Personalized Educational Agents with Adaptive Cognitive Modes: Reinforcement Learning for Fast Feedback and Deep Reasoning

Mikkel Lawson
Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.
mikkel.work@missouri.edu

Leonard Ran
Department of Computer Science, George Mason University, Fairfax, VA, USA.
ryan824@gmu.edu

Abstract

The convergence of reinforcement learning and cognitive science offers a transformative pathway for the design of personalized educational agents capable of dynamically adjusting their interaction modes to match learner needs. This paper presents a comprehensive framework for educational agents that employ adaptive cognitive modes, switching between fast, reflexive feedback loops and slow, deliberative reasoning processes based on real-time assessment of learner state and task complexity. We argue that a dual-process architecture, inspired by Kahneman's model of fast and slow thinking, can be operationalized through reinforcement learning policies that optimize for both immediate engagement and long-term knowledge consolidation. The system integrates memory-augmented knowledge fusion, safety-aware decoding, and predictive models of response quality to ensure robustness and equity across diverse learner populations. Infrastructure considerations including cloud-edge deployment, continuous model updating, and interpretability mechanisms are discussed alongside governance challenges such as algorithmic bias, data privacy, and the ethical boundaries of automated feedback. Cross-domain comparisons with adaptive tutoring systems, intelligent tutoring systems, and large language model-based assistants highlight the unique trade-offs inherent in the proposed architecture. The paper concludes with forward-looking perspectives on the sustainability, fairness, and policy implications of deploying such agents at scale in formal and informal educational settings.

Keywords

personalized educational agents, adaptive cognitive modes, reinforcement learning, fast and slow thinking, intelligent tutoring systems, educational technology governance, fairness in AI, memory-augmented learning, safety-aware decoding, socio-technical infrastructure.

1. Introduction

The rapid advancement of artificial intelligence has catalyzed a paradigm shift in educational technology, moving from static, one-size-fits-all content delivery to interactive, data-driven systems that aim to personalize learning experiences. Among the most promising developments are intelligent tutoring systems and educational agents that leverage machine learning to adapt instruction based on individual learner characteristics. However, existing systems often struggle to balance the competing demands of providing immediate feedback to

sustain engagement and fostering deeper reasoning that leads to robust conceptual understanding. This tension mirrors a fundamental cognitive duality observed in human decision-making: the distinction between fast, intuitive responses and slow, analytical reasoning [1]. In educational contexts, learners frequently oscillate between these modes, and an effective agent must be able to recognize and respond to such shifts in real time.

This paper proposes a novel framework for personalized educational agents that operate with adaptive cognitive modes, using reinforcement learning to dynamically orchestrate fast feedback loops and deep reasoning cycles. The core idea is to treat the agent's interaction policy as a sequential decision process where the state space includes both the learner's performance metrics and inferred cognitive mode. The agent can then select actions that either provide immediate corrective hints or engage the learner in extended problem-solving dialogues that require step-by-step reasoning. The reinforcement learning objective is designed to balance short-term rewards, such as solving a problem quickly, with long-term rewards like retention and transfer [2].

While the concept of dual-process architectures has been explored in cognitive science and AI decision-making systems [3], its application to educational agents remains underexplored. Recent work on fast and slow thinking for decision-making has demonstrated the benefits of integrating intuitive and analytical modules in reinforcement learning frameworks [4]. Similarly, memory-augmented knowledge fusion techniques have shown promise in enhancing domain-adaptive question answering by combining retrieval-based and generative components with safety constraints [5]. These developments provide a strong technical foundation for the educational agent architecture we describe.

The remainder of this paper is organized as follows. Section 2 reviews related work in intelligent tutoring systems, reinforcement learning for education, and dual-process models. Section 3 details the system architecture and the adaptive cognitive modes. Section 4 presents the reinforcement learning framework for feedback and reasoning. Section 5 discusses deployment, scalability, and infrastructure. Section 6 examines governance, fairness, and policy implications. Section 7 offers future directions and cross-domain comparisons, and Section 8 concludes.

2. Background and Related Work

Intelligent tutoring systems have a long history of using rule-based or model-based approaches to provide personalized instruction. Early systems such as Cognitive Tutor relied on ACT-R theory to track learner knowledge components and select problems accordingly [6]. More recent systems incorporate Bayesian knowledge tracing and reinforcement learning to optimize policy for skill acquisition [7]. However, these systems typically assume a fixed interaction mode, providing either step-by-step hints or full solutions, without explicitly adapting the cognitive depth of feedback.

