems, including the tendency of LLMs to hallucinate incorrect information or generate inconsistent instructional guidance. To address these concerns, many researchers are integrating symbolic reasoning mechanisms and pedagogical constraints into conversational tutoring architectures.

One common approach involves coupling LLMs with *educational knowledge graphs* that represent subject matter concepts, prerequisite relationships, and the learner's current progress (Brusilovsky, 1994; Hitzler, 2021; Hu et al., 2024; Skjæveland et al., 2024). In this framework, the LLM consults the symbolic knowledge structure to ensure that generated responses remain factually accurate, pedagogically appropriate, and aligned with curricular objectives. Another strategy employs multi-agent neuro-symbolic architecture, in which an LLM-based tutor works alongside additional AI agents responsible for reasoning verification, logical consistency checks, or pedagogical monitoring. This design reflects the broader concept of AI systems that combine rapid neural intuition with slower, deliberate symbolic reasoning.

These developments demonstrate that neuro-symbolic learning companions have become a major interdisciplinary research area that integrates advances in artificial intelligence, cognitive science, and educational technology. Contemporary systems build upon decades of research in

intelligent tutoring systems and adaptive hypermedia while leveraging modern AI capabilities to achieve deeper levels of personalization, contextual understanding, and instructional adaptability. Importantly, these systems are designed not merely to answer questions or solve problems, but also to model the *learner's cognitive state*, knowledge progression, and developmental trajectory to continuously adjust instructional strategies to individual learning needs.

Despite substantial advances in intelligent tutoring systems, large language model-based educational agents, and adaptive learning technologies, contemporary AI learning companions remain limited in their ability to provide genuinely developmentally aligned, cognitively adaptive, and explainable instructional support (Jaldi et al., 2024; Nesbit et al., 2014; Prather et al., 2023). Existing systems frequently personalize learning at a superficial performance level, such as adjusting task difficulty or sequencing content based on correctness, without modeling the deeper conceptual misconceptions, reasoning strategies, or evolving cognitive states of individual learners. Consequently, many tutoring systems continue to employ coarse-grained adaptation mechanisms that fail to distinguish between learners with fundamentally different misunderstandings despite similar observable errors. At the same time, purely neural architectures, particularly large language model-based tutors, suffer from limited reasoning transparency, susceptibility to hallucination, and weak pedagogical explainability, as their instructional outputs are generated through statistical pattern prediction rather than interpretable cognitive reasoning (Wan et al., 2024; Zhang & Sheng, 2024). Although symbolic tutoring systems offer greater interpretability, they often lack the flexibility, scalability, and contextual adaptability required for authentic educational dialogue. Current neuro-symbolic frameworks such as NEOLAF represent important progress toward hybrid educational intelligence; however, the integration of neural semantic understanding with symbolic pedagogical reasoning remains methodologically

fragmented and insufficiently grounded in cognitive-developmental theory. Furthermore, many existing systems inadequately implement the broader dimensions of educational scaffolding identified in learning sciences research, particularly motivational regulation, affect-sensitive support, reflective prompting, and lifelong adaptive learning. As a result, contemporary AI tutors often optimize procedural correctness rather than deeper conceptual understanding, reflective reasoning, or self-regulated learning, thereby reinforcing concerns regarding shallow instructional dependency and limited pedagogical alignment (Tsakeni et al., 2025; Wei et al., 2024).

The present research addresses these limitations through the development of an adaptive neuro-symbolic learning companion framework explicitly grounded in cognitive-developmental theory, educational scaffolding principles, and explainable AI reasoning. Unlike conventional tutoring architectures that rely exclusively on neural language generation or rigid symbolic rule systems, the proposed framework integrates transformer-based conversational intelligence with symbolic learner modeling, dynamic knowledge representation, and interpretable pedagogical reasoning. Central to the proposed architecture is the use of structured learner knowledge graphs capable of modeling conceptual mastery, misconceptions, cognitive progression, and Zone of Proximal Development transitions in real time. This integration enables the system not merely to respond to learner inputs, but to reason about why misconceptions occur and adapt scaffolding strategies accordingly. Moreover, the framework operationalizes theoretically grounded scaffolding mechanisms, including reflective prompting, conceptual reinforcement, motivational support, and adaptive hint escalation, thereby extending beyond transactional answer delivery toward cognitively meaningful developmental support. This direction is further reinforced by the findings of the present study, which revealed the overwhelming dominance of direct-answer scaffolding and the near absence of reflective meta-cognitive prompting within existing tutoring

interactions. The study additionally contributes a self-improving tutoring paradigm in which pedagogical strategies evolve continuously through interaction-based reinforcement and meta-learning mechanisms, enabling the system to refine instructional effectiveness across diverse learner profiles. Equally important, the framework incorporates explainable human-AI collaboration mechanisms that allow educators to inspect, guide, and intervene within the AI's reasoning processes, thereby addressing longstanding concerns regarding transparency, trust, and pedagogical control in educational AI systems. The proposed approach contributes a novel interdisciplinary synthesis of neuro-symbolic reasoning, scaffolding theory, affect-aware tutoring, lifelong adaptation, and explainable educational intelligence, advancing the field toward AI learning companions capable of authentically understanding and scaffolding human cognitive development (Ben Chaabene & Hammami, 2026; Colelough & Regli, 2022).

This study contributes to the emerging field of explainable educational artificial intelligence through five major contributions. First, the study proposes an integrated neuro-symbolic tutoring framework combining transformer-based conversational intelligence with symbolic pedagogical reasoning for adaptive educational interaction. Second, the framework operationalizes cognitively aligned scaffolding strategies grounded in Vygotsky's Zone of Proximal Development (ZPD), including reflective prompting, conceptual reinforcement, adaptive hint escalation, and meta-cognitive instructional support. Third, the study introduces explainability evaluation mechanisms capable of transforming conversational tutoring behaviors into interpretable symbolic instructional representations. Fourth, the framework integrates conversational coherence analysis, emotional detection, and pedagogical scaffolding evaluation within a unified neuro-symbolic educational pipeline. Finally, the study provides empirical evidence demonstrating substantial limitations in current tutoring interactions, particularly the near

absence of reflective meta-cognitive scaffolding, thereby highlighting the need for developmentally aligned AI learning companions capable of supporting deeper cognitive development.

To address these theoretical and methodological limitations, this study proposes an adaptive neuro-symbolic learning companion architecture integrating neural conversational intelligence, symbolic pedagogical reasoning, developmental scaffolding mechanisms, and explainable educational evaluation (see Figure 1).

