

Large Language Model Agents for Personalized Education: Adaptive Tutoring and Reinforcement Learning-Based Learning Path Optimization

Vince Heere

Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.

vincemail@unr.edu

Jinthe Kalman

Department of Computer Science, University of New Hampshire, Durham, NH, USA.

kalman1999@unh.edu

Abstract

Personalized education seeks to approximate the responsiveness, diagnosis, and pacing of expert human tutoring at scale, yet traditional intelligent tutoring systems often remain constrained by brittle domain models, limited natural language flexibility, and static sequencing logic. Large language model agents introduce a new system-level possibility: they can interpret open-ended learner language, coordinate pedagogical tools, maintain extended interaction histories, and generate context-sensitive explanations while remaining embedded in measurable instructional workflows. This paper examines the architecture, deployment, and governance of large language model agents for adaptive tutoring and reinforcement learning-based learning path optimization. It argues that the central challenge is not merely model capability but the socio-technical integration of learner models, curriculum representations, reinforcement learning controllers, retrieval infrastructure, assessment instruments, and human oversight. The paper analyzes adaptive tutoring architectures, reward design for sequencing, orchestration trade-offs across cloud and edge environments, robustness under hallucination and distribution shift, fairness and privacy obligations, and sustainability concerns associated with large-scale inference. It also considers case illustrations across K-12 mathematics, higher education, and professional learning, and compares educational deployments with other high-stakes service infrastructures. The conclusion emphasizes that durable personalization requires interpretable learner representations, auditable optimization objectives, resilient fallback pathways, and policy frameworks that treat educational agents as accountable public-interest systems rather than isolated software products.

Keywords

personalized education; large language model agents; adaptive tutoring; reinforcement learning; learning path optimization; educational governance; socio-technical systems.

1. Introduction

Personalized education has long been animated by the aspiration to provide every learner with instruction as responsive as expert one-to-one tutoring. Bloom's influential account of the two-sigma problem framed the gap between conventional group instruction and individualized tutoring as both a pedagogical and an infrastructural challenge [1]. Subsequent research on intelligent tutoring systems demonstrated that targeted feedback, mastery learning,

and fine-grained task sequencing can produce meaningful learning gains, though often within bounded domains and carefully engineered curricula [2]. Systems such as cognitive tutors and mixed-initiative dialogue tutors showed that computational instruction can diagnose errors, scaffold problem solving, and sustain learner engagement when the underlying domain model is sufficiently structured [3], [4]. At the same time, systematic reviews of chatbots and artificial intelligence in education have repeatedly noted that technological enthusiasm often outpaces evidence of durable learning, equitable access, and institutional integration [5], [7].

Large language model agents mark a qualitative shift in this landscape. Unlike earlier tutoring systems that relied primarily on hand-authored rules or narrowly scoped natural language interfaces, contemporary agents can interpret informal learner explanations, generate analogies, summarize misconceptions, and coordinate external tools such as calculators, retrieval services, and assessment engines [8], [9], [10], [11]. When combined with reinforcement learning for sequencing, such agents could adapt not only the wording of feedback but also the trajectory of tasks, difficulty, modality, and review. Yet this promise introduces complex structural trade-offs. Open-ended generation can produce fluent but incorrect explanations, reinforcement learning can optimize proxies that diverge from learning, and deployment at scale can create privacy, fairness, and sustainability risks. This paper therefore treats personalized education as a socio-technical infrastructure problem rather than a single-model problem. It examines how large language model agents can support adaptive tutoring and learning path optimization while remaining governable, robust, and pedagogically accountable.

The paper proceeds through conceptual foundations, adaptive tutoring architectures, reinforcement learning for sequencing, infrastructure and deployment trade-offs, governance, robustness, case illustrations, and future directions. The argument is that personalization is most credible when agentic flexibility is constrained by explicit learner models, validated curriculum representations, transparent reward structures, and human oversight. The educational value of large language model agents will depend less on benchmark performance than on whether institutions can integrate them into accountable routines that protect learners and preserve instructional integrity [6].