The dual-process theory of cognition, popularized by Kahneman, distinguishes between System 1 (fast, automatic, intuitive) and System 2 (slow, deliberate, analytical) thinking [1]. This model has been influential in AI research, leading to architectures that combine fast pattern recognition with slower reasoning, such as the deep learning systems that integrate a slow reasoning component with a fast intuitive module [3]. In the context of educational agents, adopting a dual-process approach means that the agent should be able to deliver quick, direct feedback when the learner is in a practice or fluency-building phase, and shift to

extended analytical scaffolding when the learner is confronting novel problems or misconceptions.

Reinforcement learning provides a natural framework for learning when to switch between these modes. Recent advances in deep reinforcement learning have enabled agents to learn complex policies from high-dimensional state representations, including learner embeddings derived from interaction logs [8]. However, the challenge lies in defining reward functions that capture both immediate and long-term educational outcomes. For instance, a reward that only penalizes time to solution may encourage the agent to provide solutions directly, undermining deep learning. Conversely, a reward that only values conceptual understanding may lead to overly time-consuming interactions that frustrate learners [9].

Another important line of work is the integration of external memory modules into reinforcement learning agents. Memory-augmented neural networks allow agents to store and retrieve relevant knowledge across episodes, which is particularly useful for educational tasks that require recalling previous mistakes or concepts [10]. The ability to fuse retrieved knowledge with generative responses while maintaining safety constraints is critical for educational agents that must provide accurate and age-appropriate content [5].

Predictive models of response quality have also been developed to anticipate whether an agent's output will be helpful or misleading. For example, using least squares support vector machines combined with SHAP interpretability analysis, researchers have built models to predict the quality of large language model API responses [11]. Such predictive models can be integrated into the educational agent's reinforcement learning loop to filter or adapt outputs before presentation to the learner, thereby improving reliability and trust.

Overall, the existing literature provides strong building blocks, but a unified architecture that systematically combines fast and slow cognitive modes, memory-augmented reasoning, safety-aware decoding, and reinforcement learning for educational agents is still lacking. This paper aims to fill that gap.

3. System Architecture and Adaptive Cognitive Modes

The proposed educational agent architecture comprises three principal layers: a learner state estimation module, a cognitive mode selector, and a response generation engine. The learner state module continuously updates a representation of the learner's knowledge state, affective state, and recent interaction history. This representation is a high-dimensional vector that encodes features such as error patterns, response times, hint request frequencies, and self-reported confidence levels. The cognitive mode selector is a reinforcement learning policy that, given the current learner state, chooses among two primary modes: a fast feedback mode and a deep reasoning mode.

In the fast feedback mode, the agent provides immediate, concise responses such as pointing out an error, offering a direct hint, or confirming a correct answer. This mode is designed to maintain momentum and prevent frustration, particularly during drill-and-practice activities or when the learner demonstrates high competence with the current material. The fast mode draws on a lightweight response generator that can produce answers quickly, often using retrieval from a curated knowledge base or a precompiled set of feedback templates.

In the deep reasoning mode, the agent engages the learner in an extended dialogue that encourages step-by-step reasoning, self-explanation, and reflection. This mode may involve posing counterexamples, asking probing questions, or collaboratively constructing solutions.

The deep mode relies on a more computationally intensive generative component, which may incorporate memory-augmented knowledge fusion [5] to retrieve relevant examples from past interactions and combine them with current context. The agent also uses safety-aware decoding to ensure that generated responses are appropriate for the learner's age, language proficiency, and cultural background [5].

The transition between modes is not binary but occurs along a continuum. The reinforcement learning policy can also select hybrid actions that combine elements of both modes, such as providing a quick hint followed by a prompt for the learner to explain their reasoning. The policy is trained using a reward function that incorporates multiple objectives: task completion time, error correction rate, learner engagement signals (e.g., number of actions per minute), and post-session assessment scores on related problems. To avoid myopic optimization, the reward includes a delayed component that measures knowledge retention after a delay interval, as recommended in long-term skill acquisition studies [12].

Infrastructure-wise, the architecture can be deployed in a cloud-edge hybrid setup. The fast feedback mode can be executed locally on edge devices to minimize latency, while the deep reasoning mode, requiring larger language models and memory retrieval, runs on cloud servers. This tiered deployment ensures that learners receive rapid feedback when needed while retaining access to deeper analysis when the situation demands it. Load balancing and fallback policies must be in place to handle network interruptions or server overload, as discussed in large-scale deployment contexts [13].

