Fuzzmonte-rag: A hybrid fuzzy logic and monte carlo approach for personalized learning optimisation in llm-assisted it education

Wenhao Yan[1*] Chaofeng Zhang[2] Kazuo Yamada[3]

[1]Sophia University
[2]Advanced Institute of Industrial Technology
[3]SoftUsing Co., Ltd.
[*]Corresponding author: Wenhao Yan, wh_yan@eagle.sophia.ac.jp

Abstract Recent advancements in AI-driven education have led to widespread adoption of large language models for personalized learning. While retrieval-augmented generation enhances LLMs' ability to provide accurate and relevant content, current systems face significant challenges in handling learning uncertainties and optimizing learning paths, especially for learners with diverse backgrounds. In particular, existing approaches often treat learning capabilities and content difficulty as discrete states, failing to capture the continuous nature of learning processes and struggling to adapt to individual learning patterns. In this paper, we present FuzzMonte-RAG, a novel framework that enhances RAG with fuzzy logic and Monte Carlo methods for personalized IT education. Our approach incorporates fuzzy logic for modeling learning states and Monte Carlo simulation for optimizing learning paths. The framework has been deployed at a real-world IT business company for technical staff training with varying educational and professional backgrounds. Evaluation results shows that over 80% of the participants reported improved learning efficiency compared to traditional methods, with particularly positive feedback on adaptive difficulty adjustment and learning path coherence.

Keywords e-learning; fuzzy modeling; path optimization; retrieval-augmented generation

1 Introduction

In recent years, the fields of e-learning and educational methodology have faced significant revolution due to the explosive advancement of various technologies in the realm of Information Technology (IT)[1]. Specifically, technologies such as Augmented Reality (AR), Virtual Reality (VR), and related innovations have fundamentally altered traditional face-to-face interaction paradigms[2][3][4], ushering in a new era of digital engagement. Moreover, various mobile learning applications have liberated education from the confines of classrooms and libraries, which also empower individuals to create personalized learning paths[5]. Users can find courses that match their needs, interests, and ways of learning. The traditional boundaries of time and space in education have been dissolved, allowing for a truly flexible and adaptive learning ecosystem[6].

While numerous technological advancements have reshaped our digital landscape, Artificial Intelligence (AI) indisputably stands as the pinnacle of innovation, captivating both public imagination and professional discourse. Its transformative potential across myriad sectors has positioned AI as the keystone of our technological future. Traditional approaches to student or trainee tutoring frequently encounter inherent limitations, including constrained accessibility, variability in pedagogical efficacy, and obstacles to widespread implementation. In light of these persistent challenges, Large Language Models (LLMs) represent a quantum leap in educational technology, bridging the gap between the aspirational goals of personalized instruction and the practical limitations of human-centric tutoring systems[7]. By leveraging their ability to understand context, generate human-like responses, and adapt to individual learning styles, LLMs are uniquely positioned to address challenges like availability, consistency, and scalability that have historically constrained the efficacy of traditional tutoring approaches[8].

However, relying on only general LLMs (like ChatGPT) for skill-training or education purposes may face several significant drawbacks. These tools can provide inaccurate information and aren't able to offer guidance tailored to specific course content[9][10]. Without extra customization, those general LLMs can hardly meet the unique requirements of individual courses or students.

In order to harness AI's potential and improve user learning experiences, researchers and related workers in educational industries explore the application in the field of Retrieval Augmented Generation (RAG) in LLMs [11]. RAG is an AI technique that enhances language models with external knowledge, which combines the power of LLMs and the ability to access specific and latest information. After retrieving relevant data from a knowledge base, AI can provide more accurate, current, and contextually relevant responses with the guidance of external information, especially in domains requiring specialized or frequently updated knowledge [12].

While RAG has demonstrated significant potential in educational LLMs, it is not without limitations. RAG excels at retrieving and generating relevant educational content, but it faces challenges in handling the inherent uncertainties in learner abilities and generating truly optimal learning paths. Actually, education is inherently characterized by uncertainty. Factors such as learners' ability levels, learning styles, and the suitability of content are not precise and quantifiable values but rather fuzzy concepts [13]. Furthermore, the selection of a reasonable learning path involves exploring a vast space of possibilities. For example, in a comprehensive Python programming course, there are numerous knowledge points to cover, such as programming language basics, data structures, object-oriented programming, web development, and even machine learning theories, in which most of topics are not independent. More specifically, when explaining why NumPy is more efficient than native Python for numerical operations, an LLM should clarify that this efficiency stems from NumPy's use of contiguous memory blocks and optimized C implementations. Similarly, when discussing why sequential structures often outperform linked lists, the LLM should elucidate concepts like cache locality and memory access patterns. By exposing these reasoning paths, LLMs can help students understand not just the "what" but the critical "why" behind important computing concepts, fostering deeper comprehension and analytical thinking skills.