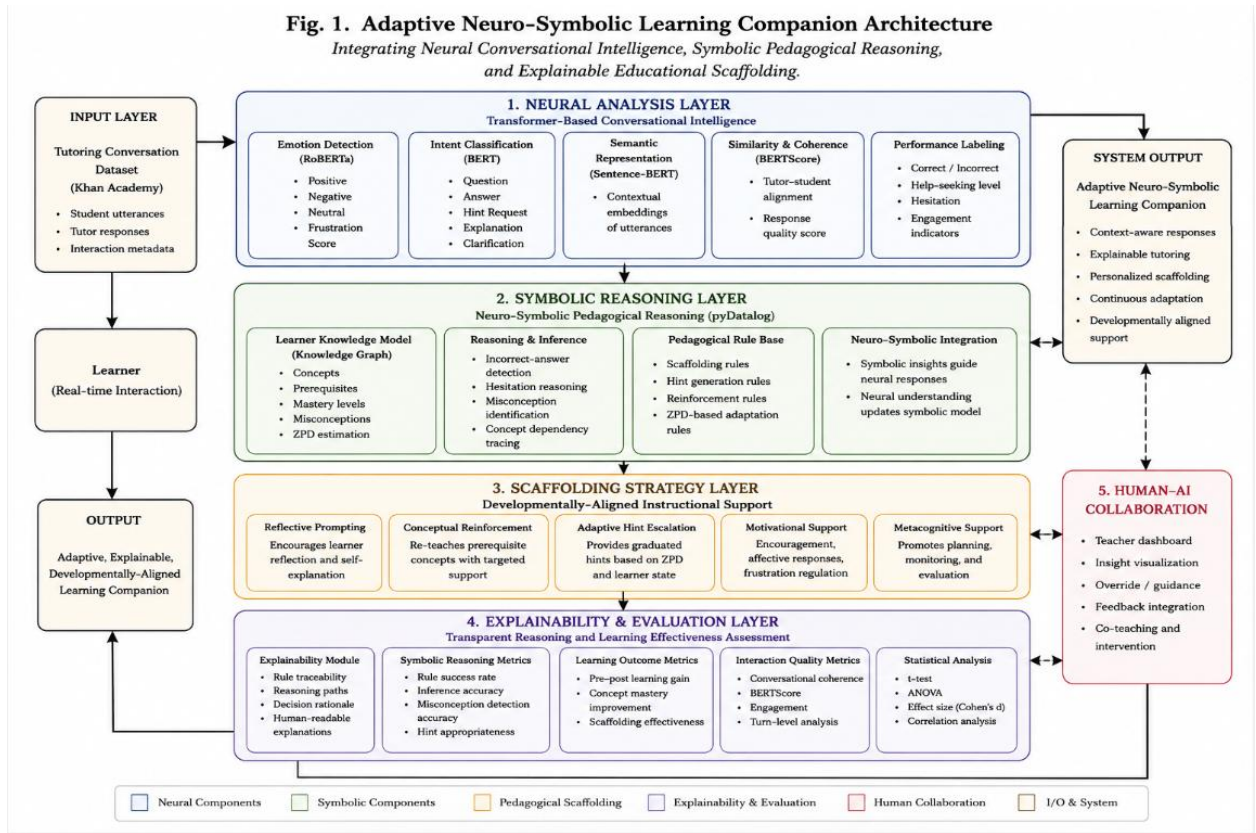


Figure 1 presents the proposed adaptive neuro-symbolic learning companion architecture integrating neural conversational intelligence, symbolic pedagogical reasoning, cognitively aligned scaffolding mechanisms, and explainable educational evaluation. The framework consists of four interconnected layers: (1) transformer-based neural conversational analysis for emotion detection, intent classification, semantic representation, and contextual dialogue understanding;

(2) symbolic pedagogical reasoning for learner modeling, tutoring inference, and interpretable instructional decision-making; (3) developmental scaffolding mechanisms grounded in Vygotskian educational theory, including adaptive hint escalation, conceptual reinforcement, and meta-cognitive prompting; and (4) explainability and evaluation components supporting conversational coherence assessment, symbolic interpretability, and instructional effectiveness analysis. The architecture additionally incorporates human-AI collaboration mechanisms enabling teacher oversight, pedagogical transparency, and adaptive instructional intervention.

3. Methods

This study employed a quantitative computational research design integrating neural language modeling, symbolic reasoning, conversational analysis, and educational scaffolding evaluation to investigate how adaptive neuro-symbolic AI learning companions can understand and scaffold human cognitive development. The methodology combined transformer-based natural language processing techniques with symbolic tutoring logic to evaluate instructional behaviors, emotional dynamics, explainability, conversational quality, and learning effectiveness within authentic tutoring interactions. The overall methodological framework was grounded in neuro-symbolic artificial intelligence, educational scaffolding theory, and Vygotskian cognitive-developmental principles.

The methodological pipeline consisted of six major stages: dataset acquisition and preprocessing, neural conversational analysis, symbolic tutoring reasoning, scaffolding and Zone of Proximal Development (ZPD) evaluation, conversational quality and explainability assessment, and statistical analysis of instructional effectiveness. The analysis was implemented entirely in Google Colab using Python-based machine learning, natural language processing, and statistical computing libraries.

The methodological implementation followed the neuro-symbolic architecture presented in Figure 1. Specifically, the framework consisted of neural conversational analysis, symbolic pedagogical reasoning, scaffolding evaluation, and explainability assessment layers.

3.1 Research Design

The study adopted an exploratory computational research design using secondary conversational tutoring data. The research design was specifically developed to evaluate adaptive tutoring interactions through a neuro-symbolic analytical framework capable of combining neural semantic understanding with symbolic pedagogical reasoning. Unlike purely neural conversational analyses, the proposed methodology integrated interpretable symbolic instructional structures to investigate how tutoring systems scaffold learner cognition, provide adaptive support, and maintain conversational coherence.

The study focused on identifying emotional and cognitive states within tutoring dialogue, tutoring intent structures, symbolic tutoring behaviors, scaffolding strategies, explainability patterns, conversational quality, and learning effectiveness.

The methodological framework was designed to align with theoretical foundations from neuro-symbolic AI, educational psychology, cognitive science, and intelligent tutoring systems research.

3.2 Dataset and Data Source

The study utilized the Khan Academy Tutoring Accuracy Dataset publicly available through GitHub: <https://github.com/Khan/tutoring-accuracy-dataset> (Miller & Dicerbo). The dataset contains authentic tutoring interactions collected from educational conversational exchanges involving tutoring assistance and instructional problem-solving activities. Specifically,

the study analyzed the CoMTA_dataset.json file, which contains structured tutoring conversations consisting of multiple conversational turns between learners and tutoring agents.

The dataset was selected for several reasons. First, it contains authentic educational tutoring interactions rather than synthetic or simulated dialogue. Second, the conversational structure enables analysis of adaptive instructional behaviors, scaffolding strategies, and learner-tutor interaction patterns. Third, the dataset supports both neural conversational analysis and symbolic pedagogical reasoning, making it appropriate for neuro-symbolic educational AI research.

The final expanded conversational dataset contained 2,022 conversational turns following preprocessing and dialogue expansion procedures.

3.3 Computational Environment

All analyses were conducted in Google Colab using Python 3 with GPU acceleration to support transformer-based neural language processing and semantic embedding computations. The computational framework utilized several machine learning, natural language processing, statistical, and visualization libraries, including transformers, datasets, sentence-transformers, evaluate, pyDatalog, bert-score, torch, scipy, scikit-learn, seaborn, matplotlib, pandas, and numpy. These libraries enabled neural conversational analysis, transformer-based emotion detection, semantic classification, symbolic pedagogical reasoning, explainability assessment, conversational coherence evaluation, statistical testing, and visualization generation throughout the study.

3.4 Data Preprocessing

The dataset preprocessing stage involved multiple procedures to transform the raw conversational JSON structures into analyzable tutoring interactions.

First, the repository containing the tutoring dataset was cloned directly within Google Colab using GitHub integration. The CoMTA_dataset.json file was then loaded into a Pandas DataFrame structure.

The tutoring dialogues were originally nested within structured conversational objects. Consequently, the dataset required expansion of nested conversational turns using the `explode()` function to separate each tutoring exchange into individual observations. Conversational metadata including speaker roles and utterance content were extracted from the nested structures and converted into separate columns labeled `role` and `content`.

The preprocessing stage additionally included: lowercasing all text, whitespace normalization, string conversion, and removal of unnecessary formatting artifacts. These procedures ensured consistency across neural language processing and symbolic reasoning stages.

3.4.1 Neural Conversational Analysis

The neural analysis stage utilized transformer-based language models to evaluate emotional states and conversational intentions within tutoring interactions.

3.4.2 Emotion Detection

Emotion detection was implemented using the transformer-based model: `j-hartmann/emotion-english-distilroberta-base`. This model was selected because it supports multi-class emotional classification within conversational text and has demonstrated strong performance in educational and affective language analysis tasks. Each conversational utterance was analyzed independently, and the model classified interactions into one of several emotional categories including: neutral, joy, surprise, sadness, anger, fear, and disgust. The emotional analysis enabled examination of affective dynamics within tutoring interactions and supported evaluation of emotionally supportive scaffolding environments.

3.4.3 Intent Classification

Conversational intent classification was implemented using rule-assisted neural categorization procedures. The analysis classified conversational utterances into four instructional intent categories: question, help request, reflection, and elaboration. Intent classification relied on semantic and syntactic conversational features including interrogative structures, reflective language indicators, explicit help-seeking phrases, and elaborative response patterns. This stage enabled identification of dominant instructional interaction patterns within the tutoring conversations and supported evaluation of learner engagement behaviors.