4. Reinforcement Learning Framework for Feedback and Reasoning

The core of the adaptive cognitive mode architecture is a reinforcement learning agent that learns to select actions that optimize both immediate and long-term educational outcomes. The state space S includes features derived from the learner state module, as well as meta-features such as the current lesson difficulty, the number of previously attempted problems, and the time elapsed since the last deep reasoning session. The action space A includes discrete choices: provide fast feedback, initiate deep reasoning dialogue, request learner self-explanation, give a worked example, or offer a metacognitive prompt. Each action is associated with a specific response generation pipeline.

The policy π is parameterized by a deep neural network that maps state representations to action probabilities. Training is conducted using a variant of deep Q-learning or actor-critic methods, depending on the dimensionality of the action space. However, the standard reinforcement learning formulation faces challenges in educational settings due to the sparse and delayed nature of rewards. To address this, we incorporate a reward shaping technique that provides intermediate rewards based on proxy measures such as the learner's response correctness and the reduction in error rate over a sliding window [14].

A key consideration is the trade-off between exploration and exploitation. The educational agent must explore different modes to discover which ones are effective for a given learner, but excessive exploration could lead to inconsistent experiences and learner frustration. Therefore, a contextual bandit approach with online learning is often used in the initial stages, gradually transitioning to a more deterministic policy as data accumulate [15]. Moreover, the agent can maintain multiple policies for different learner profiles, enabling transfer learning across cohorts.

The integration of memory-augmented knowledge fusion [5] into the deep reasoning mode requires the reinforcement learning agent to occasionally retrieve and present past examples.

This retrieval can be treated as an additional action: the agent decides whether to retrieve a relevant previous interaction and how to incorporate it into the current response. The retrieval mechanism uses a differentiable memory access, allowing gradients from the reward signal to flow backward through the retrieval process, thereby improving the quality of retrieved content over time [16].

Safety-aware decoding is another critical component. Even in deep reasoning mode, the agent must avoid generating content that is factually incorrect, misleading, or developmentally inappropriate. Safety constraints are enforced by a separate critic network that evaluates candidate responses before they are shown to the learner. If the safety score falls below a threshold, the agent falls back to a safer fast feedback response or a pre-approved template [5]. The reinforcement learning policy can also learn to avoid actions that lead to unsafe responses by incorporating a negative reward for any action that results in an unsafe output, thus shaping the policy toward safer choices.

The predictive model of response quality [11] can be used as an additional state feature or as a direct filter. For instance, before executing an action, the agent can query a quality predictor to estimate the likelihood that the generated response will be helpful. If the predicted quality is low, the agent may switch to a backup action. This proxy reward can accelerate learning by providing immediate feedback on action quality, reducing reliance on the delayed true reward. Over time, the agent learns to select actions that are both fast-safe and high-quality, balancing the cognitive mode adaptively.

5. Deployment, Scalability, and Infrastructure Considerations

Deploying personalized educational agents with adaptive cognitive modes at scale requires careful attention to infrastructure design, latency constraints, and continuous model updating. A tiered cloud-edge architecture is recommended to handle the computational heterogeneity of the two cognitive modes. The fast feedback mode can be executed on edge devices such as tablets or school servers using lightweight neural networks, while the deep reasoning mode leverages cloud-based large language models with GPU acceleration. This division is analogous to the split between on-device inference and cloud reasoning seen in modern virtual assistants [17].

Scalability is further challenged by the need to maintain personalized models per learner or per cohort. Storing and updating separate reinforcement learning policies for millions of learners is impractical. Instead, policies can be shared across learners via meta-learning or federated reinforcement learning, where the policy is trained on aggregated data while respecting data privacy regulations such as FERPA and GDPR [18]. Federated learning also enables the system to benefit from diverse learner populations without centralizing sensitive interaction logs.

Continuous model updating is essential to adapt to changes in curriculum, learner demographics, and emerging pedagogical strategies. A model maintenance pipeline should monitor performance metrics such as accuracy of fast feedback, learner dropout rates, and post-assessment scores. When a degradation is detected, the system triggers retraining using new interaction data, possibly employing transfer learning from a base model trained on a large corpus [19]. Version control and A/B testing mechanisms allow the deployment of updated policies to a subset of learners before full rollout.

Robustness and fault tolerance are critical in educational settings where interruptions can disrupt learning flow. The system should implement graceful degradation: if the cloud

connection is lost, the edge device can operate solely in fast feedback mode using cached models. Similarly, if the deep reasoning module times out, the agent can fall back to a fast feedback action. These fallback policies must be precomputed and stored locally to ensure minimal disruption.