2. Foundations of Large Language Model Agents in Education

Large language models acquire broad linguistic and procedural competence through pretraining on large text corpora, followed by instruction tuning and preference optimization that align outputs with human expectations [8], [9]. In educational settings, this capability enables agents to explain concepts in multiple registers, respond to learner affect, and shift between procedural hints and conceptual discussion. However, an agent is not simply a language model. Agentic behavior requires planning, memory, tool use, and interaction control. Frameworks that combine reasoning with acting allow models to decide when to retrieve information, query a calculator, call an assessment function, or ask the learner for clarification [10]. Tool-augmented models can also learn to invoke external APIs, which is essential for education because arithmetic, symbolic manipulation, curriculum lookup, and record keeping should not be entrusted to generative fluency alone [11]. Surveys of autonomous language model agents further emphasize that memory design, action space definition, and environment feedback shape whether agents remain useful over extended tasks [12].

For education, these foundations imply that the agent should be understood as a coordinating layer rather than an oracle. The learner interacts with a pedagogical interface, but behind that

interface may sit a planning module that selects instructional strategies, a retrieval module that surfaces curriculum-aligned content, a learner model that estimates mastery, and a reinforcement learning controller that sequences activities. The language model contributes flexible explanation and dialogue management, but it must be bounded by domain constraints, assessment evidence, and institutional policy. This is particularly important because educational interactions occur over weeks or months, not isolated turns. Long-horizon memory can support continuity but also risks storing inaccurate inferences, reinforcing stale misconceptions, or exposing sensitive information. Therefore, the foundation of educational agents must include explicit representations of what the system knows, what it merely suspects, and what it should not claim. The difference between a compelling conversational interface and a reliable tutoring infrastructure lies in these representational and governance choices.

3. Adaptive Tutoring Architectures and Pedagogical Interaction

Adaptive tutoring architectures typically combine a domain model, a learner model, a pedagogical model, and an interface layer. In earlier intelligent tutoring systems, the domain model encoded correct procedures and common errors, while the learner model tracked skill acquisition through knowledge tracing or related techniques [2], [17]. Deep knowledge tracing extended this tradition by using sequence models to estimate mastery from learner responses, allowing finer-grained prediction of future performance [16]. Large language model agents can enrich these components by interpreting natural language explanations and generating adaptive feedback, but they do not eliminate the need for structured learner models. Without such models, personalization can drift into stylistic adaptation rather than pedagogically meaningful adjustment. A system that changes tone but not difficulty, feedback specificity, or review schedule is not truly adaptive in the instructional sense.

Pedagogical interaction is equally important. Effective tutoring depends on scaffolding, diagnosis, and metacognitive support. AutoTutor demonstrated that mixed-initiative dialogue can emulate aspects of human tutoring by prompting learners to articulate reasoning and by responding to incomplete or erroneous explanations [4]. Large language model agents can extend this capacity, but they must avoid the temptation to provide answers too quickly. A well-designed agent should distinguish between procedural hints, conceptual prompts, worked examples, and full solutions, and it should choose among them based on learner state and task context. Reviews of educational chatbots show that interaction quality, perceived usefulness, and learning outcomes vary widely depending on how dialogue is integrated with instructional goals [5]. The architecture must therefore encode pedagogical policies, not merely conversational fluidity. It should also support teacher visibility, because instructors need to understand why a learner received a particular sequence, how mastery was inferred, and where intervention may be necessary. In this sense, adaptive tutoring is a shared human-machine activity rather than a fully autonomous replacement for teaching.

4. Reinforcement Learning for Learning Path Optimization

Learning path optimization can be framed as a sequential decision problem in which the system chooses activities, difficulty levels, and review intervals to maximize long-term learning outcomes. Reinforcement learning provides a natural formalism for such sequential choice, because it balances immediate reward with future consequences and can adapt policies through interaction [13]. Deep reinforcement learning has demonstrated that complex control problems can be learned from high-dimensional feedback, though often at substantial computational cost and with sensitivity to reward design [14]. In education, the reward

function is the central design artifact. If the system rewards correct answers only, it may select easy tasks that inflate performance and undermine learning. If it rewards engagement time, it may optimize for persistence rather than understanding. If it rewards immediate test gains, it may neglect transfer, retention, and deeper conceptual growth. A robust learning path optimizer must therefore combine multiple signals, including accuracy, response time, hint use, error type, retention after delay, and learner confidence.

