A Neuro-Symbolic Approach to Extracurricular Activity Recommendation Based on Cognitive Profiling

Tatyana Bukina[1(✉)], Marina Khramova[2], Nikita Smirnov[1], Semen Kurkin[1], and Alexander Hramov[1]

[1] Research Institute of Applied Artificial Intelligence and Digital Solutions, Plekhanov Russian University of Economics, Moscow, Russia
bukinatatyanav@gmail.com

[2] Higher School of Cybertechnology, Mathematics and Statistics, Plekhanov Russian University of Economics, Moscow, Russia
hramova.mv@rea.ru

Abstract. The evolution of AI in education calls for systems that transition from being mere tools to becoming collaborative teammates, capable of synergistic partnership with educators. We present ADES (Activity-recommendation Decision-support Expert System), a neuro-symbolic system that takes a step toward this vision by providing explainable, personalized extracurricular activity recommendations based on EEG-derived cognitive profiles. ADES implements a Kautz Type 3 (Neuro | Symbolic) architecture, which cleanly separates a symbolic layer (fuzzy-logic cognitive classification, production-rule reasoning, temporal trend analysis) from a neural layer that conducts LLM-driven query expansion, embedding-based retrieval, and cross-encoder reranking. This design ensures that recommendations are both cognitively relevant and interest-aligned. We evaluate three configurations (full ADES, symbolic-only, and LLM-only baselines) using an LLM-as-a-judge protocol across five pedagogical dimensions. Results show that the neuro-symbolic approach significantly outperforms both ablations in cognitive relevance and interest alignment, while preserving the explainability crucial for human-in-the-loop collaboration. By providing a transparent reasoning trace, ADES shifts from being an automated tool toward acting as an auditable partner that supports shared decision-making between educators, students, and AI, advancing the vision of AI as an adaptive teammate in augmented learning.

Keywords: Neuro-symbolic AI · Educational recommendation systems · Cognitive profiling · EEG · Fuzzy logic · LLM · Kautz Type 3 architecture

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
E. G. Blanchard et al. (Eds.): AIED 2026, LNAI 16581, pp. 309–324, 2027.
https://doi.org/10.1007/978-3-032-29744-0_21

1 Introduction

Portable electroencephalography (EEG) devices are increasingly used in naturalistic school environments to monitor cognitive processes such as attention, working memory, and cognitive load [11,16]. However, a systematic review by García-Monge et al. [11] reveals that most EEG-based educational studies focus on real-time task-difficulty adaptation (e.g., adjusting game complexity [14,32]) rather than generating long-term personalized recommendations for cognitive development outside the classroom. This leaves a critical gap: how to translate neurophysiological cognitive profiles into concrete, evidence-based extracurricular activity recommendations.

The challenge is compounded by two factors. First, matching cognitive deficits to appropriate activities requires structured domain knowledge about which activities develop which cognitive functions—knowledge that is inherently symbolic and rule-governed. Second, ensuring that recommended activities resonate with student interests demands semantic understanding of natural-language descriptions, a task at which large language models (LLMs) excel but for which they lack pedagogical guardrails [25,30].

Recent work on second-classroom recommendation [31] uses GPT-4 and generative adversarial networks as black-box components, achieving high accuracy but sacrificing interpretability. Similarly, purely symbolic approaches such as knowledge-graph-based tutors [10] provide transparency but cannot bridge the *referential similarity gap* [29] between students' informal interest descriptions (e.g., "Minecraft") and formal activity taxonomies (e.g., "spatial reasoning training").

We address these limitations with ADES, a neuro-symbolic recommendation system classified as Kautz Type 3 (Neuro | Symbolic) [15]—an architecture in which symbolic and neural modules operate as independent subsystems with a defined interface, allowing each to contribute its strengths without tight coupling. Our contributions are:

1. A neuro-symbolic architecture that separates symbolic decision logic (fuzzy classification, production rules, temporal analysis) from neural semantic understanding (query expansion, embeddings, LLM reranking), enabling both pedagogical rigor and interest-aware personalization.
2. A curated knowledge base of 35 evidence-based extracurricular activities mapped to four cognitive functions assessed via EEG paradigms.
3. An LLM-as-a-judge evaluation protocol comparing four system configurations across five pedagogical quality dimensions.
4. Empirical evidence that the hybrid approach outperforms both symbolic-only and LLM-only baselines, particularly in cognitive relevance and interest alignment.

Finally, while ADES in its current implementation operates primarily as a decision-support tool, its neuro-symbolic architecture is fundamentally designed to enable a more collaborative, teammate-like role in the educational process.

