

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

CogEvolution: A Human-like Generative Educational Agent to Simulate Student's Cognitive Evolution

Permalink

<https://escholarship.org/uc/item/5rq5s10c>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Zhang, Wei
cheng, yihang
Ye, zhirong
[et al.](#)

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

CogEvolution: A Human-like Generative Educational Agent to Simulate Student's Cognitive Evolution

Wei Zhang^{1,2}, Yihang Cheng^{*1,2}, Zhirong Ye^{1,3} & Kezhen Huang^{1,3}

¹Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan, Hubei, China

²National Engineering Research Center for Educational Big Data, Central China Normal University, Wuhan, Hubei, China

³Central China Normal University Wollongong Joint Institute, Central China Normal University, Wuhan, Hubei, China

*Corresponding author: yihangcheng@mails.ccnuc.edu.cn

Abstract

Generative Agents, owing to their precise modeling and simulation capabilities of human behavior, have become a pivotal tool in the field of Artificial Intelligence in Education (AIED) for uncovering complex cognitive processes of learners. However, existing educational agents predominantly rely on static personas to simulate student learning behaviors, neglecting the decisive role of deep cognitive capabilities in learning outcomes during practice interactions. Furthermore, they struggle to characterize the dynamic fluidity of knowledge internalization, transfer, and cognitive state transitions. To overcome this bottleneck, this paper proposes a human-like educational agent capable of simulating student cognitive evolution—**CogEvolution**. Specifically, we first construct a cognitive depth perceptron based on the Interactive, Constructive, Active, Passive (ICAP) taxonomy from cognitive psychology, achieving precise quantification of learner cognitive engagement. Subsequently, we propose a memory retrieval method based on Item Response Theory (IRT) to simulate the connection and assimilation of new and prior knowledge. Finally, we design a dynamic cognitive update mechanism based on evolutionary algorithms to simulate the real-time integration of student learning behaviors and cognitive evolution processes. Comprehensive evaluations demonstrate that CogEvolution not only significantly outperforms baseline models in behavioral fidelity and learning curve fitting but also uniquely reproduces plausible and robust cognitive evolutionary paths consistent with educational psychology expectations, providing a novel paradigm for constructing highly interpretable educational agents.

Keywords: Generative Agents; Cognitive Evolution; Educational Psychology; Artificial Intelligence in Education

Introduction

In recent years, breakthrough advancements in Large Language Models (LLMs) have provided novel opportunities for the deep integration of cognitive science and educational technology (Dai et al., 2025; Luo et al., 2025). As a significant application form of LLMs, Generative Agents demonstrate immense potential in simulating complex human behaviors due to their superior capabilities in memory retrieval, planning, reasoning, and reflection (Park et al., 2023; Summers et al., 2024; Xi et al., 2025). In cognitive science research, these agents are widely used to construct computational social models to verify sociological hypotheses or reproduce human decision-making processes in specific contexts (Aher et al., 2023; Chuang et al., 2024).

With deepening research, the application of generative agents in educational scenarios has become increasingly

prominent, particularly in simulating dynamic interactions within teaching relationships. Researchers have begun dedicating efforts to utilizing such agents to build high-fidelity "Student Simulations" (Markel, Gorgun, et al., 2023; Xu et al., 2024). By learning from real student response logs and interaction histories, these agents can assist in evaluating the effectiveness of instructional strategies or serve as standardized patients for training novice teachers (Xu et al., 2024), thereby enhancing educational quality in low-risk environments.

However, despite progress in simulating explicit behaviors, agents still face severe challenges in modeling deep cognitive mechanisms. Frontier research on EvoAgents (Li et al., 2025) has acutely pointed out the issue of "**Cognitive Dynamics Disconnect**" present in generative agents. Specifically, agents lacking theoretical support from cognitive psychology struggle to possess the capability to truly simulate human internal mental activities. For example, mainstream architectures such as Smallville (Park et al., 2023) neglect the cognitive architectures driving these behaviors (Summers et al., 2024).

In educational scenarios, this "cognitive disconnect" phenomenon is particularly pronounced, which we specifically define as "**Missing Cognitive Features**" and "**Cognitive Path Stagnation**." Current educational agent architectures largely rely on **Static Personas** constructed based on response behavior features. Whether they are base agents constructed directly using native large models like Llama 3 (Meta AI, 2024), or behavioral modeling by LLM combined with deep learning neural networks (Lv et al., 2025), their essence lies in the parametric fitting of a snapshot of student ability. These static parametric methods neglect the essence of learning—a dynamic process of continuously reconstructing cognitive schemas through Assimilation and Accommodation (Piaget, 1976). They struggle to capture the dynamic fluidity of knowledge internalization, ability transfer, and cognitive states (such as from confusion to insight) generated by deep cognitive engagement (such as constructive and interactive modes in ICAP theory (Chi & Wylie, 2014)) during practice. To bridge this gap, this paper proposes the **CogEvolution**. This is the first computational agent based on educational psychology theories that focuses on the dynamic evolution of **cognitive states** and **cognitive paths** in agent learning activities.