The inherent uncertainties of users in the learning process and the importance of clear reasoning paths in educational LLMs underscore the need for more sophisticated approaches in AI-driven education. While traditional methods struggle to capture the nuanced nature of learning and the complexity of optimal content delivery, in this paper, we offer a promising solution with fuzzy modeling and Monte Carlo searching approach. Fuzzy logic provides a framework to handle the imprecise nature of learner abilities, content difficulty, and topic relevance, allowing for more nuanced modeling of educational concepts. For instance, a learner's understanding of a certain skill or knowledge

Received: 2024-10-30; Published online: 2025-02-28

might be better represented as ”mostly proficient” rather than a binary ”understands” or ”doesn’t understand”. This fuzzy representation allows for more accurate learner modeling and adaptive content delivery. Complementing this, Monte Carlo methods excel at exploring vast possibility spaces, crucial for generating and evaluating potential learning paths. By simulating numerous learning scenarios, Monte Carlo techniques can help identify optimal sequences of topics and exercises that maximize learning outcomes while accounting for individual learner characteristics. For example, when deciding whether to introduce NumPy before or after certain native Python concepts, a Monte Carlo simulation could evaluate thousands of potential curriculum structures, considering factors like prerequisite knowledge, learning speed, long-term retention, and the most important, the learners themselves. Our proposed method can enhance the LLM’s ability to provide clear reasoning paths by simulating various explanation strategies and selecting the most effective ones based on the learner’s profile. By integrating fuzzy logic for uncertainty handling and Monte Carlo methods for generation optimization, educational LLMs can offer more personalized, adaptive, and effective learning experiences, bridging the gap between the complexity of human learning and the capabilities of AI-driven educational systems.

Our main contributions can be summarised as follows:

- We propose FuzzMonte-RAG, an innovative approach integrating fuzzy logic and Monte Carlo methods into educational LLMs.
- We utilize fuzzy logic to better handle the inherent uncertainties in the learning process, such as learner abilities and content difficulty, which are inherently fuzzy concepts.
- We apply Monte Carlo methods to explore and evaluate a vast number of potential inferring paths, thereby generating optimal personalized learning sequences.
- Our method enhances the LLM’s capability to provide clear reasoning paths through this approach, facilitating students’ understanding of the reason behind key concepts.
- Our works offer a method for delivering more personalized, adaptive, and effective learning experiences, bridging the gap between the complexity of human learning and the capabilities of AI-driven educational systems.

Literature review

As is widely known to all, recent years have witnessed a revolutionary transformation in educational methodologies, primarily driven by rapid advancements in various Information Technologies [14]. This transformation has fundamentally reshaped traditional learning paradigms through various technological innovations. Notably, emerging technologies such as AR and VR have significantly altered conventional face-to-face interaction models in education [3]. Another particularly significant development has been the proliferation of mobile learning applications, which have effectively liberated education from traditional spatial constraints. These applications have enabled learners to transcend the physical boundaries of classrooms and libraries, fostering the creation of personalized learning pathways [15] [16]. This shift has fundamentally altered the temporal and spatial dimensions of education, establishing a more flexible and adaptive learning ecosystem. As for e-learning development in Japan, there are also a quantity of works deserving our attention. K. Ota and M. Dong et al. have proposed a novel approach by fusing a real world and virtual space to let a learner have dual experiences in both real and virtual worlds with ubiquitous computing technologies[17]. S. Yokota et al. have developed an e-learning system for studying embedded systems that combines three key components: an online textbook, a graphical simulator, and a remote real system connected to the internet, allowing students to learn embedded systems programming regardless of their location while eliminating the need for physical equipment [18]. All these efforts laid a solid foundation in Japan for the future introduction of intelligent technologies.