The system does not merely output recommendations; it provides a transparent, auditable reasoning trace that serves as a shared context for deliberation between educators, students, and the AI. This design intentionally supports a human-in-the-loop paradigm where the AI's analytical precision and semantic understanding augment human expertise and oversight. We posit that this foundation of explainability and structured interaction is a critical prerequisite for evolving AIED systems from tools into adaptive teammates that co-evolve with their human partners towards shared educational objectives.

2 Related Work

2.1 Neuro-Symbolic Architectures in Education

Pure LLMs produce fluent but often hallucinated or pedagogically unsound outputs [25]. Recent work addresses this through neuro-symbolic designs where a symbolic control layer constrains LLM generation. Dass et al. [8] introduce Ivy, which separates Task-Method-Knowledge (TMK) models from LLM interpretation, showing that symbolic constraints improve explanation quality. ARGUS [30] combines GNN-based structural analysis with LLM feedback generation for argumentative writing. CyberBOT [34] uses domain ontologies to verify LLM-generated answers in cybersecurity education.

Particularly relevant to our work is Figueiredo [9], who demonstrates that encoding fuzzy scaffolding logic into LLM prompt schemas enables pedagogical adaptivity without fine-tuning. ADES extends this insight: rather than using fuzzy logic to modulate LLM prompts, we use it as a diagnostic front-end that classifies continuous cognitive scores into actionable categories, which then drive a production-rule expert system. The LLM components operate downstream, enhancing candidate retrieval through semantic matching.

2.2 EEG-Based Cognitive Profiling in Education

The cognitive functions targeted in EEG-based educational systems are grounded in established theoretical frameworks: working memory follows Baddeley's multi-component model [3], visual search draws on Treisman's feature integration theory [26], and cognitive flexibility is conceptualized within Miyake et al.'s executive function framework [20]. Wearable EEG sensors can reliably identify neurocognitive markers associated with specific learning abilities. Lekati et al. [16] profiled 30 sixth-graders during fraction learning, finding that EEG features (gamma and beta oscillations) distinguished three learner profiles. Gasimov et al. [12] used EEG to classify memorization modalities and personalize content delivery, reporting significant grade improvements.

The effectiveness of EEG-informed interventions is supported by several studies. Li et al. [18] demonstrated that EEG-guided adaptive learning improves cognitive control in children with ADHD over 8 weeks of training. Chaouachi et al. [6] showed that adapting learning strategies based on real-time EEG-derived

cognitive load and engagement indices improves both learning outcomes and emotional states.

Most existing systems implement real-time closed-loop adaptation within a single task (e.g., [14,32]). Our work extends the open-loop neuroadaptive system of Grubov et al. [13]—a pipeline that uses EEG assessments to build cognitive profiles of students and guide the selection of extracurricular activities aimed at developing specific cognitive functions. That system demonstrated improved cognitive outcomes in schoolchildren through EEG-guided interventions but lacked an automated recommendation component. ADES fills this gap, operating at a longer time horizon: EEG sessions produce cognitive profiles that inform activity recommendations applied over weeks or months, with the feedback loop closing through periodic re-assessment.

Zammouri et al. [33] provide a key architectural precedent by using fuzzy logic to combine learning outcomes and estimated mental effort (from EEG power spectral density) into a unified learner evaluation. ADES adopts a similar fuzzy classification approach but extends it with production rules for recommendation generation and LLM-based semantic matching.

2.3 LLM-Enhanced Recommender Systems in Education

LLMs have transformed educational recommendation from static filtering to dynamic content generation [19]. However, relying solely on LLMs introduces risks. Xu et al. [29] identify the *referential similarity gap*—the semantic mismatch between how learners describe interests and how educational content is taxonomized—and show that LLM-augmented retrieval (analogous to our query expansion) significantly improves exercise recommendation for language learning.

Hybrid architectures combining LLMs with structured logic are emerging as the standard for reliable educational AI. Troussas et al. [27] integrate fuzzy memory retention models with ChatGPT for personalized tutoring. Sampaio et al. [22] combine LLMs with fuzzy logic for content-level classification and recommendation. Tabarsi [24] demonstrates that rule-based content control alongside RAG is essential for trustworthy educational agents.

Knowledge-based recommendation using competency mapping provides a complementary perspective. Beutling and Spahic-Bogdanovic [4] link career goals to courses through competency ontologies extracted by LLMs, validating the approach of mapping abstract goals to concrete activities through an intermediate knowledge structure—directly paralleling our mapping of cognitive deficits to extracurricular activities through target functions.