Specifically, the main contributions of this paper are summarized as follows:

1. Refined the definition of the cognitive dynamics disconnect

problem in the educational domain and proposed the CogEvolution, pioneering a research paradigm for exploring cognitive evolutionary paths in generative agents.

2. Innovatively integrated the ICAP cognitive perception module as the adapter of CogEvolution, designed a memory retrieval based on IRT theory (Lord, 1980), and a state update mechanism inspired by evolutionary algorithms, realizing a paradigm shift from "static persona" to "dynamic cognitive flow."
3. Constructed the CogMath-948 dataset containing cognitive state annotations and demonstrated through comprehensive evaluation that CogEvolution significantly outperforms existing baseline models in learning curve fitting and knowledge transfer capabilities.

Related Work

With Large Language Models (LLMs) demonstrating powerful reasoning and generalization capabilities, research on Generative Agents has experienced explosive growth (Dai et al., 2026; Luo et al., 2025; Xi et al., 2025). The "Stanford Town" simulation proposed by Park et al. (Park et al., 2023) is a milestone work in this field, constructing an architecture containing Memory Stream, Reflection, and Planning, enabling agents to simulate long-term human social interactions. Subsequently, this paradigm was widely extended: Wang et al. (Wang et al., 2024) systematically reviewed the potential of LLM-based autonomous agents in task planning and tool use; Li et al. (Li et al., 2025) proposed EvoAgents, which further introduced psychodynamic mechanisms, resolving the issue of rigid personalities in agent social simulations. Additionally, the emergence of open-source evaluation platforms like AgentSims (Sumers et al., 2024) has lowered the threshold for constructing high-fidelity social simulation environments. However, these general architectures mostly focus on the breadth of social behaviors and the coherence of interpersonal interactions, lacking fine-grained modeling of individual internal cognitive processes (such as knowledge construction and conceptual change), making them difficult to directly transfer to educational learning scenarios that require high cognitive depth.

In the field of education, utilizing agents to simulate students (Student Simulation) to assist in instructional evaluation and teacher training has become a core topic. The GPTeach platform developed by Markel et al. (Markel, Gorgun, et al., 2023) uses LLMs to generate virtual students with different knowledge backgrounds for training teaching assistants' instructional response capabilities. To enhance simulation realism, Xu et al. (Xu et al., 2024) proposed EduAgent, which achieves multimodal cloning of student classroom behaviors (such as raising hands, disruptive behavior) by integrating audiovisual and text data from real classrooms. Although these works have achieved success in simulating explicit instructional behaviors, their core mostly relies on Static Personas or behavioral cloning (Mandyam et al., 2025), meaning agents tend to generate fixed response patterns based on pre-

set labels of "low-performing" or "high-performing" students, rather than dynamically evolving their knowledge states based on Cognitive Engagement during interactions, leading to the phenomenon of "Missing Cognitive Features."

In summary, although generative agents have made significant progress in behavioral simulation and instructional assistance, existing research still has gaps in simulating the dynamics of learner cognitive development and psychological interpretability. There is an urgent need to establish a computational method capable of real-time alignment between external learning behaviors and internal cognitive state evolution.

Model Design

The CogEvolution agent aims to simulate a cognitive subject capable of dynamic evolution. As shown in Figure 1, its core logic lies in perceiving cognitive depth through the ICAP mechanism and utilizing evolutionary algorithms to drive the restructuring of knowledge structures. The Cogevolution consists of three highly coupled components: Cognitive Adapter for Depth Perceptron, an agent input adapter responsible for environmental perception and cognitive feature initialization; IRT-Driven Memory Retrieval Module, executing memory retrieval, state generation, and knowledge evolution; Evolutionary Algorithm-Based Cognitive State Update Mechanism, used to couple the simulation of learning behaviors with their corresponding cognitive paths and dynamically update cognitive evolution.

Symbol Definition

To provide a rigorous formal description of the cognitive evolution process, we define the symbol system as shown in Table 1.