3.4.4 Neuro-Symbolic Reasoning Framework

A central contribution of the methodology involved the integration of symbolic reasoning mechanisms with neural conversational analysis. Symbolic tutoring logic was implemented using the pyDatalog framework to operationalize interpretable pedagogical tutoring behaviors. The symbolic framework modeled multiple tutoring interaction categories including incorrect-answer detection, hesitation detection, guidance-oriented tutoring, escalated hinting, and conceptual reinforcement behaviors.

3.4.5 Symbolic Rule Construction

Symbolic pedagogical reasoning was implemented through a neuro-symbolic rule construction framework designed to operationalize interpretable tutoring behaviors and adaptive instructional scaffolding. The symbolic component utilized expanded keyword detection and semantic reasoning procedures to identify pedagogically meaningful learner states and tutoring interactions within conversational exchanges. Incorrect-answer patterns included instructional indicators such as “incorrect,” “wrong,” “not correct,” “try again,” and “check your work,” while hesitation-related conversational states were identified through expressions including “maybe,”

“not sure,” “I think,” “perhaps,” and “guess.” Guidance-oriented tutoring behaviors were similarly modeled using scaffolding-oriented instructional phrases such as “step by step,” “remember that,” “think about,” and “consider this.” These symbolic structures enabled the transformation of conversational tutoring interactions into interpretable pedagogical tutoring actions grounded in adaptive instructional reasoning.

To operationalize symbolic pedagogical inference, the study implemented interpretable tutoring rules representing adaptive scaffolding behaviors and learner-state transitions. These symbolic inference structures were expressed as:

$$\textit{Incorrect}(x) \rightarrow \textit{GiveHint}(x)$$

$$\textit{Hesitation}(x) \rightarrow \textit{GuidingQuestion}(x)$$

$$\textit{RepeatedMistake}(x) \rightarrow \textit{EscalatedHint}(x)$$

where x represents a learner interaction state within the tutoring dialogue. These symbolic inference equations model how specific learner cognitive conditions trigger corresponding pedagogical interventions within the neuro-symbolic tutoring framework. Specifically, incorrect responses activate hint-generation mechanisms, hesitation states initiate guidance-oriented questioning strategies, and repeated conceptual failure prompts escalated scaffolding support. The symbolic rules therefore operationalize adaptive tutoring logic through interpretable pedagogical reasoning structures capable of representing instructional decision-making processes transparently.

The integration of these symbolic inference mechanisms constitutes a central component of the neuro-symbolic architecture because it enables the system not only to process conversational language semantically but also to reason pedagogically about learner cognition and instructional

adaptation. Unlike purely neural conversational systems that generate responses through statistical prediction alone, the symbolic reasoning framework provides explicit instructional traceability and pedagogical interpretability. Consequently, the symbolic rules support explainability, adaptive scaffolding, tutoring transparency, and cognitively aligned instructional intervention, thereby reinforcing the study's broader objective of developing neuro-symbolic AI learning companions capable of understanding and scaffolding human cognitive development.

3.4.6 Semantic Symbolic Classification

To improve symbolic reasoning coverage beyond simple keyword matching, the study additionally implemented zero-shot semantic classification using the transformer model: facebook/bart-large-mnli. The semantic reasoning stage classified tutoring utterances into instructional categories including incorrect answer, student hesitation, tutor guidance, and concept reinforcement. A confidence threshold of 0.60 was applied to reduce false classifications and improve symbolic interpretability reliability.

3.5 Symbolic Tutoring Action Assignment

The outputs from keyword-based symbolic detection and semantic classification were integrated into final symbolic tutoring actions: hint, guide, escalated hint, reinforce, and none.

These symbolic tutoring actions represented interpretable pedagogical scaffolding behaviors capable of supporting explainability analysis and instructional evaluation.

3.5.1 Scaffolding and ZPD Analysis

The study evaluated scaffolding behaviors using theoretical foundations from Vygotsky's Zone of Proximal Development (ZPD). Specifically, the analysis examined how tutoring interactions provided adaptive instructional support aligned with learner cognitive progression.

3.5.2 Hint-Type Classification

Hint classification was implemented using transformer-based semantic sentence embeddings generated through the all-MiniLM-L6-v2 model to identify pedagogical scaffolding patterns within tutoring interactions. Four primary scaffolding categories were defined: direct-answer scaffolding, step-by-step guidance, concept reinforcement, and meta-cognitive prompting. Unlike conventional keyword-based classification approaches, the present framework employed contextual semantic similarity analysis to evaluate the conceptual alignment between tutoring utterances and predefined scaffolding representations. This enabled the system to capture semantically meaningful instructional behaviors beyond explicit lexical matching, thereby supporting more adaptive and contextually grounded tutoring analysis.

Semantic similarity between tutoring utterances and predefined scaffolding categories was computed using cosine similarity over transformer-generated sentence embeddings, expressed as:

$$CosSim(A, B) = \frac{A \cdot B}{\|A\| \|B\|}$$

where A represents the embedding vector of a tutoring utterance, B represents the embedding vector of a predefined scaffolding category, $A \cdot B$ denotes the vector dot product, and $\|A\|$ and $\|B\|$ represent the magnitudes of the respective vectors. The cosine similarity function measures semantic similarity, contextual alignment, and conceptual proximity between tutoring responses and instructional scaffolding categories. Higher cosine similarity values indicate stronger semantic correspondence between conversational tutoring behaviors and specific pedagogical scaffolding strategies.

The integration of cosine similarity within the scaffolding classification process strengthens the computational rigor of the neuro-symbolic framework by enabling semantically informed instructional interpretation rather than relying solely on surface-level keyword detection.

This is particularly important within the context of adaptive tutoring systems because pedagogical meaning often emerges contextually rather than through isolated lexical indicators. Consequently, the semantic embedding framework supports the neural conversational intelligence component of the proposed neuro-symbolic architecture by facilitating context-aware instructional classification, adaptive tutoring analysis, and meaningful pedagogical interpretation of learner–tutor interactions.

3.6 Meta-Cognitive Prompt Detection

The study additionally examined reflective instructional support through explicit meta-cognitive prompt detection. Meta-cognitive prompts included reflective instructional phrases such as: “think about,” “why do you think,” “how did you get,” “reflect on,” and “consider why.” This stage evaluated the extent to which tutoring interactions encouraged learner self-monitoring and reflective reasoning.

3.6.1 Wait-Time Approximation

Conversational timing dynamics were approximated using simulated latency intervals ranging from one to five seconds. Although precise timestamp metadata were unavailable within the dataset, the approximation enabled preliminary modeling of conversational pacing and scaffolding timing.

3.7 Learning Effectiveness Evaluation

Learning effectiveness was evaluated through inferred conversational correctness patterns derived from tutoring interactions. Correctness detection utilized instructional confirmation indicators such as “correct,” “exactly,” “well done,” and “that is right” to identify successful learner responses within conversational exchanges. Tutoring interactions were subsequently grouped into instructional sequences using sequential conversation identifiers, enabling comparison between pre-interaction and post-interaction learner performance across tutoring

sessions. This approach allowed the study to estimate conversational learning progression and evaluate the instructional impact of adaptive scaffolding behaviors within authentic tutoring dialogue.

To operationalize instructional effectiveness quantitatively, learning gain was computed as the difference between post-interaction correctness and pre-interaction correctness across tutoring exchanges, expressed as:

$$LG = PostAccuracy - PreAccuracy$$

where LG represents learning gain, $PostAccuracy$ denotes post-interaction learner correctness, and $PreAccuracy$ represents pre-interaction learner correctness. The learning gain equation measures the extent of learner improvement following tutoring intervention and serves as an indicator of scaffolding effectiveness within the neuro-symbolic tutoring framework. Specifically, the metric enables evaluation of whether adaptive instructional strategies contributed to measurable learning progression and whether different pedagogical scaffolding approaches influenced learner performance differently. Consequently, the learning gain formulation supports the broader objective of assessing how adaptive neuro-symbolic tutoring mechanisms scaffold cognitive development and instructional effectiveness within conversational educational environments.

