Koh, E., et al. (Eds.) (2026). Proceedings of the 2nd International Conference on Learning Evidence and Analytics. Asia-Pacific Society for Computers in Education.

# Bridging Learning Analytics and Generative AI through Retrieval-Augmented Generation: A Conceptual Framework for Self-Regulated Learning Support

**Tzu-Chi Yang[a] & Jiun-Yu WU[b]**
[a]*Institute of Education, National Yang Ming Chiao Tung University, Taiwan (R.O.C)*
[b]*Department of Teaching and Learning, Southern Methodist University, USA*
*tcyang.academic@gmail.com

**Abstract:** Learning Analytics (LA) extracts meaningful behavioral patterns from digital learning environments; however, the interpretation and application of its outputs remain heavily dependent on data science expertise. While Generative AI (GenAI) offers real-time dialogue and content generation capabilities, it is constrained by insufficient domain knowledge and a lack of individualized learner context. This paper proposes a conceptual framework integrating LA, Retrieval-Augmented Generation (RAG), and GenAI, with the core innovation of leveraging learners' learning trajectories as a RAG knowledge source, thereby enabling GenAI to generate personalized recommendations grounded in authentic learning contexts. Grounded in Zimmerman's three-phase Self-Regulated Learning model, this paper delineates how the integrated framework provides differentiated support across the Forethought, Performance, and Self-Reflection phases, forming a dynamic learning support loop. Future work are further discussed.

**Keywords:** Learning Analytics, Generative Artificial Intelligence, Retrieval-Augmented Generation, Self-Regulated Learning, Personalized Learning Support

## 1. Introduction

Learning Analytics (LA) systematically collects and analyzes learners' behavioral data, including actions, assessments, and outcomes, throughout the learning process, aiming to comprehensively understand and optimize learning. Meanwhile, Generative Artificial Intelligence (GenAI) possesses powerful real-time interaction and content generation capabilities, offering dialogue-based support approaching human expert levels in educational contexts, and has emerged as one of the most significant educational tools of the contemporary era (Yan et al., 2024a). Despite this, both technologies have practical limitations. LA lacks general-purpose data interpretation capabilities and can only operate within structured workflows. GenAI, lacking learner's context, cannot generate feedback that is truly contextual, personalized, or situation specific. To address this, Retrieval-Augmented Generation (RAG) technology introduces external knowledge bases and domain-specific information to produce grounded and targeted responses, offering a solution that simultaneously addresses the limitations of both LA and GenAI (Li et al., 2025). Integrating the strengths of these three technologies LA, GenAI, and RAG opens new possibilities for truly learner-centered, personalized precision learning support. This paper proposes a LA+RAG+GenAI integrated framework supporting personalized Self-Regulated Learning (SRL), describing the benefits, potential challenges, and future directions of this integrated framework.

## 2. Related work

## 2.1 Learning Analytics

Learning Analytics is defined as “the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs” (Clow, 2013). Compared with traditional educational assessment that focuses on final outcomes or standardized tests, LA places greater emphasis on insights into the learning process. LA research focused on mining meaningful patterns from log data in digital learning environments (e.g., learning management systems) to provide timely feedback and intervention. By analyzing students’ clicks, settings, discussion forum posts, and other behaviors within the system, educators can gain a more comprehensive understanding of students’ engagement levels and potential difficulties. LA possesses several important characteristics, including evidence-driven approaches, multidimensional analysis, real-time monitoring and predictive capabilities. Despite the opportunities LA presents, its results require teachers and students with a certain level of data science literacy to utilize effectively, limited LA's advantages.