Table 1: Symbol Definitions and Parameter Descriptions

Symbol	Description and Physical Meaning
$\mathcal{A}$	Agent object
I_t	Input vector at time t (question, logs, environmental features)
S_t	Agent comprehensive state, $S_t = \{P, C, K\}$
K_t	Knowledge Structure , long-term stable cognitive schema
C_t	Cognitive State , transient state (e.g., confusion, flow)
$\hat{p}_t$	Mastery confidence probability, used for AUC/RMSE calculation
L_t	Cognitive Engagement Depth, $L \in \{Passive, Active, Constructive, Interactive\}$
$\hat{y}_{icap}$	Probability distribution vector over ICAP levels
v_i	Baseline gain coefficient for the i -th ICAP level
ω_L	Cognitive Evolution Rate , weighted continuous ICAP intensity
$\mathcal{M}$	Memory bank storing quadruples $(I, Response, C, L)$
Sim_{struct}	Structural similarity based on IRT characteristic curve distance
Sim_{sem}	Semantic similarity based on text embedding
θ_t	Learner's latent ability parameter in IRT
$P_i(\theta)$	Item Characteristic Function (ICC) for question i
$\mathcal{G}_{evo}$	Mutation Operator: Hypothesis Generation Function
$\mathcal{H}$	Set of generated cognitive hypotheses (offspring)
σ	Mutation intensity parameter modulated by emotion
ZPD	Zone of Proximal Development , constrains effective search space
S_{eval}	Selection Operator: Consistency Evaluation Function
γ	Penalty coefficient for deviating from ZPD
δ_{align}	Behavior-Cognition Alignment Coefficient

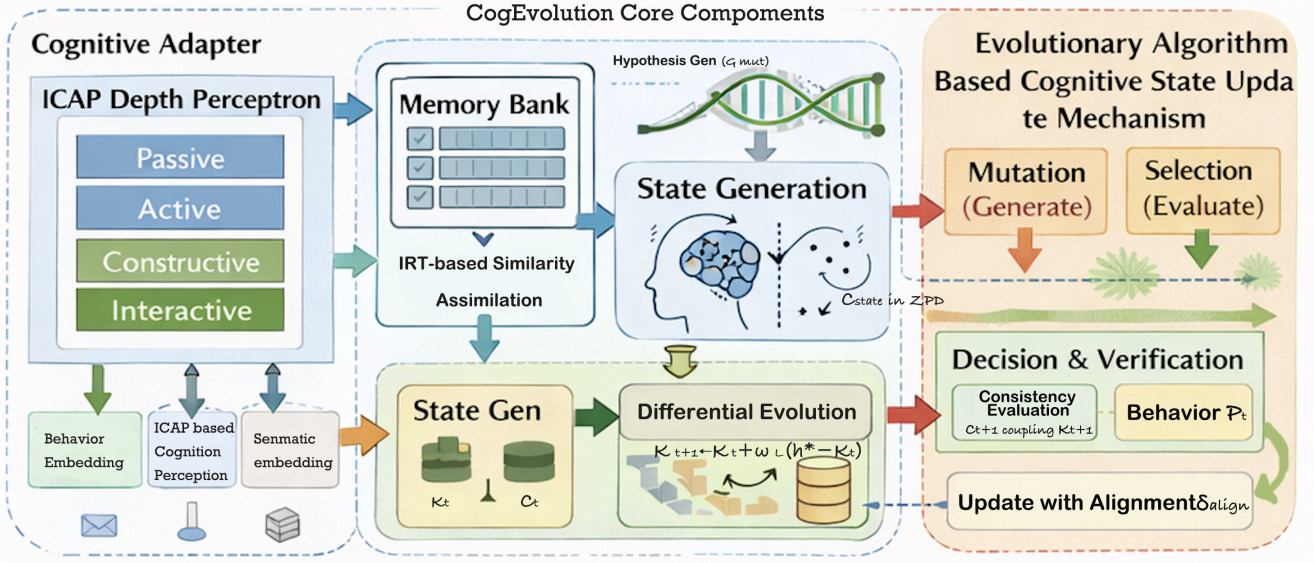


Figure 1: Overview of the CogEvolution framework.

Cognitive Adapter

The adapter serves as the perception entry point of the agent, designed to establish the foundational information field required for simulation operation. We constructed a multimodal feature extractor to extract feature vectors I_t from interaction logs from the CogMath-948 dataset into the embedding of Agent's Adapter, including:

1. Cognitive Semantic Features: Utilizing Sentence-BERT to extract cognitive features and the textual information of the interactive practice in the dataset, such as incorrect classifications and reflections.
2. Behavioral Operation Features: Including interaction types (reading, questioning, self-explanation) and response latency.