3.8 Explainability and Conversational Quality Evaluation

3.8.1 Explainability Metrics

Explainability was evaluated by measuring the proportion of tutoring interactions that could be successfully represented through symbolic pedagogical reasoning structures within the neuro-symbolic tutoring framework. Specifically, the analysis examined the extent to which conversational tutoring behaviors could be interpreted through explicit symbolic tutoring actions generated from pedagogical inference rules. Unlike purely neural conversational systems that

operate primarily as opaque statistical models, the proposed framework incorporated interpretable symbolic instructional logic capable of tracing pedagogical decision-making processes and adaptive scaffolding behaviors.

The explainability score was computed as:

$$ES = \frac{N_{symbolic}}{N_{total}}$$

where ES represents the Explainability Score, $N_{symbolic}$ denotes the number of tutoring interactions represented through symbolic pedagogical reasoning structures, and N_{total} represents the total number of tutoring interactions analyzed. This metric quantified the degree to which tutoring interactions could be interpreted, traced, and explained through symbolic instructional reasoning mechanisms embedded within the neuro-symbolic framework.

The explainability formulation operationalizes instructional interpretability by evaluating how much of the tutoring process can be transparently represented through explicit pedagogical reasoning structures rather than opaque neural prediction alone. Consequently, the metric supports the study’s broader objective of advancing explainable educational AI systems capable of providing transparent, cognitively aligned, and pedagogically accountable instructional support. The explainability equation therefore mathematically reinforces the central argument of the study that neuro-symbolic tutoring systems offer greater transparency, interpretability, and educational trustworthiness than purely neural conversational tutoring architectures, thereby aligning with broader objectives in explainable AI, educational accountability, and trustworthy adaptive learning systems.

3.9 Conversational Quality Assessment

Conversational coherence and contextual instructional alignment were evaluated using BERTScore semantic similarity metrics. The bert-score framework computed semantic correspondence between tutor and learner utterances through contextual transformer-based embeddings capable of capturing deeper semantic meaning beyond surface lexical similarity. Specifically, BERTScore precision, recall, and F1 similarity metrics were calculated between paired tutoring interactions to evaluate the degree of conversational continuity, semantic coherence, and instructional relevance maintained throughout the tutoring dialogue. The mean BERTScore F1 value served as an overall indicator of conversational quality within the adaptive neuro-symbolic tutoring framework.

Conversational coherence between tutor and learner interactions was operationalized as:

$$BERTScore = F1(P, R)$$

where P represents semantic precision, R denotes semantic recall, and $F1$ represents the contextual semantic similarity score between tutoring utterances. The BERTScore formulation measures semantic alignment, conversational continuity, and contextual instructional coherence by evaluating the similarity between transformer-generated semantic representations of tutor and learner interactions.

The integration of BERTScore within the evaluation framework demonstrates that the proposed neuro-symbolic architecture assesses not only instructional correctness and symbolic reasoning transparency, but also the quality and contextual consistency of educational dialogue. This is particularly important within adaptive learning environments because effective AI learning companions must maintain coherent conversational flow, contextually relevant instructional support, and meaningful interaction continuity throughout tutoring exchanges. Consequently, the

conversational quality metric reinforces the study's broader argument that adaptive neuro-symbolic tutoring systems should simultaneously support explainable pedagogical reasoning and conversational intelligence to facilitate cognitively aligned and developmentally meaningful educational interaction.

3.10 Statistical Analysis

The study employed inferential statistical procedures to evaluate whether scaffolding strategies significantly influenced learning effectiveness.

3.10.1 Independent Samples T-Test

An independent samples t-test was conducted to compare learning gains between step-by-step scaffolding, and direct-answer tutoring strategies. The t-test evaluated whether statistically significant differences existed between instructional approaches.

3.10.2 One-Way ANOVA

A one-way analysis of variance (ANOVA) was additionally conducted to compare learning gains across multiple scaffolding categories including direct-answer scaffolding, step-by-step guidance, and concept reinforcement. The ANOVA enabled evaluation of broader differences in instructional effectiveness across tutoring strategies.

3.10.3 Visualization Procedures

Multiple visualizations were generated to support interpretability of the findings, including symbolic tutoring action distributions, scaffolding hint-type distributions, and learning-gain histograms. The visualizations were implemented using matplotlib and seaborn libraries.

3.11 Ethical Considerations

The study utilized publicly available secondary educational conversational data and did not involve direct human participant recruitment or intervention. No personally identifiable

information was collected, accessed, or analyzed during the study. All analyses complied with ethical standards for secondary computational educational research.

The methodology contributes a comprehensive neuro-symbolic analytical framework integrating transformer-based neural conversational analysis, symbolic pedagogical reasoning, emotional detection, scaffolding evaluation, explainability assessment, conversational coherence analysis, and statistical instructional evaluation within a unified educational AI pipeline. Unlike prior educational AI methodologies that rely exclusively on either neural conversational modeling or standalone symbolic tutoring systems, the present framework operationalizes an integrated neuro-symbolic approach using authentic tutoring interactions to simultaneously examine conversational intelligence, adaptive scaffolding behaviors, pedagogical interpretability, and learning effectiveness. By combining neural semantic understanding with symbolic instructional reasoning grounded in cognitive-developmental theory, the methodology advances the study of adaptive AI learning companions capable of understanding, explaining, and scaffolding human cognitive development in more transparent and pedagogically aligned ways.

4. Results

This section presents the empirical findings from the neuro-symbolic tutoring analysis conducted using the Khan Academy Tutoring Accuracy Dataset. The analysis evaluated conversational tutoring interactions through a combined neural and symbolic framework designed to assess adaptive scaffolding, emotional dynamics, explainability, conversational quality, and learning effectiveness. The findings are presented in relation to the study's core objective of investigating how adaptive neuro-symbolic AI learning companions can understand and scaffold human cognitive development.

The results are organized into six major categories: (1) emotional state detection, (2) intent classification, (3) neuro-symbolic tutoring behaviors, (4) scaffolding and Zone of Proximal Development (ZPD) analysis, (5) explainability and conversational quality evaluation, and (6) statistical analyses of tutoring effectiveness.

4.1 Emotional State Detection in Tutoring Interactions

The neural emotion-detection component was applied to all conversational turns to identify affective patterns during tutoring interactions. The analysis (see table 1) revealed that emotionally neutral conversational states dominated the dataset, accounting for 1,531 interactions. Positive affective states such as surprise ($n = 256$) and joy ($n = 158$) were also observed, while negative emotional states including sadness ($n = 41$), fear ($n = 17$), anger ($n = 16$), and disgust ($n = 3$) occurred relatively infrequently.

These findings suggest that the tutoring interactions generally maintained emotionally stable and supportive learning environments. The prevalence of surprise and joy may indicate moments of conceptual realization, successful scaffolding, and positive reinforcement during instructional exchanges. Conversely, the low occurrence of negative emotional states implies that the tutoring environment minimized emotionally adverse experiences that could disrupt cognitive engagement.

The findings align with prior educational AI research emphasizing the importance of emotionally supportive tutoring environments for sustained engagement and learning effectiveness. The results further support the proposition that adaptive neuro-symbolic tutoring systems can facilitate cognitively productive interactions while preserving emotional stability.

Table 1

Distribution of Detected Emotional States in Tutoring Interactions

Emotion	Frequency
Neutral	1531
Surprise	256
Joy	158
Sadness	41
Fear	17
Anger	16
Disgust	3

4.2 Intent Classification and Conversational Purpose

The neural intent-classification framework identified four dominant conversational intentions (see table 2) within the tutoring interactions: questions, elaborations, help requests, and reflective statements. Question-oriented interactions constituted the majority of utterances ($n = 1,123$), followed by elaborative responses ($n = 867$). Explicit help requests ($n = 17$) and reflective statements ($n = 15$) occurred substantially less frequently.

The dominance of question-oriented interactions indicates that the tutoring sessions were primarily inquiry-driven and focused on immediate problem solving. Similarly, the high number of elaborative responses suggests that learners actively engaged with instructional content rather than limiting responses to short-answer interactions.

However, the comparatively limited presence of reflective and meta-cognitive utterances reveals a critical weakness in existing tutoring interactions. Although learners frequently asked questions, they rarely engaged in deeper self-reflective reasoning or explicit meta-cognitive dialogue. This finding directly supports the study's theoretical argument that current AI tutoring systems prioritize transactional instructional support over developmental cognitive scaffolding.