For embedding-based retrieval in education, Li et al. [17] show that LLM-derived contextual embeddings outperform static knowledge-graph embeddings for educational content recommendation, supporting our use of `text-embedding-3-small` for activity retrieval.

3 System Architecture

3.1 Overview and Neuro-Symbolic Classification

ADES extends the open-loop neuroadaptive system for enhancing students' cognitive abilities developed by Grubov et al. [13]. In the original system, activity recommendations were generated manually by neuropsychologists based on cognitive profiles. ADES automates this step, providing a decision-support tool that suggests evidence-based activities while preserving human oversight. The full pipeline comprises: (1) EEG-based cognitive testing using validated paradigms; (2) biomarker computation from spectral analysis; (3) ADES recommendation generation (this work); (4) teacher/neuropsychologist verification; (5) multi-week activity execution; (6) re-assessment.

This design differs from real-time closed-loop BCI systems: no physiological feedback modifies recommendations during activity execution. Instead, the feedback loop operates at the course horizon (weeks to months) through periodic EEG re-assessment, temporal trend analysis, and strategy adaptation.

Following the taxonomy of Kautz [15], ADES is a Type 3 (Neuro | Symbolic) system: the symbolic and neural modules operate as independent subsystems with a defined interface. The symbolic layer handles diagnostic classification, constraint enforcement, and scoring logic. The neural layer handles semantic understanding of student interests and activity descriptions. Their outputs are fused in the expert system's scoring function, as shown in Fig. 1.

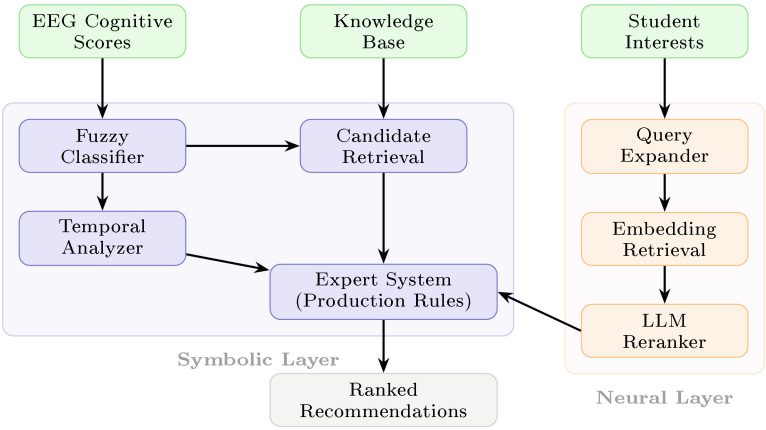


Fig. 1. ADES architecture. The symbolic layer (left, blue) handles diagnostic classification and rule-based scoring. The neural layer (right, orange) performs semantic interest matching. Both feed into the expert system for score fusion. (Color figure online)

3.2 Symbolic Layer

Fuzzy Logic Classification. Following the approach of Grubov et al. [13], cognitive test scores are classified relative to the peer group distribution. For each cognitive function, scores are categorized into three levels based on quartile boundaries:

- **Low**: scores in the first quartile (Q1)—below-average performance requiring intervention
- **Medium**: scores in the second and third quartiles (Q2–Q3)—average performance
- **High**: scores in the fourth quartile (Q4)—above-average performance

This relative classification adapts to the specific cohort, avoiding arbitrary absolute thresholds. A cognitive function is flagged as *deficient* when classified as "Low," triggering activity recommendation.

Temporal Trend Analysis. When historical assessment data is available, the system compares current scores against prior sessions. Meaningful improvement triggers an EXPLOIT strategy (continue similar activities), while stagnation across multiple functions triggers EXPLORE (diversify recommendations). Previously assigned activities for stagnated functions receive a scoring penalty.

Knowledge Base. The knowledge base contains 35 evidence-based extracurricular activities [5], each annotated with a target cognitive function, schedule, description, and resource requirements. Activities are drawn from published intervention studies and mapped to four cognitive domains assessed in the EEG battery: Working Memory (Sternberg paradigm), Visual Search (selective attention), Mental Arithmetic, and Combined Functions (cognitive flexibility). Table 1 shows representative entries.

Table 1. Sample entries from the activity knowledge base (35 unique activities).