The perceptron Φ_{ICAP} maps input features to a probability distribution across the four ICAP levels. The specific mathematical implementation is an attention-based hierarchical probability mapping:

$$\hat{y}_{icap} = \text{Softmax}(\mathbf{W}_c \cdot \text{Encoder}(I_t) + \mathbf{b}_c) \quad (1)$$

Based on the probability distribution, we calculate the weighted continuous **Cognitive Evolution Rate** ω_L to achieve smooth state transitions:

$$\omega_L = \text{SoftPlus} \left(\sum_{i \in \{P, A, C, I\}} v_i \cdot \hat{y}_i \right) \quad (2)$$

where v_i represents the baseline gain coefficient for each level (e.g., Constructive corresponds to 1.5). ω_L serves as a key parameter determining the step size of subsequent cognitive state updates, thereby achieving precise quantification of learner cognitive investment.

Memory Retrieval

This module aims to simulate the psychological mechanisms of "connecting new and old knowledge" and "Assimilation" in human learning. We introduce Item Response Theory (IRT) to define deep structural similarity.

The system calculates a hybrid similarity between the current question q and memory $m \in \mathcal{M}$:

$$\text{Score}(m) = \alpha \cdot \text{Sim}_{sem}(q, m) + \beta \cdot \text{Sim}_{struct}(q, m) \quad (3)$$

where Sim_{struct} measures the isomorphism of questions in the cognitive ability space. We define it as the integral distance between the IRT Characteristic Curves (ICC) of two questions near the learner's current ability θ_t :

$$\text{Sim}_{struct}(e_i, e_j) \propto \exp \left(- \int_{\theta_t - \delta}^{\theta_t + \delta} |P_i(\theta) - P_j(\theta)| d\theta \right) \quad (4)$$

This formula ensures that two questions are judged as structurally similar only when they require similar levels of cognitive ability to solve. If a memory with high structural similarity is retrieved ($\text{Score} > \tau$), the agent directly invokes the old schema to solve the new problem, at which point the cognitive state remains stable. If structural similarity is low but the perceived ICAP engagement ω_L is high, the system determines that prior knowledge is insufficient to explain the new situation, generating a cognitive state C_t of "confusion" or "exploration," triggering subsequent evolutionary updates.

Evolutionary Cognitive State Update

This is the core engine of CogEvolution, designed to simulate the cognitive evolution process and achieve dynamic updates. We draw upon the "Mutation-Selection-Update" paradigm of evolutionary algorithms to couple learning behaviors with cognitive evolutionary paths in real-time.

Mutation Operator: Cognitive Hypothesis Generation (Mutation) The current knowledge structure K_t is treated as the parent individual. When cognitive conflict is perceived (input I_t cannot be fully explained by K_t), the system triggers a mutation operation. Utilizing the generative capability of LLMs as the **Mutation Operator** $\mathcal{G}_{mut}$, λ offspring hypotheses are generated based on K_t :

$$\mathcal{H} = \{h_1, h_2, \dots, h_\lambda\} = \mathcal{G}_{mut}(K_t, I_t, \sigma) \quad (5)$$

where σ is the mutation intensity, modulated by the current emotional state. These hypotheses represent the multiple possibilities generated by the learner when confused (e.g., "Is the formula wrong?" or "Is this a new question type?"), constituting a temporary population at the thought level.

Fitness Evaluation & Selection We define a fitness function $\mathcal{F}(h)$ to evaluate the quality of each hypothesis. To ensure the evolutionary process conforms to human cognitive laws, we introduce Vygotsky’s **Zone of Proximal Development (ZPD)** theory as an evolutionary constraint (Vygotsky, 1978). In the algorithm, ZPD is formalized as an effective search radius centered on the current knowledge structure K_t . The fitness function evaluates whether hypothesis h falls within this region:

$$\mathcal{F}(h) = \text{Consistency}(h, I_t) - \gamma \cdot \text{Distance}(h, K_t, \text{ZPD}) \quad (6)$$

Here, the first term measures logical self-consistency, and the second term penalizes "excessive leaps" or "stagnation" deviating from the ZPD. The system executes tournament selection to screen the hypothesis h^* with the highest fitness:

$$h^* = \underset{h \in \mathcal{H}}{\text{argmax}} \mathcal{F}(h) \quad (7)$$

Evolutionary Update & Inheritance The final step is to generate the next-generation knowledge structure K_{t+1} . We adopt a Differential Evolution approach to simulate continuous evolution on the local population, integrating the features of the winning hypothesis h^* into the parent with the ICAP evolution rate ω_L as the step size:

$$K_{t+1} \leftarrow K_t + \omega_L \cdot (h^* - K_t) \quad (8)$$

This formula indicates that the magnitude of knowledge evolution depends on the depth of cognitive engagement (ω_L): passive participation leads to minor mutations, while constructive participation allows for significant structural restructuring.