Knowledge tracing models can supply state estimates that inform reinforcement learning controllers, but they also introduce uncertainty [16], [17]. A learner model may misestimate mastery when learners guess, disengage, or receive help from outside the system. The optimizer must account for this uncertainty rather than treating estimated mastery as ground truth. Exploration is another challenge. A reinforcement learning agent needs to try alternative sequences to discover effective policies, but excessive exploration can frustrate learners or waste instructional time. Safe exploration, constrained policy updates, and teacher-approved action spaces are therefore necessary. In practice, hybrid approaches may be more defensible than end-to-end reinforcement learning. A constrained optimizer can propose candidate paths, while teachers or domain experts define boundaries, prerequisites, and unacceptable sequences. This preserves the adaptive advantages of reinforcement learning while keeping optimization aligned with curricular and ethical goals. The key structural trade-off is between flexibility and legitimacy: the more autonomous the optimizer, the greater the need for transparent objectives, audit trails, and mechanisms for appeal or correction.

5. System Architecture, Infrastructure, and Deployment Trade-offs

A large language model agent for personalized education requires an orchestration architecture that coordinates inference, retrieval, learner modeling, reinforcement learning, assessment, and logging. Retrieval-augmented generation can ground explanations in approved materials, while tool use can outsource computation and record keeping to deterministic services [11]. Agent frameworks provide planning and action selection, but they also increase latency and failure surface area [12]. Recent work on scaling agentic models suggests that broader task coverage and multi-scenario competence can improve transfer across applications, yet such scaling also intensifies orchestration complexity, evaluation burden, and infrastructure cost [15]. For education, this means that architecture choices cannot be separated from deployment realities. A cloud-hosted system may offer stronger models and centralized updates, but it raises data residency, network dependency, and recurring cost concerns. An edge or on-premise deployment may improve privacy and latency, but it can constrain model size, update frequency, and tool availability.

Interoperability is another architectural concern. Educational institutions rely on learning management systems, student information systems, assessment platforms, and content repositories. An agent that cannot exchange data through standard interfaces will create fragmented workflows and duplicate records. Conversely, deep integration increases the consequences of error, because a mistaken mastery estimate may influence course placement, advising, or credentialing. The architecture should therefore separate generative components from authoritative records. Language model outputs can be treated as suggestions or explanations, while mastery records, grades, and formal progress decisions should remain governed by validated assessment and human review. Interpretability infrastructure is also necessary. Even if the full model is not interpretable, the system can expose why a path was chosen, which evidence was used, and which alternative paths were rejected [18]. This is not merely a technical reporting feature; it is a condition for institutional trust. The most

sustainable deployments will likely be modular, allowing schools to replace models, retrieval sources, or reinforcement learning policies without rebuilding the entire educational environment.

6. Governance, Fairness, Privacy, and Policy

Governance for educational agents must address privacy, fairness, transparency, accountability, and pedagogical legitimacy. Learners generate sensitive data about performance, behavior, affect, disability, language background, and family circumstances. Regulations such as the Family Educational Rights and Privacy Act in the United States and the General Data Protection Regulation in Europe establish obligations around consent, access, retention, and purpose limitation. Yet legal compliance alone is insufficient. Educational agents can also reproduce or amplify bias if training data, reward functions, or interaction patterns disadvantage particular groups. For example, a system optimized on average response speed may penalize learners who need more processing time, while a system trained primarily on dominant linguistic norms may misinterpret dialect or multilingual reasoning. Governance must therefore include disaggregated evaluation, participatory design with teachers and learners, and continuous monitoring after deployment [6], [7].