Among various technological innovations, AI has emerged as a transformative force in educational technology. Traditional tutoring approaches have long faced inherent limitations, including restricted accessibility, inconsistent pedagogical effectiveness, and challenges in scaling implementation [8] [19]. LLMs have emerged as a potential solution to these persistent challenges, offering a significant advancement in educational technology. LLMs demonstrate remarkable capabilities in understanding context, generating human-like responses, and adapting to diverse learning styles. These characteristics position them uniquely to address long-standing challenges in traditional tutoring systems, particularly in terms of availability, consistency, and scalability. However, the implementation of general-purpose LLMs like ChatGPT in educational contexts reveals significant limitations [20]. These systems often struggle with providing accurate, course-specific information and lack the ability to offer truly personalized guidance aligned with specific educational objectives.

In response to the limitations of general-purpose LLMs, researchers and education industry practitioners have increasingly turned to RAG as a promising solution. Modran et al. have proposed integrating RAG with educational LLMs to address the limitations of general-purpose models, demonstrating significant improvements in response accuracy and domain specificity [21]. Building on similar ideas, Liu et al. have developed a new system called HITA has been deployed to over 400 students spanning entry-level to advanced courses at the Colorado School of Mines for two semesters and earned more than 97 percent positive responses [22].

However, in some latest works, researchers have identified several limitations in current RAG implementations. Song et al. have proposed Trust-Score, a holistic metric that evaluates the trustworthiness of LLMs within the RAG framework, and found that most of the existing prompting methods like in-context learning, fail to effectively adapt LLMs to the RAG task as measured by Trust-Score [23]. Agrawal et al. have investigated this dilemma by analyzing error patterns in existing RAG methods which are based on knowledge graphs and identifying eight critical failure points [24]. From these works, we can easily observe the challenges in handling user uncertainty and generating reasonable and practical suggestions. As for IT education aspects, researchers’ findings revealed that while RAG effectively retrieves relevant information, it struggles to adapt this knowledge to individual learning contexts [25]. And what is even worse is that RAG systems often fail to account for the complex interdependencies between different learning topics, particularly in structured STEM curricula like programming education.

These limitations have sparked new research directions in the field [26]. To show promising results in personalizing content delivery, adaptive learning algorithms should be applied to RAG methods. In addition, researchers should also explore the integration of uncertainty quantification methods with RAG, aiming to better handle the probabilistic nature of student learning patterns.

2 Proposed Method

Here we propose FuzzMonte-RAG, a novel framework that integrates fuzzy logic and Monte Carlo methods with retrieval-augmented generation to create a more adaptive and personalized learning experience in IT education.

System Overview

The core workflow of FuzzMonte-RAG consists of three sequential stages:

1. **Knowledge Retrieval and Processing**: Initially, the RAG module retrieves relevant educational content from the knowledge base. This content serves as the foundation for subsequent processing and personalization.

2. **Uncertainty Modeling**: The fuzzy logic module then processes both the retrieved content and learner information to model various uncertainties in the learning process, such as learner proficiency levels and content difficulty. This step transforms discrete educational concepts into continuous fuzzy variables that better reflect the reality of learning.

3. **Path Optimization**: Finally, the Monte Carlo simulation module takes the fuzzy-processed information to generate and evaluate potential learning paths, identifying optimal sequences that maximize learning outcomes while respecting topic dependencies and learner characteristics.

Our framework addresses three fundamental challenges in AI-driven education that current RAG systems struggle with. Fig 1 demonstrates the processing overview of FozzMonte-RAG framework. First and foremost, while traditional RAG systems may retrieve relevant content, they often fail to adapt it to individual learning contexts. Our approach ensures that content is not only relevant but also appropriately matched to learner capabilities through sophisticated content adaptation mechanisms. Furthermore, standard systems typically treat learning capabilities and content difficulty as discrete, deterministic values, yet the reality of learning is far more nuanced. Our fuzzy logic approach better captures the continuous, uncertain nature of learning processes, acknowledging that understanding exists on a spectrum rather than in binary states. Most critically, existing systems often struggle with complex topic interdependencies, particularly in IT education where concepts build upon each other in intricate ways. Our Monte Carlo method addresses this limitation by systematically exploring and evaluating learning sequences while carefully considering these dependencies, ensuring that learners encounter topics in an order that optimizes their understanding and retention.

Enhanced RAG Module

The foundation of FuzzMonte-RAG is an enhanced retrieval-augmented generation module specifically designed for educational contexts. Unlike conventional RAG systems that primarily focus on information accuracy, our enhanced RAG module incorporates educational metadata and structural information to support more effective learning experiences.

At the core of our enhanced RAG module is a hierarchical knowledge base that maintains not only the educational content but also critical metadata such as prerequisite relationships, difficulty levels, and learning objectives. The retrieval process operates on both content similarity and educational relevance through a dual-encoder architecture. The first encoder processes traditional content-based features, while the second encoder specifically handles educational metadata, allowing the system to consider both semantic relevance and pedagogical appropriateness simultaneously.