Decision and Verification

The agent generates final external behaviors based on the updated knowledge structure K_{t+1} . To support the precise quantification of learning states, this module outputs two types of information:

Behavioral Decision: Generates specific problem-solving steps or natural language responses D_{out} .

Mastery Confidence: To simulate the student’s anticipation of their own ability and support subsequent Knowledge Tracing assessments (such as AUC/RMSE calculation), the model outputs the predicted probability $\hat{p}_t$ that the current response is correct. We obtain this value utilizing LLM Logits or Self-Reflection Prompts:

$$\hat{p}_t = P(\text{Correct} | K_{t+1}, I_t) \in [0, 1] \quad (9)$$

Alignment Check: The system calculates the semantic alignment coefficient δ_{align} between cognitive state C_{t+1} and behavior D_{out} . If δ_{align} is below the threshold τ , indicating "cognitive dissonance" (e.g., lacking confidence but forcing an answer), the system marks this interaction step as low confidence:

$$\delta_{align} = \text{CosSim}(\text{Embed}(D_{out}), \text{Embed}(C_{t+1})) \quad (10)$$

Experimental Set Up

To systematically verify the effectiveness of CogEvolution in simulating real learner cognitive dynamics, we designed a series of comparative experiments aimed at answering three core research questions:

- **RQ1** Does CogEvolution outperform traditional static persona and knowledge tracing models in terms of ability prediction accuracy and mistake fidelity?
- **RQ2** Can CogEvolution reproduce the non-linear learning trajectories characteristic of human learners, specifically regarding conformity to the Power Law of Practice?
- **RQ3** What is the validity of the simulated cognitive evolution process, and how does each core module contribute to the overall performance?

Dataset and Baseline Models

Given the lack of suitable open datasets in the current research field, we constructed the **CogMath-948** dataset collected from real educational scenarios to facilitate in-depth research on fine-grained cognitive evolution. This dataset covers real learning data from 1,245 eighth-grade students over six months. For each student interaction, we collected rich information across the following four dimensions: Question and Response Information: Includes question stem, student-selected answer, and standard correct answer; Reflective Explanation: Textual explanations provided by students after answering, elucidating their reasoning process or causes of error, which is a key basis for inferring cognitive states. Cognitive Engagement Label: Cognitive depth labels derived based on the ICAP taxonomy; Fine-grained Misconception Classification: Specific error concept categories associated with incorrect answers.

Baseline Models and Parameter Settings

To comprehensively evaluate model performance, we selected three types of representative baseline models.

General LLMs including Llama 3 (70B) (Meta AI, 2024) and GPT-4o(OpenAI, 2025), representing high-performance base capabilities using standard Persona Prompting.

Reasoning-Enhanced Models such as Gemini 3 Pro Think(Google DeepMind, 2025), used to investigate the impact of long Chain-of-Thought on the simulation process.

Knowledge Tracing (KT)-augmented Agent (Discriminative Assisted Generative Baseline(Arana et al., 2025)), we reproduced the PEERS framework as a strong baseline. This model integrates classic Bayesian Knowledge Tracing (BKT)(Corbett & Anderson, 1995) with LLMs, utilizing external data-driven methods to predict student mastery probabilities and generate responses accordingly.

Using this as a comparison group aims to demonstrate the differences in simulation authenticity between "external data-driven (KT)" and the "internal cognitive evolution (CogEvolution)" proposed in this paper.

Metrics

This study adopts a two-stage evaluation system, covering ability simulation accuracy and cognitive evolutionary scientificity:

Task Performance:

- **AUC & RMSE:** Adopting recognized metrics in the field of knowledge tracing. AUC (Area Under Curve) measures the model’s ability to distinguish between "mastery" and "non-mastery," while RMSE (Root Mean Square Error) quantifies the deviation between predicted probabilities and actual response results.
- **Mistake Precision(Piech et al., 2015):** Based on established standards, this evaluates not only whether the agent answers incorrectly but more importantly whether it commits "real and representative errors." This metric calculates the consistency between agent-generated incorrect options and typical student Misconceptions.