Transparency and accountability require more than explaining a single recommendation. Institutions need to know who is responsible when an agent misdiagnoses a misconception, recommends an inappropriate sequence, or discloses private information. A unified framework of ethical principles for artificial intelligence emphasizes beneficence, nonmaleficence, autonomy, justice, and explicability, but these principles must be translated into operational controls [20]. Operationally, this means model cards, data provenance records, approval workflows, human review thresholds, and clear escalation paths. It also means avoiding inappropriate automation in high-stakes decisions. Interpretable models or constrained decision rules may be preferable when decisions affect access, advancement, or credentialing [19]. In practice, educational agents should be designed as advisory systems unless rigorous evidence supports greater autonomy. The policy question is not whether agents can make decisions, but which decisions they should be permitted to make, under what evidentiary standards, and with what mechanisms for contestation. Fairness in personalized education is thus a governance property of the entire system, not a metric attached to a model.

7. Robustness, Sustainability, and Evaluation

Robustness is essential because educational agents operate in noisy, unpredictable environments. Learners misspell, change languages, express frustration, attempt prompt injection, or misunderstand instructions. Retrieval may return outdated content, tools may fail, and reinforcement learning policies may encounter states not represented in historical data. Large language models can hallucinate citations, invent procedures, or provide confidently incorrect explanations [5]. A robust system needs multiple defenses: grounded retrieval from vetted sources, uncertainty detection, contradiction checks, fallback to human help, and conservative behavior when confidence is low. Evaluation cannot rely solely on short-term quiz scores. It must examine retention, transfer, self-efficacy, equity of access, and effects on teacher workload. Longitudinal studies are especially important because personalization effects may appear only after weeks or months.

Sustainability introduces another dimension. Large-scale inference, frequent model updates, and continuous logging consume energy and financial resources. Schools with limited budgets may find that highly personalized agents are available only to wealthier districts,

reproducing existing inequalities. Sustainable design therefore includes model distillation, caching, selective inference, offline capabilities, and careful limits on agent autonomy. The environmental and economic costs of reinforcement learning exploration must also be considered, particularly when policies are trained through extensive simulation or live experimentation [14]. Interpretability and evaluation infrastructure can reduce long-term costs by making failures easier to diagnose and correct [18], [19]. A sustainable evaluation framework should compare personalized agents not only with traditional instruction but also with lower-cost alternatives such as improved teacher dashboards, targeted small-group tutoring, and structured peer learning. The question is not whether personalization is desirable in principle, but whether a given architecture delivers sufficient learning value relative to its cost, risk, and maintenance burden.

8. Case Illustrations and Cross-Domain Comparisons

In K-12 mathematics, an agent might combine step-by-step problem solving with a knowledge tracing model and a constrained sequencer. Cognitive tutor research showed that embedded assessment and just-in-time feedback can support learning in real classrooms, but implementation depends on teacher training, curriculum alignment, and technical support [3]. A large language model agent could reduce authoring burden by generating varied practice problems and explanations, while a reinforcement learning controller could schedule review. However, the system must avoid exposing students to unvetted content or optimizing for test preparation at the expense of conceptual understanding. In higher education, agents may support large introductory courses by answering routine questions, generating formative assessments, and recommending resources. Reviews of artificial intelligence in higher education caution that pedagogical and institutional questions are often secondary to technological novelty [7]. A university deployment would need to integrate with learning management systems, honor accessibility requirements, and preserve academic integrity.

Professional and corporate learning offers a different comparison. Here, learners may be adults with specific performance goals, and the system may optimize for competency acquisition, compliance, or job readiness. The reward function might include workplace performance indicators, but those indicators can be noisy and politically contested. Healthcare education illustrates the stakes: an agent that misunderstands a clinical procedure could reinforce dangerous practice. Cross-domain comparison shows that the same technical components produce different risk profiles depending on context. In all cases, the most effective systems are likely to be blended. Teachers, mentors, and peers remain essential for motivation, ethical judgment, and complex problem framing. Large language model agents can extend instructional capacity, but they should not be treated as substitutes for the relational and contextual dimensions of education [4], [5], [12]. The strongest case illustrations therefore emphasize augmentation, not replacement.