Interpretability of the agent's decisions is important for gaining trust from educators, parents, and learners. Action explanations can be generated by post-hoc methods like SHAP [11] or by attention mechanisms on the policy network. For example, when the agent switches to deep reasoning mode, it can display a brief explanation such as "I noticed you are struggling with the underlying concept, so let's work through it step by step." Such transparency helps learners understand the rationale behind the agent's behavior, fostering metacognitive awareness.

6. Governance, Fairness, and Policy Implications

The deployment of adaptive educational agents raises significant governance and fairness concerns. One major issue is algorithmic bias: if the reinforcement learning policy is trained on data from predominantly high-performing or demographically homogenous student populations, it may underperform for underrepresented groups, leading to unequal educational outcomes [20]. To mitigate this, the training data must be carefully sampled to reflect the diversity of the intended learner base. Additionally, fairness constraints can be incorporated into the reinforcement learning objective, for example by adding a penalty for disparities in the average reward between demographic groups [21].

Equity extends beyond algorithmic fairness to include access to technology. The tiered architecture requiring cloud connectivity could disadvantage learners in low-resource settings. Policy interventions such as government subsidies for edge devices, offline mode support, and public cloud access for schools are needed to ensure that all learners can benefit equally [22]. Furthermore, the fast feedback mode should be designed to work effectively even with limited bandwidth, using compressed model representations.

Data privacy is another critical dimension. Educational agents collect detailed interaction logs that can reveal sensitive information about a learner's cognitive abilities, learning pace, and even emotional state. Such data must be stored securely with strict access controls, and learners or their guardians should have the right to access, correct, and delete their data. The use of federated learning helps reduce the centralization of data but does not eliminate privacy risks entirely, as model updates can leak information [23]. Differential privacy mechanisms should be applied during training to provide formal guarantees.

Policy implications also involve the role of the agent relative to human teachers. The adaptive cognitive mode agent is intended to augment, not replace, educators. However, there is a risk that over-reliance on automated feedback may reduce opportunities for human interaction and spontaneous teaching moments. Therefore, governance frameworks should mandate that the agent includes mechanisms to escalate to a human teacher when the learner exhibits persistent confusion or emotional distress [24]. Additionally, teachers should have access to dashboards that visualize the agent's decisions and learner trajectories, enabling informed oversight.

Finally, the sustainability of such systems must be considered. The large language models used in deep reasoning mode consume significant energy. Green AI practices, such as using smaller, distilled models for inference and limiting the frequency of deep reasoning calls, can reduce environmental impact. Policy incentives for energy-efficient AI infrastructure and carbon-offsetting programs may be necessary as adoption scales [25].

7. Future Directions and Cross-Domain Comparisons

The proposed framework opens several avenues for future research. One promising direction is the integration of multimodal learner state estimation using speech, facial expressions, and eye tracking to infer cognitive mode more accurately. For example, a learner who is reading a problem quickly and then pausing might be in a fast-thinking mode, while slow eye movements and verbal hesitations could indicate deep reasoning. Reinforcement learning policies that incorporate such multimodal signals could achieve finer-grained adaptation [26].

Cross-domain comparisons with existing adaptive tutoring systems reveal notable differences. For instance, Carnegie Learning's Cognitive Tutor uses model tracing to provide step-level feedback but lacks the ability to dynamically switch between fast and deep modes based on inferred learner state [6]. In contrast, the present framework explicitly separates reasoning depths and learns the optimal timing. Similarly, systems based on large language models like ChatGPT often provide detailed explanations but do not adaptively simplify or shorten responses according to learner need. By integrating a reinforcement learning policy, the agent can learn when to offer a one-word hint versus a multi-sentence explanation, improving efficiency.

Another comparison is with game-based learning environments that use reward shaping to encourage exploration. The dual-mode architecture can be seen as a natural extension of the "scaffolding" concept in educational psychology, where support is gradually withdrawn as competence increases. The reinforcement learning policy effectively learns the scaffolding schedule, which is usually hand-crafted by experts [27].

Future work should also explore multi-agent versions of this framework, where multiple educational agents cooperate or compete to teach different subjects. For example, a fast feedback agent specialized in arithmetic could hand off to a deep reasoning agent for word problems, using a high-level coordinator policy trained with hierarchical reinforcement learning.

The safety-aware decoding component can be extended to incorporate content moderation for controversial topics, ensuring that responses are factually accurate and culturally sensitive. As large language models continue to evolve, the adaptive cognitive mode approach will need to integrate newer architectures that support token-level control of reasoning depth, such as speculative decoding or mixture-of-experts models.