Cognitive Dynamics:

- **Learning Curve Fitting (R_{LC}^2):** Based on relevant research(Newell & Rosenbloom, 1981), real human learning follows the Power Law of Practice. We calculate the goodness of fit R^2 between the agent’s performance improvement trajectory and the power function. An $R^2 > 0.9$ typically represents extremely high psychological validity of the simulation data.
- **Behavior-Cognition Alignment (Alignment):** Evaluates the logical self-consistency between the agent’s generated internal state (e.g., confidence/confusion) and its external performance (correct/incorrect).

Evaluation and Discussion

Table 2 presents the comprehensive performance of each model on CogMath-948. Experimental results reveal sig-

nificant differences in fidelity across different modeling paradigms.

Table 2: Comparison of Ability Simulation and Cognitive Dynamics Metrics for Different Models on CogMath-948

Model	Task Performance			Cognitive Dynamics	
	AUC $\uparrow$	RMSE $\downarrow$	Mistake Prec. $\uparrow$	R_{LC}^2 $\uparrow$	Align $\uparrow$
<i>Static Agents</i>					
Llama 3 (70B)	0.65	0.48	42.1%	0.45	0.72
GPT-4o	0.71	0.43	55.4%	0.52	0.78
Gemini 3 Pro Think	0.73	0.41	58.2%	0.61	0.82
<i>KT-augmented</i>					
PEERS (BKT+LLM)	0.82	0.35	64.5%	0.78	0.75
<i>Ours</i>					
CogEvolution	0.80	0.37	76.8%	0.92	0.91

Analysis of Task Performance Accuracy (RQ1)

In terms of mastery perception precision (AUC/RMSE), CogEvolution performs comparably to the knowledge tracing-based PEERS model (AUC: 0.80 vs 0.82), and both significantly outperform static persona models. This indicates that after introducing dynamic state update mechanisms, the agent can accurately capture fluctuations in student ability, similar to KT models. More importantly, CogEvolution achieves a significant advantage in **Mistake Precision** (76.8%). Traditional Static Agents (such as GPT-4o) tend to generate "hallucinatory errors" or "logical leap errors," while PEERS generates error content that still relies on the base model despite predicting error probabilities. In contrast, through ICAP perception and memory retrieval, CogEvolution can reproduce typical Misconceptions consistent with real student cognitive gaps, demonstrating high fidelity in micro-level behavioral simulation.

Learning Curve Fitting and Dynamics Analysis (RQ2)

To quantitatively answer RQ2—whether the learning trajectories generated by CogEvolution conform to the fundamental laws of human cognition—we conducted a rigorous **Learning Curve Fitting** analysis based on the *Power Law of Practice*.

Metric Calculation Protocol In educational psychology, the learning rate of human students typically follows a non-linear decay pattern, where performance improves rapidly during early practice stages and gradually plateaus. We formalize this trajectory using the power function model:

$$E(n) = A \cdot n^{-\alpha} + \epsilon \quad (11)$$

where $E(n)$ represents the error rate at the n -th practice opportunity, A is the initial error rate, and α denotes the learning rate decay parameter. To evaluate the fidelity of the agents, we fit the error rate sequences generated by each model to this power function. The evaluation metric, **Learning Curve Fitting Degree (R_{LC}^2)**, is defined as the coefficient of determination (R^2) between the agent’s simulated trajectory and the

ground truth human trajectory from the CogMath-948 dataset:

$$R_{LC}^2 = 1 - \frac{\sum_n (y_{human}^{(n)} - y_{agent}^{(n)})^2}{\sum_n (y_{human}^{(n)} - \bar{y}_{human})^2} \quad (12)$$

A higher R_{LC}^2 indicates that the agent’s learning dynamics more closely mirror the "fast-then-slow" nature of human skill acquisition.

Visualization and Result Discussion Figure 2 visualizes the error rate trajectories of different agents over 100 practice opportunities. The visualization provides compelling evidence for the cognitive validity of CogEvolution.

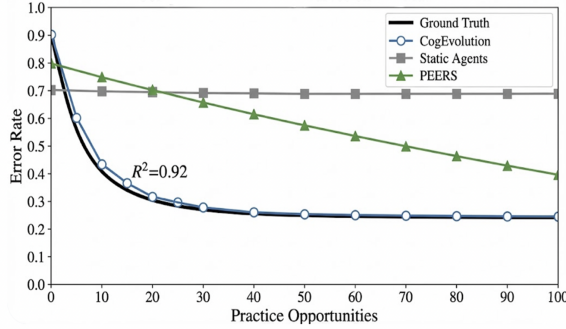


Figure 2: Comparison of Learning Curves on CogMath-948. The black solid line represents the ground truth student performance, which strictly follows the Power Law of Practice. Specifically, the error rate of exercises related to the same knowledge point decreases as the number of practice opportunities increases until it approaches a convergence value based on the difficulty of the questions. The closer the learning curve simulated by the agent is to the actual situation, the more accurately it reflects the learner’s knowledge level. This value is the computational indicator of learning curve fitting R^2 .