The scarcity of reflective interactions also highlights a significant opportunity for adaptive neuro-symbolic learning companions to introduce self-monitoring, reflective prompting, and cognitively guided scaffolding mechanisms.

Table 2*Intent Classification Distribution Across Tutoring Interactions*

Intent Type	Frequency
Question	1123
Elaboration	867
Help request	17
Reflection	15

4.3 Neuro-Symbolic Tutoring Behaviors

The symbolic reasoning framework identified multiple categories of tutoring actions across the tutoring conversations (see figure 2), including guidance-oriented scaffolding, hint generation, escalated hints, and reinforcement behaviors. The majority of interactions were classified as non-symbolic instructional exchanges ($n = 1,352$). Among detected symbolic tutoring behaviors, guidance-oriented scaffolding represented the most frequent category ($n = 438$), followed by hint-based interventions ($n = 154$), escalated hints ($n = 52$), and reinforcement strategies ($n = 26$).

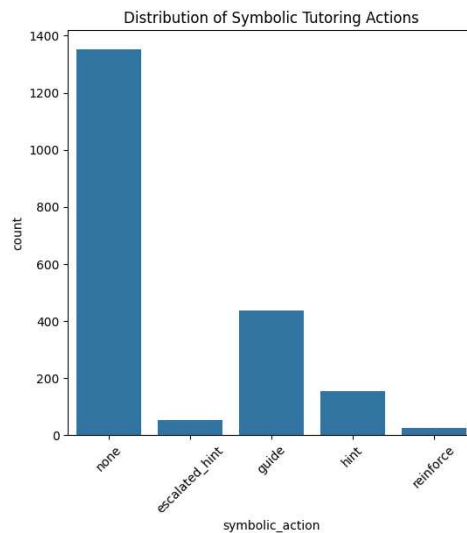
These findings indicate that tutoring interactions frequently relied on guidance-oriented support rather than explicit corrective interventions. The prevalence of guidance behaviors suggests that conversational tutoring strategies attempted to maintain learner engagement through indirect instructional prompting rather than immediate answer delivery.

Importantly, the neuro-symbolic framework successfully operationalized pedagogically meaningful tutoring actions using symbolic reasoning rules. This demonstrates the feasibility of integrating symbolic tutoring logic into conversational AI systems to improve explainability and pedagogical transparency.

The results further indicate that symbolic tutoring actions can serve as interpretable instructional representations capable of explaining how tutoring systems adapt their responses to learner states.

Figure 2

Distribution of Symbolic Tutoring Actions



4.4 Scaffolding and Zone of Proximal Development Analysis

The scaffolding analysis evaluated how tutoring interactions aligned with adaptive instructional support principles associated with Vygotsky's Zone of Proximal Development (ZPD). Four primary scaffolding categories were identified: direct-answer scaffolding, step-by-step guidance, concept reinforcement, and meta-cognitive prompting.

The findings (figure 3) revealed that direct-answer scaffolding overwhelmingly dominated the tutoring interactions ($n = 1,634$). In contrast, step-by-step guidance ($n = 279$), concept reinforcement ($n = 82$), and meta-cognitive prompting ($n = 27$) occurred substantially less frequently.

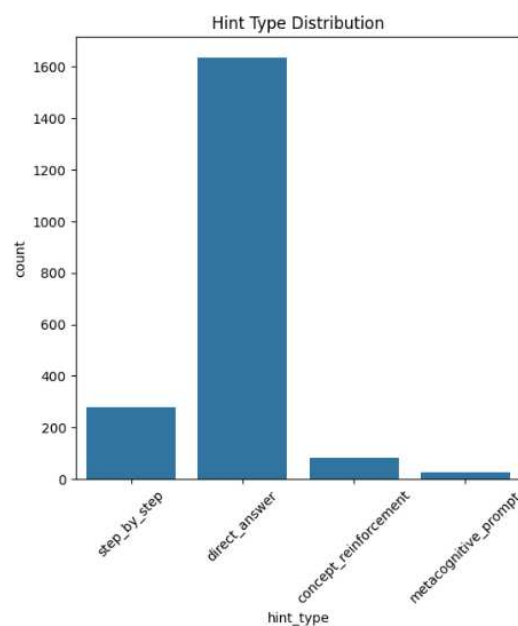
The dominance of direct-answer tutoring suggests that existing tutoring interactions emphasize immediate solution delivery rather than progressive conceptual scaffolding. Although direct answers may support short-term task completion, they may not sufficiently promote deeper cognitive development, reflective reasoning, or independent problem-solving capabilities.

The relatively low occurrence of concept reinforcement and meta-cognitive prompting further highlights the limited presence of developmental scaffolding strategies within current tutoring interactions. These findings strongly support the study’s central argument that existing AI tutoring systems insufficiently support reflective cognitive development.

The results therefore reinforce the need for adaptive neuro-symbolic learning companions capable of integrating meta-cognitive prompting, conceptual reinforcement, and progressive scaffolding strategies.

Figure 3

Distribution of Tutoring Hint Types



4.5 Meta-Cognitive Prompt Detection

The analysis identified only six explicit meta-cognitive prompts across 2,022 tutoring interactions, revealing a substantial absence of reflective instructional scaffolding within existing tutoring dialogue. This finding represents one of the most significant outcomes of the study

because it demonstrates that current tutoring interactions remain overwhelmingly procedural and solution-oriented rather than cognitively developmental or metacognitively supportive.

The extremely limited occurrence of reflective instructional prompting indicates that current tutoring interactions remain primarily transactional and solution-oriented rather than cognitively developmental. Existing tutoring systems appear to prioritize answer generation and procedural support over reflective self-monitoring, conceptual abstraction, and strategic reasoning.

This finding strongly validates the theoretical foundations of the proposed adaptive neuro-symbolic learning companion framework. Specifically, it demonstrates the necessity of integrating explicit meta-cognitive scaffolding mechanisms into AI tutoring systems to support higher-order cognitive development.

The results further suggest that future AI tutoring systems should incorporate reflective prompting strategies capable of encouraging learners to evaluate their reasoning processes, monitor conceptual understanding, and engage in self-regulated learning.

4.6 Learning Gain and Tutoring Effectiveness

Learning effectiveness was evaluated using computed learning gain measures derived from conversational correctness patterns. The analysis (see figure 4 and table 3) produced a mean learning gain score of 0.064, indicating modest but positive instructional improvement across tutoring sessions.

The distribution of learning gains revealed substantial variability among tutoring interactions, with values ranging from -1.0 to 1.0 and a standard deviation of 0.446. This variability suggests that tutoring effectiveness differed considerably across conversations and learners.

The results indicate that while some tutoring interactions successfully facilitated conceptual improvement, others produced limited gains or even negative learning outcomes. This

inconsistency highlights the importance of adaptive personalization mechanisms capable of dynamically adjusting scaffolding strategies according to learner states and cognitive needs.

The findings therefore support the study’s central argument that neuro-symbolic tutoring systems require adaptive reasoning mechanisms to optimize scaffolding effectiveness across diverse learners.

Figure 4

Learning Gain Distribution Across Tutoring Interactions

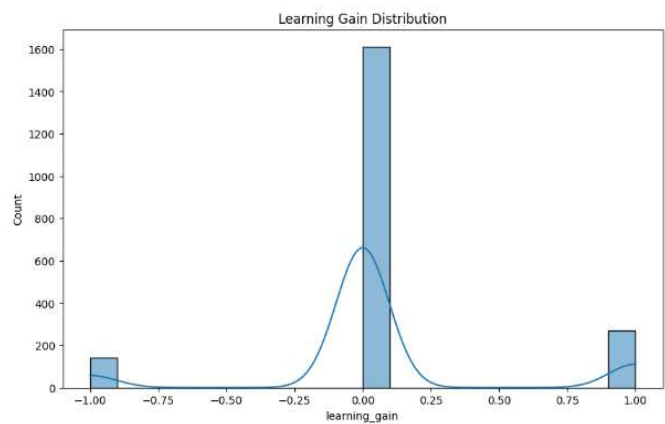


Table 3

Descriptive Statistics for Learning Gain

Statistic	Value
Count	2022.00
Mean	0.064
Standard deviation	0.446
Minimum	-1.000
25 th percentile	0.00
Median	0.000
75 th percentile	0.000
Maximum	1.000

4.7 Explainability and Symbolic Reasoning Performance

The explainability analysis (see table 4) evaluated the extent to which tutoring interactions could be interpreted through symbolic instructional reasoning. The neuro-symbolic framework

achieved an explainability score of 0.331, indicating that approximately one-third of tutoring interactions were successfully represented through interpretable symbolic tutoring actions.