Our retrieval mechanism employs a novel scoring function that combines the following three aspects: First, traditional content similarity measures the semantic alignment between the query and stored content using dense vector representations. Then, educational relevance evaluates how well the content matches the current learning context, including the learner's progress and educational goals. Last but maybe the most important, structural compatibility assesses how the content fits within the broader learning curriculum by considering prerequisite relationships and topic dependencies.

Formally, given a query q and a candidate document d, the retrieval score is computed as:

$$R(q,d) = \alpha \cdot sim(q,d) + \beta \cdot eduRel(q,d) + \gamma \cdot strucComp(q,d)$$

where $sim(q,d)$ represents traditional semantic similarity using dense vector representations, $eduRel(q,d)$ measures educational relevance through learning progress and goals alignment, and $strucComp(q,d)$ evaluates prerequisite relationship compatibility. The weights α, β, and γ are learnable parameters being fit after being deployed with online learning to balance these components and sum to 1. The semantic similarity $sim(q,d)$ is computed by first projecting both query and document into a shared vector space through encoder E, followed by applying the cosine similarity metric, like:

$$sim(q,d) = \cos(E(q), E(d)) = \frac{E(q) \cdot E(d)}{\|E(q)\| \|E(d)\|}.$$

For content generation, our RAG module goes beyond standard language models by following educational best practices. When generating explanations, it follows a structured approach that teachers commonly use: starting with simple examples, gradually introducing abstract concepts, and regularly checking if students understand the material. This ensures that the content is not only accurate but also easy to follow and learn from.

The enhanced RAG module then feeds its output to the next stages of our system - the fuzzy logic and Monte Carlo components. Along with the educational content itself, it provides important contextual information like topic relationships and difficulty levels. This comprehensive information helps our system make better decisions about what content to show each student and in what order.

Fuzzy Logic Modeling

While traditional educational systems often treat learning states as discrete values, real-world learning processes exhibit continuous and uncertain characteristics. Our fuzzy logic system models this uncertainty by formalizing three components of the learning process: learner proficiency, content complexity, and topic relations. In our framework, each learning concept C is represented as a triple:

$$C = (P, D, R)$$

where P denotes proficiency, D denotes difficulty, and R represents relationships. For each component, we define corresponding fuzzy sets and membership functions. The proficiency P is characterized by a set of fuzzy membership functions:

$$P = \{\mu_{novice}(x), \mu_{intermediate}(x), \mu_{advanced}(x)\}$$

where each membership function is modeled using trapezoidal fuzzy numbers:

$$\mu(x) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a < x \le b \\ 1, & b < x \le c \\ \frac{d-x}{d-c}, & c < x \le d \\ 0, & x > d \end{cases}$$

where $[a,d]$ defines the support set and $[b,c]$ represents the core interval.

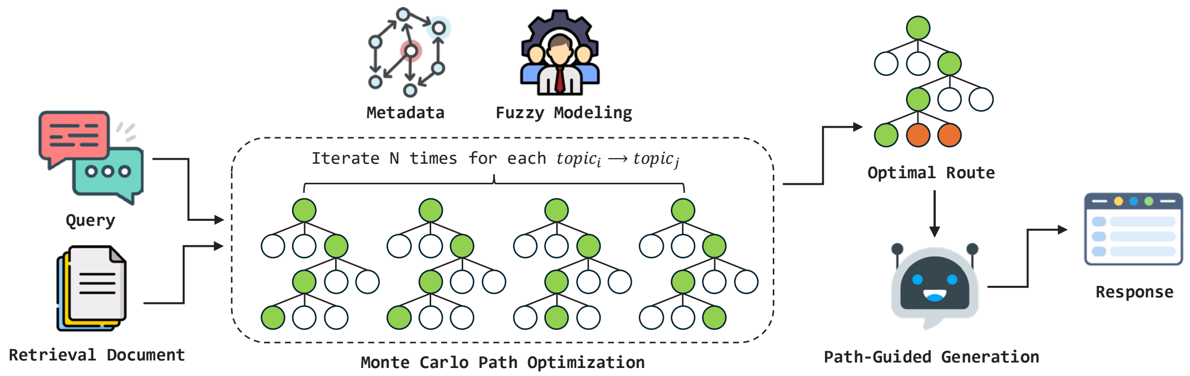

Fig. 1 The overview of proposed framework, integrating RAG retrieval with fuzzy modeling metadata, and path optimization to generate personalized learning responses