As observed in Figure 2, the **Static Agents** (Grey Square) exhibit a nearly flat trajectory ($R_{LC}^2 = 0.45$). This confirms that without a state update mechanism, standard LLMs rely solely on pre-trained knowledge, failing to manifest any "learning" effect from practice. The **PEERS** model (Green Triangle) shows a downward trend ($R_{LC}^2 = 0.78$), but its trajectory is overly linear. This suggests that while BKT-based probability transitions can simulate performance improvement, they struggle to capture the *diminishing returns* characteristic of cognitive saturation.

In sharp contrast, **CogEvolution** (Blue Circle) produces a curve that tightly aligns with the Ground Truth ($R_{LC}^2 = 0.92$). The trajectory displays a steep drop in the first 20 steps, followed by a gradual stabilization. This high-fidelity simulation is attributed to our evolutionary update mechanism: early "Cognitive Spikes" (caused by Mutation) lead to rapid schema restructuring (Assimilation), while later updates become more refined (Fine-tuning), perfectly replicating the transition from "novice" to "proficient" as predicted by the Power Law.

Ablation Study (RQ3)

To verify the contribution of each module, we conducted ablation experiments (Table 3).

- **w/o ICAP:** Removing cognitive depth perception renders the agent unable to distinguish between shallow and deep learning, causing R_{LC}^2 to drop sharply to 0.58, demonstrating the validity of cognitive features in simulating agent cognitive psychological validity.
- **w/o Meta-Ret:** Removing structured retrieval leads to a 12.3% decrease in Mistake Precision, indicating that knowledge assimilation is crucial for reproducing specific cognitive misconceptions.
- **w/o Evo-Update:** Removing the evolutionary update mechanism results in significantly reduced behavioral mistake fidelity and cognitive alignment (Align), causing the agent’s simulation capability to fall back to static persona modeling.

Table 3: Ablation Study Results for Core CogEvolution Modules

Variant	Mistake Prec.	R_{LC}^2	Align
CogEvolution (Full)	76.8%	0.92	0.91
w/o ICAP	68.5% (↓ 8.3)	0.58 (↓ 0.34)	0.73
w/o Meta-Ret	64.5% (↓ 12.3)	0.85 (↓ 0.07)	0.88
w/o Evo-Update	55.1% (↓ 21.7)	0.51 (↓ 0.41)	0.76

Conclusion

Addressing the pervasive issues of missing cognitive ability simulation and dynamic fluidity in generative educational agent research, this paper proposes the CogEvolution agent, which integrates an ICAP cognitive depth perceptron, structured memory retrieval, and a state update mechanism based on evolutionary strategies. Extensive experimental and simulation evaluations demonstrate that CogEvolution successfully realizes a paradigm shift from static behavioral simulation to dynamic cognitive evolution, marking a solid step toward constructing digital learners with "human-like cognition" and dynamic evolution.

Acknowledgments

This research was supported by the General Program of the National Natural Science Foundation of China (No. 62377024), titled *Discovering Interpretable Cognitive Evolution Paths Based on Causal Reasoning*, running from January 1, 2024 to December 31, 2027.

References

- Aher, G. V., Arriaga, R. I., & Kalai, A. T. (2023). Using large language models to simulate multiple humans and replicate human subject studies. *International Conference on Machine Learning*, 337–371.