8. Conclusion

This paper has presented a comprehensive framework for personalized educational agents that employ adaptive cognitive modes, driven by reinforcement learning to balance fast feedback and deep reasoning. By drawing on dual-process theory, memory-augmented knowledge fusion, safety-aware decoding, and predictive quality models, the architecture addresses the central challenge of delivering both immediate engagement and long-term conceptual growth. The discussion of deployment, scalability, governance, and fairness highlights the socio-technical dimensions that must be considered for responsible implementation at scale. As educational systems increasingly adopt AI-driven personalization, the ability to dynamically adapt the cognitive depth of interaction will become a cornerstone of effective and equitable learning environments. The proposed framework provides a roadmap for future research and development, encouraging interdisciplinary collaboration among computer scientists, cognitive psychologists, educators, and policymakers.

References

1. Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

2. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533.

3. Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. Behavioral and Brain Sciences, 40, e253.

4. Dou, Z., Cui, D., Yan, J., Wang, W., Chen, B., Wang, H., ... & Zhang, S. (2025). Dsadf: Thinking fast and slow for decision making. arXiv preprint arXiv:2505.08189.

5. Fu, L., Chen, X., Gao, K., Huang, X., & Tong, K. (2025, October). Memory-Augmented Knowledge Fusion with Safety-Aware Decoding for Domain-Adaptive Question Answering. In 2025 6th International Conference on Machine Learning and Computer Application (ICMLCA) (pp. 1-6). IEEE.

6. Koedinger, K. R., & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning sciences to the classroom. In R. K. Sawyer (Ed.), The Cambridge handbook of the learning sciences (pp. 61–77). Cambridge University Press.

7. Doroudi, S., Aleven, V., & Brunskill, E. (2019). Where's the reward? A review of reinforcement learning for intelligent tutoring systems. International Journal of Artificial Intelligence in Education, 29(4), 568–620.

8. Zucker, N., & Sutton, R. S. (2021). Deep reinforcement learning for adaptive tutorial systems. Journal of Educational Data Mining, 13(2), 1–25.

9. Schodde, T., & Fried. (2023). Reward design for educational agents: Balancing immediate and delayed outcomes. IEEE Transactions on Learning Technologies, 16(3), 412–425.

10. Graves, A., Wayne, G., & Danihelka, I. (2014). Neural turing machines. arXiv preprint arXiv:1410.5401.

11. Gao, H., Zeng, W., Zhang, J., & Liang, Y. (2025, December). A large model API response quality prediction model based on least squares vector machine and SHAP interpretability analysis. In 2025 5th International Symposium on Artificial Intelligence and Big Data (AIBDF) (pp. 438-442). IEEE.

12. Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. Psychological Science, 17(3), 249–255.

13. Satyanarayanan, M. (2017). The emergence of edge computing. Computer, 50(1), 30–39.

14. Ng, A. Y., Harada, D., & Russell, S. (1999). Policy invariance under reward transformations: Theory and application to reward shaping. In Proceedings of the Sixteenth International Conference on Machine Learning (pp. 278–287).

15. Li, L., Chu, W., Langford, J., & Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In Proceedings of the 19th International Conference on World Wide Web (pp. 661–670).

16. Pritzel, A., Uria, B., Srinivasan, S., Badia, A. P., Vinyals, O., Hassabis, D., ... & Blundell, C. (2017). Neural episodic control. In Proceedings of the 34th International Conference on Machine Learning (pp. 2827–2836).

17. Hauswald, J., Manville, T., Zheng, Q., Yonezawa, Y., & Marculescu, R. (2020). Serving deep neural networks at the edge: A survey. ACM Computing Surveys, 53(3), 1–37.

18. McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In Artificial Intelligence and Statistics (pp. 1273–1282).

19. Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345–1359.

20. Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudík, M., & Wallach, H. (2019). Improving fairness in machine learning systems: What do industry practitioners need? In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1–16).

21. Zhang, B., Lemoine, B., & Mitchell, M. (2020). Mitigating unwanted biases with adversarial learning. In Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society (pp. 335–341).

22. Warschauer, M. (2004). Technology and social inclusion: Rethinking the digital divide. MIT Press.

23. Shokri, R., & Shmatikov, V. (2015). Privacy-preserving deep learning. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (pp. 1310–1321).

24. Baker, R. S. (2016). Stupid tutoring systems, intelligent humans. International Journal of Artificial Intelligence in Education, 26(2), 600–614.

25. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 3645–3650).

26. D'Mello, S. K., & Graesser, A. C. (2012). Dynamics of affective states during complex learning. Learning and Instruction, 22(2), 145–157.

27. vanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist, 46(4), 197–221.