Activity	Target Function	Schedule
Chess	Visual Search	1–2×/week, 45 min
Music Lessons	Working Memory	1×/week, 45 min
Basketball	Mental Arithmetic	2×/week, 45 min
Abacus	Combined Functions	2×/week, 45 min
Dance Aerobics	Working Memory	2×/week, 45 min

Production-Rule Expert System. The expert system implements a set of production rules for candidate scoring. Given a set of deficit functions D, the system retrieves all activities whose target function matches any $d \in D$, then applies the following rules:

R1 Base score initialization: Each candidate activity receives a base score $s_0 = 50$.
R2 Resource constraint: IF all required resources (e.g., Computer_Lab, Chess_Club) are available at the school THEN $s \mathrel{+}= 10$ ELSE $s \mathrel{-}= 5$.
R3 Semantic interest match: IF the neural semantic matching score $m > 0.35$ THEN $s \mathrel{+}= 20 \cdot m$ (see Sect. 3.3).
R4 Failed intervention penalty: IF the activity was previously assigned AND the targeted function showed Stagnation THEN $s \mathrel{-}= 50$.
R5 Strategy selection: IF temporal analysis indicates overall Stagnation THEN switch to EXPLORE strategy (penalize previously tried activities more aggressively via R4).

The final ranking sorts candidates by total score s and selects the top 3 unique activities. The weight magnitudes establish a clear priority ordering among rules. The base score $s_0 = 50$ means that R4 (-50) effectively eliminates a previously failed activity from the ranking—the strongest constraint, reflecting the pedagogical priority of avoiding repeated ineffective interventions. R2's asymmetry ($+10/-5$) treats resource unavailability as a soft penalty (the activity may still be feasible outside the school), while confirmed availability provides a stronger positive signal. These magnitudes were chosen to reflect domain priorities. This production-rule formalism provides complete transparency: each score component is traceable to a specific rule and input condition, enabling educators to understand and override the system's reasoning.

3.3 Neural Layer: Semantic Interest Matching

The neural layer addresses the referential similarity gap [29] between students' informal interest descriptions and formal activity descriptions through a three-stage pipeline (Fig. 2).

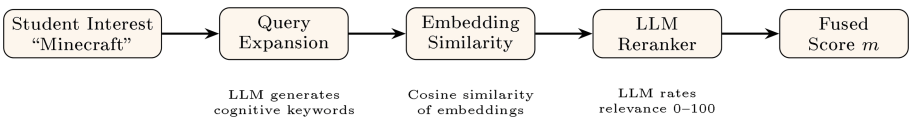


Fig. 2. Semantic matching pipeline. Student interests are expanded, embedded, and reranked against activity descriptions.

Query Expansion. Raw student interests (e.g., "Video Games", "Sport") are semantically expanded using an LLM (Gemini 2.0 Flash via OpenRouter API). The LLM generates educational and cognitive keywords: for instance, "Minecraft" expands to "Minecraft, spatial reasoning, planning, resource management, creativity, problem-solving, 3D navigation." This expansion bridges the vocabulary gap between informal interest labels and educational activity descriptions.

Embedding-Based Retrieval. Activity descriptions are pre-encoded into vector representations using `text-embedding-3-small` [17]. Expanded interest queries are similarly embedded, and cosine similarity identifies the best-matching interest for each candidate activity:

$$s_{\text{emb}}(q, a) = \frac{\mathbf{e}_q \cdot \mathbf{e}_a}{\|\mathbf{e}_q\| \cdot \|\mathbf{e}_a\|} \tag{1}$$

where $\mathbf{e}_q$ and $\mathbf{e}_a$ are the embedding vectors of the expanded query and activity description, respectively.

LLM Reranking. The top embedding candidates are reranked by prompting an LLM to rate the relevance of each (interest, activity) pair on a 0–100 scale. This captures nuances that embedding similarity alone may miss (e.g., that chess is more relevant to a "strategy games" interest than a generic "games" interest).

Score Fusion. The final semantic matching score fuses embedding and reranker signals:

$$m = \alpha \cdot s_{\text{emb}} + (1 - \alpha) \cdot s_{\text{rerank}}, \quad \alpha = 0.4 \tag{2}$$

This score is injected into the expert system via rule R3 (Sect. 3.2), completing the neuro-symbolic integration. The weight $\alpha = 0.4$ was set empirically, giving greater influence to the reranker which provides more contextual judgment.

3.4 Data

The system was evaluated using data collected as part of the Grubov et al. [13] study: 45 students (aged 9–12) from two schools who underwent two rounds of EEG-based cognitive testing separated by approximately 3 months. Each session measured four cognitive functions (0–100 scale): Working Memory, Visual Search, Mental Arithmetic, and Combined Functions. Students also reported their interests. The second-round data was used for evaluation. The code and anonymized data will be available upon request.