The difficulty D is also modeled as a fuzzy set, but multi-dimensional:

$$D = \{D_{\text{complexity}}, D_{\text{abstraction}}, D_{\text{dependency}}\}$$

Each dimension is associated with its own fuzzy rule base. The rules are formulated as conditional statements that map input characteristics to fuzzy difficulty levels. For example, the fuzzy rules for complexity assessment might include: *"Topics that contain nested concepts and require multi-step reasoning are considered highly complex, while concepts that can be directly applied without reasoning steps are assessed as having low complexity."*

While D and P may appear similar in their fuzzy set formulations, they fundamentally differ in their nature and application. D represents inherent properties of learning materials determined through expert evaluation, whereas P captures the dynamic state of learner understanding through observable metrics. In addition, P values change dynamically with learning progress, D values remain relatively fixed.

Topic relationships R are represented using a fuzzy relation matrix:

$$R_{ij} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{pmatrix}$$

where $r_{ij} \in [0, 1]$ indicates the strength of relationship between topics i and j.

These fuzzy components are integrated through a Mamdani-type fuzzy inference system, which consists of fuzzification, rule evaluation, aggregation, and defuzzification stages. The system processes input variables through the fuzzy rule base to generate learning recommendations, effectively handling the inherent uncertainty in educational decision-making. The fuzzy inference results provide essential input for the Monte Carlo path generation process. Specifically, we transform the fuzzy evaluations into probabilistic weights:

$$w_{ij} = f(r_{ij}, \mu_P(x_i), D_j)$$

where w_{ij} represents the transition probability from topic i to topic j, r_{ij} stands for the fuzzy relationship strength, $\mu_P(x_i)$ denotes the learner's current proficiency level, and D_j the difficulty of the target topic. The probabilistic interpretation of fuzzy results have bridged the gap between uncertainty modeling and path optimization, setting the foundation for our Monte Carlo approach described in the next part.

Monte Carlo Learning Path Optimization

Learning path generation in educational systems must balance between optimal sequencing and exploration of diverse learning possibilities. We have approached this challenge through a Monte Carlo-based optimization framework that leverages the fuzzy evaluations from the previous stage.

The core of our approach lies in the construction of a transition probability matrix W, where each element w_{ij} represents the likelihood of transitioning from topic i to topic j:

$$w_{ij} = \theta \cdot norm(r_{ij}) + \phi \cdot h(\mu_P(x_i), D_j)$$

where:

$$h(\mu_P(x_i), D_j) = \begin{cases} 1, \text{if } \mu_P(x_i) \text{ matches } D_j \\ \exp(-\eta \cdot \Delta(\mu_P(x_i), D_j)), \text{otherwise} \end{cases}$$

Based on these transition probabilities, we generate candidate instruction paths through an iterative sampling process. Starting from an initial topic s asked by a certain learner at the beginning, a path p is constructed by sampling subsequent topics according to W:

$$p = (s, t_1, t_2, \ldots, t_n)$$

For each generated path, we evaluate its quality using a reward function that considers multiple objectives:

$$R(p) = \sum_{i=1}^{n} [w_{i,i+1} \cdot (1 + \lambda \cdot co(t_i, t_{i+1}))]$$

where $co(t_i, t_{i+1})$ captures the educational coherence between consecutive topics, and λ balances the importance of transition probabilities and coherence.

Then, we employ a modified Monte Carlo Tree Search (MCTS) algorithm for path generation. Unlike traditional MCTS which builds a complete search tree, our method focuses on sampling meaningful educational sequences. For each decision point, we perform N simulations to evaluate potential future sequences. The sampling process follows:

Algorithm 1 Monte Carlo Path Optimization

Given: current topic t_c, proficiency state P, horizon H
for i = 1 to N **do**
  p_i = SimulatePath(t_c, P, H)
  v_i = EvaluatePath(p_i)
end for
Estimate expected value: $E[v|t] = \frac{1}{N}\sum_{i=1}^{N} v_i$

In each simulation step, we first generate a candidate path by forward sampling H steps according to the transition probabilities, where each step considers the fuzzy-derived weights w_{ij}. Along this candidate path, we simulate the learner's proficiency changes using fuzzy inference rules, updating the membership degrees $\mu_P(x)$ based on the encountered topic difficulties. The simulation then accumulates rewards that account for both immediate transitions and potential future learning effects, weighted by a discount factor γ to balance short-term and long-term educational benefits. The expected value of a path is estimated through repeated simulations:

$$E[R(p)|t_c] = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{H}\gamma^k \cdot [w_{k,k+1} \cdot (1 + \lambda \cdot co(t_k, t_{k+1}))]$$

Based on the Monte Carlo estimates, we select the next topic using a probabilistic policy that balances exploitation and exploration:

$$\pi(t|t_c) = \frac{\exp(E[R(p)|t_c]/\tau)}{\sum_{t'}\exp(E[R(p')|t_c]/\tau)}$$

where τ is a temperature parameter controlling the exploration-exploitation trade-off.