- Arana, J. M., Carandang, K. A. M., Casin, E. R., Alis, C., Tan, D. S., Legara, E. F., & Monterola, C. (2025, July). Foundations of PEERS: Assessing LLM role performance in educational simulations. In J. Zhao, M. Wang, & Z. Liu (Eds.), *Proceedings of the 63rd annual meeting of the association for computational linguistics (volume 4: Student research workshop)* (pp. 908–918). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.acl-srw.66>
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243.
- Chuang, Y.-S., Suresh, S., Harlalka, N., Goyal, A., Hawkins, R., Yang, S., Shah, D., Hu, J., & Rogers, T. T. (2024). The wisdom of partisan crowds: Comparing collective intelligence in humans and LLM-based agents. <https://arxiv.org/abs/2311.09665>
- Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *Proceedings of the International Conference on User Modeling*, 1–23.
- Dai, C., Hu, J., Shi, H., Li, Z., Yang, X., & Wang, M. (2025). Psyche-r1: Towards reliable psychological llms through unified empathy, expertise, and reasoning. *arXiv preprint arXiv:2508.10848*.
- Dai, C., Shen, Y., Hu, J., Gao, Z., Li, J., Jiang, Y., Wang, Y., Liu, L., & Ge, Z. (2026). Tears or cheers? benchmarking llms via culturally elicited distinct affective responses. <https://arxiv.org/abs/2601.13024>
- Gao, H.-a., Geng, J., Hua, W., Hu, M., Juan, X., Liu, H., Liu, S., Qiu, J., Qi, X., Wu, Y., Wang, H., Xiao, H., Zhou, Y., Zhang, S., Zhang, J., Xiang, J., Fang, Y., Zhao, Q., Liu, D., . . . Wang, M. (2026). A survey of self-evolving agents: What, when, how, and where to evolve on the path to artificial super intelligence. <https://arxiv.org/abs/2507.21046>
- Google DeepMind. (2025). Gemini 3 Pro (multimodal large language model).
- Li, Y., Wang, S., Li, J., Xu, Y., Tang, K., Li, J., Liu, H., & Tang, C. (2025). EvoAgents: A cognitive-driven framework for personality evolution in generative agent society. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 47. <https://escholarship.org/uc/item/7s3173zf>
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.
- Luo, J., Zhang, W., Yuan, Y., Zhao, Y., Yang, J., Gu, Y., Wu, B., Chen, B., Qiao, Z., Long, Q., Tu, R., Luo, X., Ju, W., Xiao, Z., Wang, Y., Xiao, M., Liu, C., Yuan, J., Zhang, S., . . . Zhang, M. (2025). Large language model agent: A survey on methodology, applications and challenges. <https://arxiv.org/abs/2503.21460>
- Lv, R., Liu, Q., Gao, W., Zhang, H., Lu, J., & Zhu, L. (2025). GenAL: Generative agent for adaptive learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(1), 577–585.
- Mandyam, A., Gottesman, O., Andaz, S., & Foster, D. (2025). TutorTest: Evaluating language model-based tutoring policies using surrogate tasks. *NeurIPS 2025 Workshop on Scaling Environments for Agents*. <https://openreview.net/pdf?id=UoJZv2YfDX>
- Markel, J. M., Gorgun, S. G., et al. (2023). GPTeach: Interactive TA training with GPT-based students. *Proceedings of the Tenth ACM Conference on Learning @ Scale*, 226–230. <https://doi.org/10.1145/3573051.3596191>
- Meta AI. (2024). Llama 3 (large language model).
- Murray, R. C., Ritter, S., Nixon, T., Schwiebert, R., Hausmann, R. G. M., Towle, B., Fancali, S. E., & Vuong, A. (2013). Revealing the learning in learning curves. *International Conference on Artificial Intelligence in Education*, 473–482.
- Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Lawrence Erlbaum Associates.
- OpenAI. (2025). GPT-4o (large multimodal model).
- Park, J. S., O’Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 1–22.
- Piaget, J. (1976). Piaget’s theory. In *Piaget and his school: A reader in developmental psychology* (pp. 11–23). Springer.
- Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. *Advances in Neural Information Processing Systems*, 28.
- Sumers, T., Yao, S., Narasimhan, K. R., & Griffiths, T. L. (2024). Cognitive architectures for language agents [Survey Certification, Featured Certification]. *Transactions on Machine Learning Research*. <https://openreview.net/forum?id=1i6ZCvflQJ>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., et al. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6), 186345.
- Wu, T., Chen, J., Lin, W., Li, M., Zhu, Y., Li, A., Kuang, K., & Wu, F. (2025, July). Embracing imperfection: Simulating students with diverse cognitive levels using LLM-based agents. In W. Che, J. Nabende, E. Shutova, & M. T. Pilehvar (Eds.), *Proceedings of the 63rd annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 9887–9908). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.acl-long.488>
- Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., et al. (2025). The rise and potential of large language model based agents: A survey. *Science China Information Sciences*, 68(2), 121101.

- Xu, S., Zhang, X., & Qin, L. (2024). EduAgent: Generative student agents in learning. *arXiv preprint arXiv:2404.07963*. <https://arxiv.org/abs/2404.07963>
- Yuan, S., Zhang, H., Wang, L., et al. (2024). EvoAgent: Large language models as evolutionary agents. *arXiv preprint arXiv:2402.11223*. <https://arxiv.org/abs/2402.11223>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64–70.