4 Evaluation

4.1 Experimental Setup

We compare four recommendation approaches:

- **ADES** (full system): Fuzzy classification + expert rules + semantic matching (query expansion, embeddings, LLM reranker).
- **Symbolic-only**: Fuzzy classification + expert rules without semantic matching. Activities are scored based on resource availability and intervention history alone.
- **LLM-only**: A direct LLM prompt receives the student's cognitive scores, deficit analysis, interests, and the full activity list, and is asked to recommend the top 3 activities with justifications. No symbolic reasoning is applied.
- **LLM-few**: A few-shot LLM baseline that additionally receives exemplar ADES recommendations for three reference students, providing the model with concrete examples of high-quality, deficit-targeted recommendations before generating its own.

This comparison follows an ablation study design: the Symbolic-only and LLM-only configurations are obtained by removing specific components of ADES, isolating the contribution of each architectural layer. Notably, the LLM-only baseline is not a naive prompt—it receives the complete student context (cognitive profile, deficit analysis, interests, and the full activity catalogue), making it an informed baseline. The LLM-few baseline further tests whether providing exemplar ADES recommendations (few-shot prompting) can close the gap without symbolic reasoning. Comparison with established educational recommender systems (e.g., collaborative filtering, knowledge-graph-based methods) is left for future work, as adapting these methods to operate under the same input constraints (EEG-derived cognitive profiles without historical interaction data) requires careful design to ensure a fair comparison.

For each of the 45 students, each approach produces a ranked list of 3 recommended activities. The recommendations are then evaluated by an LLM-as-a-judge protocol [21,35].

4.2 LLM-As-A-Judge Protocol

Each set of recommendations is evaluated by two independent LLM judges (Gemini 2.0 Flash with temperatures 0.1 and 0.7) across five dimensions on a 1–5 Likert scale:

1. **Cognitive Relevance**: Do the recommended activities target the identified cognitive deficits?
2. **Interest Alignment**: Do the recommendations match the student's stated interests?
3. **Pedagogical Appropriateness**: Are the activities evidence-based and age-appropriate?

4. **Diversity**: Are the top-3 recommendations sufficiently varied in type?
5. **Explainability**: Is the rationale for each recommendation clear and traceable?

Each judge receives the student's cognitive profile (scores, fuzzy classifications, deficit analysis), interests, and the three recommended activities with their scoring breakdowns (for ADES and Symbolic-only) or LLM-generated justifications (for LLM-only). The judge returns a JSON response with scores and brief justifications for each dimension.

We report mean scores across all students for each dimension and approach, inter-rater agreement between the two judges using Cohen's weighted κ [7], and Wilcoxon signed-rank tests [28] for pairwise comparisons. Traditional ranking metrics such as MRR or NDCG require ground-truth relevance rankings for each query. In our setting, no expert panel has yet ranked the 35 candidate activities for each student profile, and constructing such a gold standard is non-trivial given that multiple activity combinations may be equally appropriate for a given cognitive profile. Establishing expert-annotated ground-truth rankings to enable standard retrieval metrics remains an important direction for future work.

4.3 Results

Table 2 presents the mean evaluation scores across all 45 students. The full ADES system achieves the highest scores on four of five dimensions. The few-shot LLM baseline (LLM-few), which receives exemplar ADES recommendations, improves over the zero-shot LLM-only on Interest Alignment (4.03 vs. 3.90) and Explainability (4.54 vs. 3.52), but still falls short of ADES on Cognitive Relevance and Diversity, confirming that examples alone cannot substitute for structured knowledge retrieval.

Table 2. Mean LLM-as-a-judge scores (1–5 scale) across 45 students. Bold indicates best per dimension. *Significant improvement over next-best approach ($p < 0.05$, Wilcoxon signed-rank test).

Dimension	ADES	Symbolic-only	LLM-only	LLM-few
Cognitive Relevance	**4.52***	4.38	3.67	3.89
Interest Alignment	**4.31***	2.48	3.90	4.03
Pedagogical Appr.	**4.43**	4.29	3.81	3.97
Diversity	**4.10**	3.86	3.95	3.71
Explainability	4.19	4.33	3.52	**4.54**
Overall Mean	**4.31**	3.87	3.77	4.03

Key Findings. Cognitive Relevance. ADES (4.52) and the Symbolic-only baseline (4.38) both score highly, as the symbolic layer ensures activities are retrieved specifically for identified deficits. The LLM-only approach (3.67) sometimes recommends activities that are broadly beneficial but do not target the specific deficit, confirming the value of structured knowledge retrieval.

Interest Alignment. This dimension shows the largest spread. The Symbolic-only baseline (2.48) scores poorly because it has no mechanism to consider student interests. The LLM-only approach (3.90) performs well via natural language understanding but occasionally hallucinates connections. ADES (4.31) combines the best of both: the neural layer matches interests while the symbolic layer constrains recommendations to deficit-relevant activities.