Integration with Fuzzy System

The Monte Carlo simulation process actively incorporates fuzzy evaluations at multiple levels. In each simulation step, the learner's state transition follows fuzzy inference rules:

$$P'(x) = F(P(x), D_t)$$

where $P'(x)$ is the predicted next proficiency state, and F represents the fuzzy inference process considering current proficiency $P(x)$ and topic difficulty D_t.

The reward accumulation during simulation considers fuzzy membership degrees:

$$r_t = w_{ij} \cdot (1 + \lambda \cdot co(t_i, t_{i+1})) \cdot \mu_P(x_t)$$

This integration creates a dynamic feedback loop where fuzzy proficiency levels influence transition probabilities w_{ij}, while each simulated path generates predicted proficiency changes through fuzzy inference rules. The fuzzy system continuously evaluates the appropriateness of transitions, and these evaluations are incorporated into Monte Carlo estimates, creating a comprehensive uncertainty-aware planning process.

By embedding fuzzy evaluations in the simulation process, our method captures learning uncertainties in path planning while maintaining educational coherence during exploration. This tight coupling between fuzzy logic and Monte Carlo simulation enables more realistic path optimization that adapts to gradual changes in learner proficiency, ultimately respecting the continuous nature of learning progress.

In summary, our proposed FuzzMonte-RAG framework enhances traditional retrieval-augmented generation by incorporating fuzzy logic for uncertainty modeling and Monte Carlo simulation for path optimization. The enhanced RAG module provides a foundation for content retrieval and generation, while the fuzzy logic system captures the continuous nature of learning processes through carefully designed membership functions and inference rules. The Monte Carlo path optimization leverages these fuzzy evaluations to explore and evaluate potential learning sequences, creating a dynamic system that adapts to individual learning progress. Through the tight integration of these components, our framework addresses the key challenges in AI-driven education: content relevance, learning uncertainty, and path optimization. This comprehensive approach enables more personalized and effective learning experiences that respect both the inherent uncertainties in education and the need for coherent knowledge construction.

Fig. 2 Sample question and response, showing sequential topic explanations based on Monte Carlo-optimized learning paths, using local LLM model

3 Technical Implementation

System Architecture

We have implemented our system using a multi-model approach based on the Ollama platform. The system architecture employs specialized language models for different components of the educational interaction pipeline. Our implementation leverages different model capabilities. Fig. 2 illustrates a sample question and response pair of our system deployed in a real-world company, where the model accurately responds to various knowledge topics that may be involved in the query and explains them one by one.

- **Dialogue Management**: Qwen2.5($7B$) handles the primary user interaction, which includes understanding learning queries and managing educational dialogue flow. This model is responsible for interpreting user questions and learning needs and managing conversation context. It also needs to find the initial t for Monte Carlo processing.

- **Technical Content Generation**: Qwen2.5-Coder specializes in generating IT education content by producing technical explanations, creating learning examples, generating practice exercises, and adapting difficulty levels based on fuzzy inference results.

Due to the nature of our MCTS-based topic selection, each response requires multiple calls to the Technical Content Generation module, with the Dialogue Management module integrating these contents into a coherent response. Specifically, when exploring different paths in the Monte Carlo tree, the system generates and evaluates content for multiple potential topics before determining the optimal learning sequence. Ollama's architecture allows us to easily scale the number of LLM instances when sufficient GPU memory is available, with each instance handling different aspects of content generation concurrently. Moreover, by following RESTful principles in our API design, all model interactions are stateless and standardized, where each request contains all necessary context for processing. This architectural decision significantly reduces the complexity of future maintenance and scaling efforts, as new model instances can be seamlessly added or removed without affecting the overall system behavior.

Parallel Monte Carlo Simulation

Given the computational demands of our Monte Carlo simulation process discussed before, we implement a parallel processing architecture to efficiently explore learning paths. The parallelization strategy operates at two levels: simulation-level parallelism and path-level parallelism.