Similarly, symbolic reasoning success achieved a score of 0.331, demonstrating that the symbolic tutoring framework successfully captured a meaningful proportion of instructional behaviors.

These findings are particularly significant because explainability remains one of the primary limitations of purely neural conversational AI systems. By integrating symbolic tutoring logic, the proposed framework demonstrated the capacity to transform conversational tutoring behaviors into interpretable instructional representations.

The results therefore support the broader argument that neuro-symbolic architectures can improve transparency, trustworthiness, and pedagogical interpretability in educational AI systems.

Table 4

Explainability and Symbolic Reasoning Metrics

Metric	Score
Explainability score	0.331
Symbolic Reading Success	0.331

4.8 Scaffolding Efficiency Metrics

The scaffolding efficiency analysis revealed an average conversation length of approximately 10 conversational turns per tutoring interaction (see table 5). This finding suggests that the tutoring sessions maintained sustained instructional engagement rather than shallow or single-turn exchanges. Additionally, repeated mistakes occurred relatively infrequently (n = 51), indicating that tutoring interventions often succeeded in preventing prolonged misunderstanding or repeated conceptual failure. These findings suggest that guidance-oriented scaffolding strategies may have contributed to maintaining instructional progression and conversational continuity.

Table 5*Scaffolding Efficiency Metrics*

Metric	Value
Average conversational turns	9.96
Repeated mistakes	51

4.9 Conversational Quality Evaluation

Conversational quality was evaluated using BERTScore semantic similarity metrics. The analysis produced a mean BERTScore F1 value of 0.815, indicating strong semantic alignment and contextual coherence between tutor and student conversational exchanges.

The high BERTScore value suggests that tutoring interactions maintained meaningful conversational continuity and contextual relevance throughout the instructional exchanges. This finding supports the proposition that adaptive conversational tutoring systems can sustain coherent instructional dialogue while dynamically responding to learner inputs.

Importantly, the conversational coherence findings complement the symbolic explainability results by demonstrating that the neuro-symbolic tutoring framework simultaneously preserved conversational quality and interpretability.

4.10 Statistical Analysis of Tutoring Effectiveness

Statistical testing was conducted to evaluate whether different scaffolding strategies produced significantly different learning outcomes. The independent samples t-test comparing step-by-step scaffolding and direct-answer tutoring revealed statistically significant differences in learning gains ($t = 4.094$, $p < .001$) (see table 6). This finding indicates that tutoring strategy selection significantly influenced instructional effectiveness.

Similarly, one-way ANOVA analysis demonstrated significant differences among multiple hint types ($F = 8.919$, $p < .001$) (see table 7), further confirming that scaffolding approaches

produced distinct learning outcomes. These findings provide strong empirical support for the study's central hypothesis that adaptive scaffolding strategies significantly affect learning effectiveness. The results further validate the importance of integrating adaptive neuro-symbolic reasoning mechanisms capable of dynamically selecting pedagogically appropriate tutoring strategies.

Table 6

Independent Samples T-Test Results

Metric	Value
T Statistic	4.094
P-value	0.000044

Table 7

One-Way ANOVA Results for Hint Types

Independent Samples T-Test Results

Metric	Value
F Statistic	8.919
P-value	0.000139

The findings demonstrate that neuro-symbolic tutoring analysis can successfully identify emotionally relevant interactions, symbolic tutoring behaviors, scaffolding strategies, and conversational quality patterns within educational dialogues.

Most importantly, the results reveal that current tutoring interactions remain heavily dominated by direct-answer instructional strategies while exhibiting extremely limited meta-cognitive scaffolding behaviors. This finding strongly supports the need for adaptive neuro-symbolic learning companions capable of providing reflective, explainable, and developmentally aligned scaffolding mechanisms.

The statistical analyses further confirm that scaffolding strategies significantly influence learning outcomes, thereby validating the theoretical foundations of adaptive educational AI systems grounded in Vygotskian scaffolding principles and neuro-symbolic reasoning architectures.

5. Discussion

Existing intelligent tutoring system (ITS) research primarily focuses on either neural conversational intelligence or symbolic tutoring architectures independently, resulting in limited integration between conversational adaptivity, pedagogical reasoning, explainability, and cognitive-developmental scaffolding. In response, the present study proposes an integrated neuro-symbolic educational AI framework that combines transformer-based conversational intelligence, symbolic pedagogical reasoning, explainability metrics, Zone of Proximal Development (ZPD)-guided scaffolding, emotional analysis, and conversational coherence evaluation within authentic tutoring interactions. By unifying neural semantic understanding with interpretable instructional reasoning, the study advances the development of adaptive AI learning companions capable of supporting human cognitive development through transparent and developmentally aligned educational interaction.

The findings of this study provide important empirical and theoretical insights into the emerging field of adaptive neuro-symbolic educational AI. By integrating transformer-based neural analysis with symbolic reasoning mechanisms, the study demonstrates how neuro-symbolic learning companions can contribute to understanding and scaffolding human cognitive development within authentic tutoring environments. More importantly, the results reveal critical limitations in current AI tutoring interactions while simultaneously highlighting the potential of neuro-symbolic architectures to support more explainable, adaptive, and cognitively aligned

educational systems. The findings reinforce the argument that future educational AI systems must evolve beyond procedural answer generation toward architectures capable of supporting reflective reasoning, adaptive scaffolding, and developmentally meaningful learning processes. Unlike prior ITS research that examines conversational intelligence and symbolic tutoring independently, the present study demonstrates how integrated neuro-symbolic educational architectures can simultaneously support conversational coherence, pedagogical explainability, adaptive scaffolding, and cognitively aligned instructional reasoning within authentic tutoring environments.

One of the most significant findings of the study concerns the successful implementation of interpretable symbolic tutoring behaviors within conversational tutoring interactions. The neuro-symbolic framework identified multiple pedagogically meaningful instructional actions, including guidance-oriented scaffolding, hints, reinforcement behaviors, and escalated tutoring interventions. The explainability score of 0.331 indicates that approximately one-third of tutoring interactions could be represented through symbolic instructional reasoning structures. This finding is particularly important because explainability remains one of the most persistent limitations of contemporary neural conversational AI systems. Large language models frequently demonstrate impressive linguistic fluency while simultaneously lacking transparent reasoning mechanisms capable of explaining instructional decisions. Existing neuro-symbolic literature has repeatedly emphasized that explainability and trustworthiness remain underdeveloped dimensions within neuro-symbolic AI research (Colelough & Regli, 2025). The present findings therefore contribute important empirical evidence demonstrating how symbolic tutoring logic can improve interpretability within educational AI systems.

Unlike purely neural tutoring architectures that rely exclusively on statistical text generation, the proposed framework transformed conversational tutoring interactions into interpretable pedagogical representations grounded in educational scaffolding theory. This contribution aligns with broader neuro-symbolic AI research arguing that hybrid architectures are necessary to bridge neural learning with symbolic reasoning capabilities (Wan et al., 2024). Contemporary neuro-symbolic research increasingly conceptualizes hybrid architectures as computational analogies to dual-process cognitive systems in which neural networks support intuitive pattern recognition while symbolic structures facilitate deliberate reasoning and explainable inference. The present study operationalized this theoretical perspective within educational dialogue by integrating transformer-based conversational understanding with symbolic pedagogical reasoning rules. Importantly, this extends existing neuro-symbolic research by demonstrating a practical educational implementation using authentic tutoring conversations rather than simulated instructional environments.