Explainability. The Symbolic-only baseline scores highest (4.33) due to its fully transparent scoring trace. ADES (4.19) retains most of this transparency, as the expert system's production rules provide an auditable decision trail even with the neural matching component. The LLM-only approach (3.52) provides natural-language justifications that can be plausible-sounding but are not grounded in verifiable logic.

Inter-rater Agreement. The two LLM judges showed substantial agreement with a weighted Cohen's $\kappa = 0.71$ across all dimensions, indicating reliable evaluation despite temperature variation.

4.4 Expert System Scoring Rules

Table 3 summarizes the expert system's scoring components observed across the evaluation.

Table 3. Expert system scoring rule summary with observed effect ranges.

Rule	Condition	Score Effect
R1: Base	All candidates	+50
R2: Resource	Available / Unavailable	+10 / −5
R3: Interest	Semantic match $m > 0.35$	+7 to +20
R4: History	Previously failed	−50
R5: Strategy	Stagnation → Explore	Amplifies R4

5 Discussion

5.1 Neuro-Symbolic Integration Benefits

The evaluation confirms the complementary strengths of symbolic and neural components. The symbolic layer ensures *cognitive relevance*: activities are retrieved from the knowledge base by exact target-function matching, guaranteeing that recommendations address identified deficits. The neural layer ensures

interest alignment: semantic matching personalizes the selection among cognitively appropriate candidates based on the student's informal interest descriptions.

This decomposition also yields practical benefits. The symbolic layer's production rules provide a complete audit trail: for any recommendation, educators can trace the score to specific rules and verify each component. This addresses the XAI requirement that is critical in education [1,2]—teachers and neuropsychologists must understand *why* the system recommends swimming over chess to calibrate their trust in the system and make informed override decisions.

It is worth noting that the symbolic-only baseline limits interest matching capabilities, effectively shifting the burden of aligning recommendations with student interests to the teacher. This increases the educator's workload. This design choice isolates the contribution of neural semantic matching.

5.2 Open-Loop Architecture Clarification

As an extension of the system developed by Grubov et al. [13], ADES operates within an *open-loop neuroadaptive* architecture. It is important to clarify the control-theoretic semantics: the system is open-loop *during activity execution*—no real-time physiological feedback modifies the recommendation while a student practices chess or plays basketball. However, a feedback loop operates at the *course horizon* (weeks to months): periodic EEG re-assessment provides updated cognitive profiles, the temporal analysis module detects growth or stagnation, and the strategy-switching mechanism (R5) adjusts the recommendation approach accordingly. In this sense, the system implements a *sampled closed loop* with a long sampling period determined by the re-assessment schedule.

This architecture suits extracurricular activity recommendation, where interventions are multi-week programs rather than single task steps—closed-loop adaptation within a chess lesson is neither feasible nor meaningful [32].

5.3 From Decision-Support Tool to Collaborative Teammate

The ADES architecture, while evaluated in a specific decision-support context, embodies key design principles that facilitate a transition toward richer Human-AI synergy, aligning with the vision of AI as an adaptive teammate. The strict separation between the symbolic (pedagogic constraint enforcement) and neural (semantic understanding) layers provides a modular foundation for incorporating interactive and adaptive components. Currently, the synergy is realized in a structured, asynchronous manner: the AI proposes and justifies, while the human verifies and authorizes. This embodies an initial, crucial step in human-AI collaboration—establishing calibrated trust through explainability.

To evolve toward a more participatory partnership, the architecture can be extended: the symbolic production system can be integrated with a dialogue management module for interactive recommendation refinement, and the temporal trend analysis can be enhanced to model longitudinal outcomes of human-AI decision patterns. This pathway underscores that effective human-AI team-

ing in education often begins not with open-ended generative agents, but with auditable, constrained systems designed for mutual understanding.

5.4 Limitations

Several limitations should be noted. First, the LLM-as-a-judge evaluation, while increasingly accepted [21,23,35], does not replace evaluation with human domain experts or longitudinal studies measuring actual cognitive outcomes. Although we use LLM judges while noting limitations of LLM-generated recommendations, the two tasks are qualitatively different: structured rubric-based evaluation with explicit scoring anchors versus open-ended recommendation generation. Validating the LLM-judge protocol against human-expert assessments remains important future work. However, its application is justified for our preliminary assessment as Zheng et al. [35] have shown that LLM-as-a-judge exhibits a high correlation with human judgment for certain educational tasks. Second, the knowledge base is manually curated; extending it to new activities requires domain expertise. Third, the semantic matching component depends on LLM API availability and incurs latency and cost. Fourth, the current evaluation uses two schools in Kaliningrad and Moscow, Russia, which may not generalize to other educational contexts.