At the simulation level, we distribute multiple independent Monte Carlo simulations across multiple computing nodes. Each simulation starts from the current learning state and explores future paths independently, following the transition probabilities w_{ij} we defined. This approach is particularly efficient because individual simulations are inherently independent and require minimal inter-process communication.

At the path-level parallelism, when each simulation generates and evaluates candidate paths, the content generation and fuzzy evaluation processes for different topics in the path horizon H are processed concurrently. This is crucial because, as discussed before, each topic evaluation requires both fuzzy inference calculations and content quality assessments. The parallel processing at this level significantly reduces the latency of path evaluation, especially when exploring deeper horizons.

The results from parallel simulations are aggregated using a weighted averaging scheme that considers both the exploration coverage and the reliability of individual simulations. This parallel architecture enables our system to efficiently explore a broader range of potential learning paths while maintaining real-time response capabilities for interactive learning scenarios.

4 Evaluation

Experimental Setup

We have evaluated our FuzzMonte-RAG framework through a real-world IT educational LLM at SoftUsing Co., Ltd., which is also our main collaboration partner and sponsor of this project. The system has been used to assist in training new technical staff. The participant group of over 20 employees represents diverse educational and professional backgrounds: 40 percent were recent graduates with IT or engineering degrees, 35 percent were career changers with several years of work experience in non-technical fields, and notably, 25 percent came from completely non-STEM backgrounds such as liberal arts or business. This diversity in learning backgrounds presents an ideal test case for our system's adaptive capabilities, as it needed to effectively serve learners with vastly different starting points and learning needs.

In collaboration with the company's HR and technical training departments, we have conducted a three-month evaluation period in 2024. The training program covered essential IT skills including programming fundamentals, and advanced AI knowledge, challenging our system to provide appropriately scaffolded learning paths for participants of varying technical proficiency.

Evaluation Methodology

The evaluation has been conducted through a comprehensive questionnaire survey completed by participating employees. The questionnaire was designed to assess multiple aspects of the learning experience: Learning Effectiveness, System Usability, and Adaptive Learning.

Results Analysis

Participants have rated each aspect on a 5-point Likert scale (1: Strongly Disagree to 5: Strongly Agree). The survey results show:

- 87.5% of participants (21/24) reported improved learning efficiency compared to traditional training methods.
- 79.2% of new employees (19/24) found the adaptive difficulty adjustment helpful for their learning process
- 95.8% (23/24) appreciated the system's ability to provide coherent learning paths.
- 66.7% (16/24) noted that the system effectively helped them overcome learning obstacles.

Some qualitative feedback highlighted several key advantages of our system:

- "*The system's ability to adjust difficulty based on my understanding was particularly helpful.*"
- "*Content explanations were clear and well-connected to previous topics.*"
- "*The interactive nature of learning made complex technical concepts more approachable*"

Discussion

The evaluation results validate our approach's effectiveness in corporate technical training. The high satisfaction rates with adaptive difficulty adjustment and learning path coherence specifically support the benefits of our fuzzy logic and Monte Carlo optimization components. However, we also identified areas for improvement, particularly in response speed and content variety, which will be addressed in future updates.

5 Conclusion

Here we have presented FuzzMonte-RAG, a method that combines fuzzy logic and Monte Carlo optimization with retrieval-augmented generation for IT education. The system addresses key challenges in AI-driven education by modeling learning uncertainties and optimizing learning paths. Evaluation results from our system at real-world business has demonstrated its effectiveness in training employees with diverse backgrounds. While the current results are promising, future work could explore expanding the content coverage and improving the system's response time. Overall, our work contributes to the growing field of AI-driven education by providing a more adaptive and personalized learning experience.

References

1. Liu M, Yu D. Towards intelligent E-learning systems. Educ Inf Technol (Dordr). 2023;28: 7845 – 7876.

2. Alzahrani NM. Augmented reality: A systematic review of its benefits and challenges in e-learning contexts. Applied Sciences. 2020;10: 5660.

3. Effendi H, Hendriyani Y, Humaira JD. The future of e-learning: Leveraging VR, AR, and AI for more effective and engaging learning experiences. Proceedings of vocational engineering international conference. 2023. pp. 244 – 251.

4. Gligorea I, Cioca M, Oancea R, Gorski A-T, Gorski H, Tudorache P. Adaptive learning using artificial intelligence in e-learning: a literature review. Educ Sci (Basel). 2023;13: 1216.