The findings further reveal substantial limitations in the scaffolding strategies currently dominant within AI tutoring interactions. The tutoring conversations were overwhelmingly characterized by direct-answer instructional approaches, whereas step-by-step guidance, conceptual reinforcement, and meta-cognitive prompting occurred comparatively infrequently. This imbalance is highly significant because it suggests that many current tutoring systems remain primarily solution-oriented rather than developmentally oriented. Vygotsky's theory of the Zone of Proximal Development emphasizes that effective learning occurs when instructional support is adaptively calibrated to guide learners progressively beyond their independent capabilities. Effective scaffolding therefore requires graduated conceptual guidance, reflective prompting, and strategic support rather than immediate answer delivery. The dominance of direct-answer tutoring

observed in the present study suggests that many existing AI tutoring interactions prioritize procedural completion over deeper cognitive development.

This finding strongly supports longstanding concerns within educational technology research regarding the tendency of AI tutoring systems to emphasize performance optimization rather than conceptual understanding. Although direct-answer support may facilitate short-term task completion, it may not sufficiently promote higher-order reasoning, self-regulated learning, or conceptual transfer. The statistical analyses reinforce this interpretation. Both the independent samples t-test and ANOVA demonstrated statistically significant differences between scaffolding strategies and learning outcomes, indicating that tutoring strategy selection substantially influenced instructional effectiveness. Specifically, step-by-step instructional approaches produced significantly different learning gains when compared with direct-answer tutoring behaviors. These findings provide empirical support for educational scaffolding theory by demonstrating that adaptive tutoring strategies influence learning outcomes in measurable ways.

Importantly, the study extends existing neuro-symbolic educational AI research by quantitatively linking symbolic tutoring strategies with educational effectiveness. Existing neuro-symbolic literature has primarily focused on reasoning architectures, explainability, and computational integration (Wan et al., 2024; Zhang & Sheng, 2024). Comparatively little work has empirically examined how neuro-symbolic instructional behaviors influence pedagogical outcomes within authentic tutoring dialogue. The present study therefore contributes an important educational dimension to neuro-symbolic AI research by demonstrating that symbolic tutoring structures are not merely computationally interpretable but pedagogically meaningful.

Perhaps the most important finding of the study concerns the near absence of meta-cognitive prompting within the tutoring interactions. Only six explicit meta-cognitive prompts

were identified across more than 2,000 tutoring exchanges. This finding is highly significant because meta-cognition represents one of the foundational mechanisms underlying human cognitive development. Educational psychology literature has consistently emphasized that reflective reasoning, self-monitoring, strategic thinking, and conceptual abstraction are essential components of deep learning and long-term knowledge transfer. However, the present findings suggest that current tutoring interactions remain overwhelmingly transactional and procedural rather than reflective or cognitively developmental.

This result directly supports recent neuro-symbolic AI literature identifying meta-cognition as one of the least explored dimensions within neuro-symbolic systems. Colelough and Regli (2025) found that only approximately 5% of neuro-symbolic AI studies addressed meta-cognitive reasoning capabilities. Similarly, Wei et al. (2024) argued that future neuro-symbolic architectures must incorporate adaptation, reflective reasoning, and self-monitoring mechanisms to achieve genuinely metacognitive AI systems. The present study extends this body of work by empirically demonstrating the practical absence of meta-cognitive scaffolding within real tutoring interactions. Consequently, the findings provide quantitative evidence supporting the need for neuro-symbolic learning companions capable of integrating reflective instructional mechanisms into conversational tutoring systems.

The lack of reflective scaffolding identified in the present study also aligns with recent educational AI research emphasizing that many AI learning tools prioritize procedural support over self-regulated learning processes. Studies examining AI-supported metacognition in STEM education similarly report that existing tutoring systems rarely encourage learners to reflect on reasoning processes or monitor conceptual understanding (Tsakeni, 2025). The present findings therefore reinforce growing concerns that many contemporary tutoring systems facilitate

instructional assistance without necessarily promoting deeper cognitive development. This distinction is critically important because educational AI systems designed solely for procedural tutoring may fail to cultivate independent reasoning capabilities or transferable conceptual understanding.

The conversational quality findings provide additional evidence supporting the viability of adaptive neuro-symbolic tutoring architectures. The mean BERTScore F1 value of 0.815 indicates strong semantic coherence and contextual continuity throughout the tutoring interactions. This suggests that the tutoring dialogues maintained meaningful conversational alignment between tutor and learner exchanges while preserving instructional relevance. Conversational coherence remains a substantial challenge in educational dialogue systems because tutoring interactions require contextual continuity across extended exchanges involving conceptual clarification, learner uncertainty, and adaptive instructional support. The present findings therefore demonstrate that neuro-symbolic tutoring frameworks can simultaneously preserve conversational quality and symbolic interpretability.

Emotional analysis complements these findings by revealing that most tutoring interactions remained emotionally neutral while positive affective states such as surprise and joy occurred substantially more frequently than negative emotional states. These findings suggest that the tutoring interactions generally maintained emotionally stable and supportive learning conditions. The prevalence of surprise and joy may additionally indicate moments of conceptual realization, successful scaffolding, and cognitive engagement during instructional exchanges. Educational psychology literature has consistently associated positive emotional engagement with improved learning motivation, sustained attention, and deeper cognitive processing. Consequently, the

findings suggest that adaptive tutoring systems may support emotionally productive instructional environments while maintaining strong conversational coherence.

Importantly, the integration of emotional analysis within a neuro-symbolic tutoring framework contributes to emerging interdisciplinary efforts combining affective computing, explainable AI, and educational technology. Existing neuro-symbolic educational research increasingly emphasizes the importance of integrating emotional awareness, explainability, and symbolic reasoning within adaptive tutoring systems (Ben Chaabene et al., 2025). However, empirical implementations of such integrated frameworks remain comparatively limited. The present study therefore contributes a meaningful methodological advancement by combining emotional analysis, symbolic tutoring logic, scaffolding evaluation, and conversational coherence assessment within a unified neuro-symbolic analytical framework.

From a broader theoretical perspective, the study contributes to ongoing discussions regarding the future direction of educational AI systems. The findings collectively suggest that future learning companions should evolve beyond answer-generation systems toward cognitively adaptive developmental agents capable of reflective scaffolding, symbolic reasoning, emotional awareness, and explainable instructional decision-making. Existing conversational tutoring systems frequently operate as procedural assistants that provide immediate instructional support without necessarily cultivating reflective reasoning or self-regulated learning behaviors. The present findings demonstrate that neuro-symbolic architectures may provide a promising pathway toward more developmentally aligned educational AI systems capable of supporting deeper cognitive engagement.

The study additionally contributes methodological innovation to the field of educational AI research. By integrating transformer-based neural analysis, symbolic reasoning, semantic

tutoring classification, conversational coherence metrics, and statistical educational evaluation within a single analytical pipeline, the study demonstrates the feasibility of comprehensive neuro-symbolic tutoring analysis using authentic conversational data. Existing educational AI research often examines these dimensions independently. The present work therefore contributes a unified analytical framework capable of simultaneously evaluating conversational quality, instructional explainability, scaffolding effectiveness, emotional dynamics, and symbolic pedagogical reasoning.

Despite these contributions, several limitations are considered. First, the study relied on secondary tutoring conversation data rather than live adaptive tutoring interventions. Consequently, the findings primarily reflect existing tutoring interaction patterns rather than experimentally controlled neuro-symbolic tutoring systems. Second, some symbolic reasoning categories relied partially on semantic classification heuristics rather than fully supervised pedagogical annotation. Although this approach enabled scalable analysis, future research may benefit from expert educational annotation frameworks to improve symbolic tutoring categorization accuracy. Third, the learning gain measurements were inferred from conversational correctness patterns rather than longitudinal educational assessments. Future studies should therefore investigate neuro-symbolic tutoring systems using controlled experimental learning environments with direct cognitive outcome measures.

Several broader implementation challenges additionally remain. The proposed framework was evaluated using secondary tutoring interaction data rather than real-time classroom deployment, and therefore future research should investigate live adaptive neuro-symbolic tutoring environments involving direct learner interaction. Furthermore, the symbolic reasoning structures relied on predefined pedagogical inference rules that may require domain-specific adaptation

across educational disciplines. The framework also did not incorporate multimodal learning signals such as speech, gaze, or behavioral analytics, which may further strengthen learner-state modeling and affect-sensitive scaffolding. Future work should therefore investigate reinforcement learning, multimodal cognition, longitudinal learning evaluation, and real-world classroom integration to advance scalable explainable neuro-symbolic educational AI systems.