Data Privacy. Student data was anonymized before processing. The current prototype uses cloud-based LLM APIs for semantic matching; a production deployment would employ locally-hosted models to ensure data remains within the school infrastructure.

Future work includes longitudinal outcome tracking, teacher/neuropsychologist evaluation, semi-automated knowledge base expansion, and local embedding models to reduce API dependency.

6 Conclusion

We presented ADES, a neuro-symbolic system for recommending extracurricular activities based on EEG-derived cognitive profiles. The system demonstrates that combining symbolic reasoning (fuzzy classification, production rules, temporal analysis) with neural semantic matching (query expansion, embeddings, LLM reranking) yields recommendations that are both cognitively targeted and interest-aligned—outperforming both symbolic-only and LLM-only alternatives across multiple pedagogical quality dimensions.

The Kautz Type 3 architecture provides a principled separation of concerns: the symbolic layer guarantees pedagogical constraints and explainability, while the neural layer bridges the referential similarity gap between student interests and activity descriptions. This design supports the human-in-the-loop verification step that is essential in educational contexts.

The system extends the open-loop neuroadaptive pipeline of Grubov et al. [13], automating the recommendation step that was previously performed

manually, and contributes to the emerging field of neuroeducation-informed AI systems.

Our work illustrates that the path from AI-as-a-tool to AI-as-a-teammate begins with architectures designed for mutual intelligibility. The symbolic layer encodes domain-specific pedagogical knowledge in a form that teachers can audit and adapt; the neural layer interprets student voices in natural language. Together, they create a shared space for iterative, dialogic planning—a foundational step toward co-adaptive learning ecosystems where humans and AI jointly respond to learner development over time.

Looking forward, future iterations incorporating interactive dialogue and adaptation to user feedback can transform the system from a recommendation tool into a collaborative teammate—one that engages educators in joint interpretation of cognitive data and co-construction of development plans, augmenting rather than automating the educator's role.

References

1. Abu-Rasheed, H., Abdulsalam, M.H., Weber, C., Fathi, M.: Supporting student decisions on learning recommendations: an LLM-based chatbot with knowledge graph contextualization for conversational explainability and mentoring (2024). arXiv:2401.08517 arXiv preprint
2. Abu-Rasheed, H., Weber, C., Fathi, M.: Knowledge graphs as context sources for LLM-based explanations of learning recommendations. In: 2024 IEEE Global Engineering Education Conference (EDUCON), pp. 1–5. IEEE (2024)
3. Baddeley, A.: The episodic buffer: a new component of working memory? Trends Cogn. Sci. **4**(11), 417–423 (2000)
4. Beutling, N., Spahic-Bogdanovic, M.: Personalised course recommender: linking learning objectives and career goals through competencies. In: Proceedings of the AAAI Symposium Series, vol. 3, pp. 72–81. AAAI (2024)
5. Bukina, T.V.: Creation of a recommendations database for the neuroeducational software recommendation service. Nauchnoe Mnenie (Scientific Opinion (5) (2025)
6. Chaouachi, M., Jraidi, I., Lajoie, S.P., Frasson, C.: Enhancing the learning experience using real-time cognitive evaluation. Int. J. Info. Educ. Technol. **9**(10), 678–688 (2019)
7. Cohen, J.: A coefficient of agreement for nominal scales. Educ. Psychol. Measur. **20**(1), 37–46 (1960)
8. Dass, R., Bowlin, T., Li, Z., Jin, X., Goel, A.: Improving procedural skill explanations via constrained generation: a symbolic-LLM hybrid architecture (2025). arXiv:2511.20942 arXiv preprint
9. Figueiredo, V.: A fuzzy logic prompting framework for large language models in adaptive and uncertain tasks (2025). arXiv:2508.06754 arXiv preprint
10. Gao, F., Xu, S., Hao, W., Lu, T.: KA-RAG: integrating knowledge graphs and agentic retrieval-augmented generation for an intelligent educational question-answering model. Appl. Sci. **15**(23), 12547 (2025)
11. García-Monge, A., Rodríguez-Navarro, H., Bores-García, D., González-Calvo, G.: Electroencephalography in naturalistic and semi-naturalistic educational contexts: a systematic review. Rev. Educ. **12**(3) (2024)