## 2.2 GenAI and RAG

GenAI can simultaneously comprehend graphics, tables, and text, autonomously interpret complex analytical results, and generate recommendations in real time, offering a solution to these challenges (Ochoa et al., 2025). Furthermore, compared with traditional LMS that provide learning support based on pre-defined rules or fixed resources, GenAI possesses flexible dialogue and real-time content generation capabilities. These include near-expert real-time interaction that breaks the delays and limitations of pre-set question-and-answer formats common in traditional intelligent tutoring systems, as well as the generalizability of large language model responses, which eliminates the need to develop separate tools for each discipline or skill. The diverse educational applications GenAI enables are too numerous to enumerate; however, GenAI’s application in educational settings still has limitations. The primary challenge is that educational practice relies on reliable knowledge frameworks and field-specific knowledge, while GenAI’s text generation operates probabilistically, necessitating mechanisms to ensure the reliability of output content. Moreover, even though GenAI is capable of responding with domain knowledge, it cannot provide truly precise personalized feedback, because its pre-training does not include the backgrounds, current situations, and learning context of teachers and students. This is a problem that GenAI currently cannot proactively solve. To address these issues, RAG technology offers a solution. RAG enables GenAI to retrieve relevant information from external or domain-specific knowledge bases before generating responses, thereby substantially enhancing the accuracy, specificity, and timeliness of outputs (Lewis et al., 2020), providing students with more precise and personalized learning support.

## 2.3 RAG Architectures for Education

The core concept of RAG is that when a large language model generates a response, a retrieval module finds information relevant to the query from an external knowledge base and uses this information as background knowledge to ground the generated content. Through this mechanism, the model’s responses can be based on actual data, understand the real world, and improve response appropriateness. In other words, RAG constrains GenAI’s generative output. Employing RAG in educational settings enables GenAI to more appropriately serve the role of an intelligent teaching assistant. Building upon this foundation, numerous advanced RAG architectures have been proposed (Zhao et al., 2024). There are common RAG architectures can all enhance GenAI’s potential in educational applications, such as Self-RAG (Asai et al., 2024), Corrective RAG (Yan et al., 2024b).Graph RAG (Edge et al., 2024). Speculative RAG (Wang et al., 2024) and Agentic RAG (Singh et al., 2025). In summary, the emergence of RAG has opened new pathways for integrating LA and GenAI. Through RAG, we can structure the insights generated by LA (such as student behavioral patterns) into retrievable knowledge and entrust GenAI to understand and utilize them. The following section introduces how LA, RAG, and GenAI can be combined to form an architecture supporting self-regulated learning, as well as the roles each plays A three phases of self-regulated learning.

## 3. The integrated LA+RAG+GenAI Framework

Building on this, we propose an LA+RAG+GenAI (LARGEN) concept: treating learners' learning logs themselves as part of the knowledge base. A learning system can store and vectorize learners' learning traces (e.g., clickstream records, response performance, and resource-use patterns) as well as learning outcomes. At the same time, by using learning logs as one source of the RAG knowledge base, GenAI gains concrete learner background, enabling it to generate suggestions grounded in the learner's context rather than generic advice, and thereby better support their learning process. Such an integrated framework enables the GenAI to retrieve related concepts and relations (e.g., via knowledge graphs) and to tailor the depth and scope of responses based on individual learner characteristics and learning portfolios, thus moving GenAI toward truly personalized support for SRL (Zimmerman, 2002), i.e., three phases-learning-cycle. Table 1 illustrates example regarding learning suggestions for three SRL phases based on different settings.

Table 1. *Comparison of personalized SRL support by LA, GenAI, RAG, and LARGEN*

| | Forethought Phase | Performance Phase | Self-Reflection Phase |
|---|---|---|---|
| LA | Your average accuracy on the "Limits" unit is 58%, which is below your target of 75%. | Over the past week, you have repeated the same quiz three times, but your score improvement is less than 10%. | Weekly learning summary: Total time: 60 minutes. Primary behavior: repeated quizzes (75%). Final accuracy: 62%, which reaches 60% of your stated target. |
| GenAI | A study plan to improve mathematics performance includes: (1) setting weekly learning goals; (2) scheduling regular practice time; and (3) ensuring a review of foundational concepts. | L'Hôpital's rule applies when the numerator and denominator both approach 0 or $\pm\infty$. Differentiate the numerator and denominator separately, then compute the new limit… (step-by-step solution provided). | It sounds like you have been working hard, but the outcomes have been limited. Consider: (1) What do you think is the biggest difficulty? (2) What might you try differently next time? |
| LARGEN | Your data indicate that "Limits" is currently the main challenge. A senior student with a similar profile previously set a goal of raising the unit-test accuracy for "Limits" to 80% within two weeks. You may start with this proximal goal and adopt their successful strategy at the time: reviewing error notes before practicing. | I notice you are stuck in a repeated quiz loop. This is not uncommon. A senior student in a similar situation found that the issue was an unclear understanding of the applicability conditions for L'Hôpital's rule. Would you like to pause and watch the 2-minute micro-video on this concept in the learning materials? (link) | Let's review this learning report. The data show you spent substantial time on immediate retakes, but the score gains were limited. Remember the senior student's "review error notes first" strategy? What differences do you think it might make if, next time, we switch from "immediate retake" to "review error notes first"? |