5. Murtaza M, Ahmed Y, Shamsi JA, Sherwani F, Usman M. AI-based personalized e-learning systems: Issues, challenges, and solutions. IEEE access. 2022;10: 81323 – 81342.

6. Panjaburee P, Komalawardhana N, Ingkavara T. Acceptance of personalized e-learning systems: a case study of concept-effect relationship approach on science, technology, and mathematics courses. Journal of Computers in Education. 2022;9: 681 – 705.

7. Ashfaque MW, Tharewal S, Iqhbal S, Kayte CN. A Review on Techniques, Characteristics and approaches of an intelligent tutoring Chatbot system. 2020 International conference on smart innovations in design, environment, management, planning and computing (ICSIDEMPC). 2020. pp. 258 – 262.

8. Alam A. Harnessing the power of AI to create intelligent tutoring systems for enhanced classroom experience and improved learning outcomes. Intelligent communication technologies and virtual mobile networks. Springer; 2023. pp. 571 – 591.

9. Salvagno M, Taccone FS, Gerli AG. Artificial intelligence hallucinations. Crit Care. 2023;27: 180.

10. Huang L, Yu W, Ma W, Zhong W, Feng Z, Wang H, et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. 2023. Available: https://arxiv.org/abs/2311.05232

11. Gao Y, Xiong Y, Gao X, Jia K, Pan J, Bi Y, et al. Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:231210997. 2023.

12. Zhao P, Zhang H, Yu Q, Wang Z, Geng Y, Fu F, et al. Retrieval-augmented generation for ai-generated content: A survey. arXiv preprint arXiv:240219473. 2024.

13. Li D. Creating personalized higher education teaching system using fuzzy association rule mining. Int J Comput Int Sys. 2024;17: 239.

14. Gunasekaran A, McNeil RD, Shaul D. E-learning: research and applications. Ind Commer Train. 2002;34: 44 – 53.

15. Demir K, Akpinar E. The effect of mobile learning applications on students' academic achievement and attitudes toward mobile learning. Malaysian Online Journal of Educational Technology. 2018;6: 48 – 59.

16. Drigas AS, Pappas MA. A review of mobile learning applications for mathematics. International Journal of Interactive Mobile Technologies. 2015;9.

17. Ota K, Dong M, Wang J, Niher H, Jing L, Kansen M, et al. A Courseware Platform of Dual Experience for Safety Education Utilizing both a Miniaturized World and a Computerized World. Japan e-Learning Association Journal. 2010;10: 81 – 92.

18. YOKOTA S, HIRAMATSU K, OHYAMA Y, Jin-Hua S, HASHIMOTO H. E-learning System with a Real System and Graphical Simulator for Embedded System. IFAC Proceedings Volumes. 2008;41: 12649 – 12654.

19. Zhang K, Aslan AB. AI technologies for education: Recent research & future directions. Computers and Education: Artificial Intelligence. 2021;2: 100025.

20. Opara E, Mfon-Ette Theresa A, Aduke TC. ChatGPT for teaching, learning and research: Prospects and challenges. Opara Emmanuel Chinonso, Adalikwu Mfon-Ette Theresa, Tolorunleke Caroline Aduke (2023) ChatGPT for Teaching, Learning and Research: Prospects and Challenges Glob Acad J Humanit Soc Sci. 2023;5.

21. Modran H, Bogdan IC, Ursuțiu D, Samoila C, Modran PL. LLM intelligent agent tutoring in higher education courses using a RAG approach. 2024.

22. Liu C, Hoang L, Stolman A, Wu B. HiTA: A RAG-Based Educational Platform that Centers Educators in the Instructional Loop. International conference on artificial intelligence in education. 2024. pp. 405 – 412.

23. Song M, Sim SH, Bhardwaj R, Chieu HL, Majumder N, Poria S. Measuring and enhancing trustworthiness of LLMs in RAG through grounded attributions and learning to refuse. arXiv preprint arXiv:240911242. 2024.

24. Agrawal G, Kumarage T, Alghamdi Z, Liu H. Mindful-rag: A study of points of failure in retrieval augmented generation. arXiv preprint arXiv:240712216. 2024.

25. Dakshit S. Faculty perspectives on the potential of RAG in computer science higher education. arXiv preprint arXiv:240801462. 2024.

26. Wang S, Xu T, Li H, Zhang C, Liang J, Tang J, et al. Large language models for education: A survey and outlook. arXiv preprint arXiv:240318105. 2024.