12. Gasimov, H., Alibeyli, T., Hesenli, H., Ismayilov, A.: Development of a personalized learning trajectory using a brain-computer interface. Technol. Audit Product. Reserv. **5**(2(85)), 6–12 (Oct 2025)
13. Grubov, V.V., et al.: Open-loop neuroadaptive system for enhancing student's cognitive abilities in learning. IEEE Access **12**, 49034–49049 (2024)
14. Kadar, M.: Real-time attention measurement using wearable brain-computer interfaces in serious games. Appl. Syst. Innov. **8**(6), 166 (2025)
15. Kautz, H.: The third AI summer: AAAI robert s. engelmore memorial lecture. AI Mag. **43**(1), 93–104 (2022)
16. Lekati, E., et al.: Wearable EEG sensor analysis for cognitive profiling in educational contexts. Sensors **25**(20), 6446 (2025)
17. Li, X., Henriksson, A., Duneld, M., Nouri, J., Wu, Y.: Evaluating embeddings from pre-trained language models and knowledge graphs for educational content recommendation. Future Int. **16**(1), 12 (2023)
18. Li, Y., Tian, C., Xu, L., Pei, L., Huang, X., Wang, X.: EEG-guided adaptive learning: a new neuroeducational approach to the facilitation of cognitive control in ADHD children. Child Care Health Dev. **51**(4), e70113 (2025)
19. Liu, Z., Agrawal, P., Singhal, S., Madaan, V., Kumar, M., Verma, P.K.: LPITutor: an LLM based personalized intelligent tutoring system using RAG and prompt engineering. Peer J. Comput. Sci. **11**, e2991 (2025)
20. Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., Wager, T.D.: The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. Cogn. Psychol. **41**(1), 49–100 (2000)
21. Qian, K., et al.: Towards reliable generative AI-driven scaffolding: reducing hallucinations and enhancing quality in self-regulated learning support. arXiv preprint arXiv:2509.05929 (2025)
22. Sampaio, G.L., Oliveira, R.C.D., Melo, E.F., Santos, K.O.A.: Educational content recommendation platform using LLMs and fuzzy logic. Revista Ibero-Americana de Humanidades, Ciências e Educação **11**(6), 2518–2529 (2025)
23. Scarlatos, A., Liu, N., Lee, J., Baraniuk, R., Lan, A.: Training LLM-based tutors to improve student learning outcomes in dialogues (2025). arXiv:2507.06424 arXiv preprint
24. Tabarsi, B.: Developing LLM-powered trustworthy agents for personalized learning support. Proc. AAAI Conf. Artif. Intell. **39**(28), 29301–29302 (2025)
25. Tithi, S.D., et al.: The promise and limits of LLMs in constructing proofs and hints for logic problems in intelligent tutoring systems. arXiv preprint arXiv:2511.04736 (2025)
26. Treisman, A.M., Gelade, G.: A feature-integration theory of attention. Cogn. Psychol. **12**(1), 97–136 (1980)
27. Troussas, C., Krouska, A., Mylonas, P., Sgouropoulou, C., Voyiatzis, I.: Fuzzy memory networks and contextual schemas: enhancing ChatGPT responses in a personalized educational system. Computers **14**(3), 89 (2025)
28. Wilcoxon, F.: Individual comparisons by ranking methods. Biometrics Bulletin **1**(6), 80–83 (1945)
29. Xu, A., Monroe, W., Bicknell, K.: Large language model augmented exercise retrieval for personalized language learning. In: Proceedings of the 14th Learning Analytics and Knowledge Conference, pp. 284–294. ACM (2024)
30. Yang, L., Zhao, S.: ARGUS: a neuro-symbolic system integrating GNNs and LLMs for actionable feedback on English argumentative writing. Systems **13**(12), 1079 (2025)

31. Yang, Q., Liang, C.: A second-classroom personalized learning path recommendation system based on large language model technology. Appl. Sci. **15**(14), 7655 (2025)
32. Yuksel, B.F., et al.: Learn piano with BACh: an adaptive learning interface that adjusts task difficulty based on brain state. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 5372–5384. ACM (2016)
33. Zammouri, A., Moussa, A.A., Chevallier, S.: Use of cognitive load measurements to design a new architecture of intelligent learning systems. Expert Syst. Appl. **237**, 121253 (2024)
34. Zhao, C., et al.: CyberBOT: towards reliable cybersecurity education via ontology-grounded retrieval augmented generation. arXiv preprint arXiv:2509.00389 (2025)
35. Zheng, L., et al.: Judging LLM-as-a-judge with MT-bench and chatbot arena. In: Advances in Neural Information Processing Systems, vol. 36 (2023)