9. Future Directions and Research Agenda

Future research should move beyond demonstrations of fluent dialogue toward validated, long-horizon personalization. One direction is the development of multi-agent architectures in which separate agents handle diagnosis, explanation, assessment, and sequencing under a supervisory controller. Instruction-tuned models and reasoning-acting frameworks provide building blocks for such systems, but they require stronger coordination protocols and failure containment [9], [10]. Another direction is human-in-the-loop reinforcement learning, where teachers can inspect proposed paths, adjust rewards, and override policies. This would combine the adaptability of reinforcement learning with professional judgment. Research

should also examine how agentic scaling affects educational transfer, cost, and reliability, especially as systems are expected to handle many scenarios and learner populations [15]. Scaling may improve generality, but it may also obscure accountability unless governance keeps pace.

A third direction is the design of equitable personalization. This requires participatory research with diverse learners, culturally responsive content, and evaluation that disaggregates outcomes. Policy frameworks should specify when educational agents may act autonomously, what data they may retain, and how learners can contest decisions [20]. A fourth direction concerns assessment. If agents help learners produce work, traditional assessments may no longer reflect understanding. New forms of validated assessment, such as oral defense, process portfolios, and authentic tasks, may be necessary. Finally, sustainability research should compare centralized and distributed deployment models, quantify energy and cost, and identify when lighter-weight personalization is preferable. The research agenda is therefore interdisciplinary: computer science, learning science, ethics, policy, and institutional management must collaborate. Without this integration, personalized education risks becoming a collection of impressive but fragile tools rather than a durable educational infrastructure.

10. Conclusion

Large language model agents offer a powerful opportunity to advance personalized education by combining flexible dialogue, tool use, learner modeling, and reinforcement learning-based sequencing. They can potentially approximate aspects of expert tutoring that earlier systems could not easily provide, including natural language diagnosis, adaptive explanation, and dynamic path selection. However, the central lesson of this analysis is that capability alone is insufficient. Educational personalization is a system-level property that depends on architecture, infrastructure, governance, and evaluation. Agents must be grounded in validated content, constrained by pedagogical policy, integrated with authoritative records, and monitored for fairness, privacy, robustness, and sustainability. Reinforcement learning can optimize learning paths, but only if reward functions reflect deep learning goals rather than superficial engagement or short-term performance. The most credible deployments will treat agents as advisory collaborators within human-centered educational systems, not as autonomous authorities. If institutions invest in interpretable learner models, transparent optimization objectives, resilient fallback pathways, and participatory governance, large language model agents can meaningfully expand access to responsive instruction. If they do not, personalization may reproduce existing inequalities under a more sophisticated interface. The future of educational agents therefore depends less on the next model release than on the quality of the socio-technical structures built around it.

References

1. Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4–16.
<https://doi.org/10.3102/0013189X013006004>
2. VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221.
<https://doi.org/10.1080/00461520.2011.611369>

3. Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, 8, 30–43.
4. Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, 48(4), 612–618. <https://doi.org/10.1109/TE.2005.856149>
5. Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are we there yet? A systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence*, 4, 654924. <https://doi.org/10.3389/frai.2021.654924>
6. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
7. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39. <https://doi.org/10.1186/s41239-019-0171-0>
8. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
9. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35, 27730–27744.
10. Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023). ReAct: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations*. https://openreview.net/forum?id=WE_vluYUL-X
11. Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Zettlemoyer, L., ... Cancedda, N., & Scialom, T. (2023). Toolformer: Language models can teach themselves to use tools. *Advances in Neural Information Processing Systems*, 36, 68539–68551.
12. Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., ... Wen, J. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6), 186345. <https://doi.org/10.1007/s11704-024-40231-1>
13. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.
14. 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. <https://doi.org/10.1038/nature14236>
15. Jiang, Hao, et al. "AgentOmnia: Scaling Agentic Models for Full-Scenario Applications." *arXiv preprint arXiv:2607.23124* (2026).

16. 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, 505–513.
17. Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4), 253–278. <https://doi.org/10.1007/BF01099821>
18. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.
19. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
20. Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.8cd550d1>