Nevertheless, the study provides strong empirical support for adaptive neuro-symbolic learning companions as a promising direction for future educational AI systems. The findings demonstrate that explainability, scaffolding strategy selection, conversational coherence, and reflective instructional support represent critical dimensions of effective educational AI design. Most importantly, the near absence of meta-cognitive prompting within existing tutoring interactions highlights a substantial developmental gap within current tutoring architectures. Future educational AI systems must therefore move beyond procedural tutoring toward adaptive neuro-symbolic architectures capable of supporting reflective reasoning, conceptual abstraction, emotional awareness, and cognitively meaningful learning progression.

Conclusion

This study demonstrated the potential of adaptive neuro-symbolic learning companions to advance educational AI beyond procedural conversational tutoring toward explainable, developmentally aligned, and cognitively adaptive instructional systems. By integrating transformer-based conversational intelligence with symbolic pedagogical reasoning, scaffolding theory, conversational coherence evaluation, emotional analysis, and explainability mechanisms, the proposed framework contributes a unified neuro-symbolic architecture capable of supporting transparent and pedagogically meaningful educational interaction.

The findings revealed substantial limitations within existing tutoring interactions, particularly the dominance of direct-answer instructional strategies and the near absence of reflective meta-cognitive scaffolding. These findings reinforce growing concerns that many contemporary tutoring systems prioritize procedural assistance over deeper conceptual understanding, self-regulated learning, and cognitive development. In response, the proposed framework demonstrated how neuro-symbolic educational AI systems may support more adaptive, interpretable, and cognitively aligned instructional interaction through symbolic learner modeling, explainable tutoring logic, and developmentally grounded scaffolding mechanisms.

The study contributes to emerging interdisciplinary research at the intersection of neuro-symbolic AI, explainable artificial intelligence, intelligent tutoring systems, cognitive science, and educational technology. More importantly, it advances the broader vision of AI learning companions capable of functioning not merely as conversational assistants, but as transparent and adaptive cognitive partners supporting human learning, reflective reasoning, and long-term intellectual development.

Declarations

Ethics Approval and Consent to Participate

Not applicable. This study utilized fully anonymized secondary data obtained from a publicly available dataset. No direct interaction with human participants occurred, and no identifiable personal information was collected. According to institutional research ethics guidelines, ethical approval was not required for this study.

Consent to Publish

Consent to publish: Not applicable.

Clinical Trial Registration

Clinical Trial Number: Not applicable.

Availability of Data

The study utilized the Khan Academy Tutoring Accuracy Dataset publicly available through GitHub: <https://github.com/Khan/tutoring-accuracy-dataset>. The dataset contains authentic tutoring interactions collected from educational conversational exchanges involving tutoring assistance and instructional problem-solving activities.

Funding

The authors received no external funding for this research.

Acknowledgements

Not applicable.

Author Contributions Statement

PA: Conceptualization, Data curation, Methodology, Investigation, Visualization, Writing – original draft, Writing – review & editing, Formal analysis, Project administration, Software, Validation, Visualization. **MQ:** Writing – review & editing, Supervision. **OB:** Visualization, Writing – original draft, Formal analysis. **IA:** Visualization, Writing – original draft

References

- Ben Chaabene, N. E. H., & Hammami, H. (2026). Neuro-symbolic synergy in education: a survey of LLM-knowledge graph integration for explainable reasoning and emotion-aware student support. *Smart Learning Environments*, 13(1), 6.
- Brusilovsky, P. (1994). Student model centered architecture for intelligent learning environments.
- Colelough, B. C., & Regli, W. (2022). Neuro-Symbolic AI in 2024: A Systematic Review.
- Daniels, H. (2016). Vygotsky and dialogic pedagogy. *Dialogic pedagogy: The importance of dialogue in teaching and learning*, 10(3), 48-67.
- Fulbright, R. (2020). The expertise level.
- Hitzler, P. (2021). A review of the semantic web field. *Communications of the ACM*, 64(2), 76-83.
- Hu, B., Zheng, L., Zhu, J., Ding, L., Wang, Y., & Gu, X. (2024). Teaching plan generation and evaluation with GPT-4: Unleashing the potential of LLM in instructional design. *IEEE Transactions on Learning Technologies*, 17, 1445-1459.
- Hu, X., Xu, S., Tong, R., & Graesser, A. (2025). Generative AI in Education: From Foundational Insights to the Socratic Playground for Learning. *arXiv preprint arXiv:2501.06682*.
- Huynh, J., Jiao, C., Gupta, P., Mehri, S., Bajaj, P., Chaudhary, V., & Eskenazi, M. (2023). Understanding the effectiveness of very large language models on dialog evaluation. *arXiv preprint arXiv:2301.12004*.
- Jaldi, C. D., Ilkou, E., Schroeder, N., & Shimizu, C. (2024). Education in the Era of Neurosymbolic AI. *arXiv preprint arXiv:2411.12763*.
- Jaldi, C. D., Ilkou, E., Schroeder, N., & Shimizu, C. (2025). Education in the era of Neurosymbolic AI. *Journal of Web Semantics*, 85, 100857.
- Lindblom, J., & Ziemke, T. (2003). Social situatedness of natural and artificial intelligence: Vygotsky and beyond. *Adaptive Behavior*, 11(2), 79-96.

- Miller, P., & Dicerbo, K. LLM based math tutoring: challenges and dataset (2024). In.
- Nesbit, J. C., Adesope, O. O., Liu, Q., & Ma, W. (2014). How effective are intelligent tutoring systems in computer science education?
- Prather, J., Denny, P., Leinonen, J., Becker, B. A., Albluwi, I., Craig, M.,...Luxton-Reilly, A. (2023). The robots are here: Navigating the generative ai revolution in computing education. In *Proceedings of the 2023 working group reports on innovation and technology in computer science education* (pp. 108-159).
- Skjæveland, M. G., Balog, K., Bernard, N., Łajewska, W., & Linjordet, T. (2024). An ecosystem for personal knowledge graphs: A survey and research roadmap. *AI Open*, 5, 55-69.
- Sætra, H. S. (2025). Scaffolding Human Champions: AI as a More Competent Other. *Human Arenas*, 8(1), 56-78. <https://doi.org/10.1007/s42087-022-00304-8>
- Tong, J. R., & Lee, T. X. (2023). Trustworthy AI that engages humans as partners in teaching and learning. *Computer*, 56(5), 62-73.
- Tong, R. J., Cao, C. C., Lee, T. X., Zhao, G., Wan, R., Wang, F.,...Quevedo, J. (2023). NEOLAF, an LLM-powered neural-symbolic cognitive architecture. *arXiv preprint arXiv:2308.03990*.
- Tong, R. J., & Hu, X. (2024). Future of Education with Neuro-Symbolic AI Agents in Self-Improving Adaptive Instructional Systems. *Frontiers of Digital Education*, 1(2), 198-212. <https://doi.org/10.1007/s44366-024-0008-9>
- Tsakeni, M., Nwafor, S. C., Mosia, M., & Egara, F. O. (2025). Mapping the scaffolding of metacognition and learning by AI tools in STEM classrooms: A bibliometric–systematic review approach (2005–2025). *Journal of Intelligence*, 13(11), 148.

- Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard university press.
- Wan, Z., Liu, C.-K., Yang, H., Li, C., You, H., Fu, Y.,...Raychowdhury, A. (2024). Towards cognitive ai systems: a survey and prospective on neuro-symbolic ai. *arXiv preprint arXiv:2401.01040*.
- Wei, H., Shakarian, P., Lebiere, C., Draper, B., Krishnaswamy, N., & Nirenburg, S. (2024). Metacognitive AI: Framework and the case for a neurosymbolic approach.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of child psychology and psychiatry*, 17(2), 89-100.
- Xu, T., Tong, R., Liang, J., Fan, X., Li, H., & Wen, Q. (2024). Foundation models for education: Promises and prospects. *arXiv preprint arXiv:2405.10959*.
- Zhang, X., & Sheng, V. S. (2024). Neuro-symbolic AI: Explainability, challenges, and future trends. *arXiv preprint arXiv:2411.04383*.