In brief, LA extracts behavioral patterns and learning features from multisource data; RAG retrieves knowledge sources such as LA outputs, historical similar experiences, literature, and teachers' experiences; and GenAI generates personalized learning supports based on retrieved evidence and current learner characteristics. Learners' subsequent actions produce new data that feed into the next loop.

## 4. Conclusion and Future work

Overall, integrating LA, RAG, and GenAI is not only a technical advance but also a catalyst for shifts in educational practice. It provides powerful tools for realizing genuinely learner-centered support and may become a key approach for technology-enhanced SRL. In this vision, human teachers and AI can jointly serve as mentors: learners are no longer passive recipients of knowledge but can actively construct their own learning pathways with AI assistance. At a macro level, real-time feedback and dynamic adjustment throughout the learning process may become a new norm, and AI-supported SRL is likely to grow in importance. Nevertheless, although the LARGEN framework has substantial potential in theory, the framework proposed in this paper has several limitations that require attention in future work. First, this paper proposes a conceptual framework and has not yet been validated through systematic empirical studies. Future research should adopt rigorous randomized controlled trials or quasi-experimental designs in authentic educational settings to examine the effects of the integrated architecture on learning outcomes, SRL behaviors, and learning motivation. Second, at the implementation level, standardizing learning data across educational platforms remains an open challenge. Although xAPI and Caliper provide mechanisms for cross-platform compatibility, differences in granularity and semantics across institutions and systems require further evaluation.

## Acknowledgements

This research was supported by the National Science and Technology Council in Taiwan through Grant numbers 114-2410-H-A49-022 -MY2.

## References

Asai, A., Wu, Z., Wang, Y., Sil, A., & Hajishirzi, H. (2024). Self-RAG: Learning to retrieve, generate, and critique through self-reflection. In *Proceedings of the Twelfth International Conference on Learning Representations (ICLR 2024)*. OpenReview. https://openreview.net/forum?id=hSyW5go0v8

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education, 18*(6), 683–695. https://doi.org/10.1080/13562517.2013.827653

Edge, D., Trinh, H., Cheng, N., Bradley, J., Chao, A., Mody, A., Truitt, S., Metropolitansky, D., Ness, O. R., & Larson, J. (2024). From local to global: A graph RAG approach to query-focused summarization. *arXiv*. https://doi.org/10.48550/arXiv.2404.16130

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *arXiv.* https://doi.org/10.48550/arXiv.2005.11401

Li, Z., Wang, Z., Wang, W., Hung, K., Xie, H., & Wang, F. L. (2025). Retrieval-augmented generation for educational application: A systematic survey. *Computers & Education: Artificial Intelligence, 8*, Article 100417. https://doi.org/10.1016/j.caeai.2025.100417

Ochoa, X., Huang, X., & Shao, Y. (2025). Exploring the potential of generative AI to support non-experts in learning analytics practice. *Journal of Learning Analytics, 12*(1), 65–90. https://doi.org/10.18608/jla.2025.8615

Singh, A., Ehtesham, A., Kumar, S., & Khoei, T. T. (2025). Agentic retrieval-augmented generation: A survey on agentic RAG. *arXiv*. https://doi.org/10.48550/arXiv.2501.09136

Wang, Z., Wang, Z., Le, L., Zheng, H. S., Mishra, S., Perot, V., Zhang, Y., Lei, T., Rossi, R., Dernoncourt, F., Deilamsalehy, H., Park, N., Rossi, R. A., & Pfister, T. (2024). Speculative RAG: Enhancing retrieval augmented generation through drafting. *arXiv*. https://doi.org/10.48550/arXiv.2407.08223

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology, 55*(1), 90–112. https://doi.org/10.1111/bjet.13370

Yan, S., Gu, J., Zhu, Y., & Ling, Z. (2024). Corrective retrieval augmented generation. *arXiv*. https://doi.org/10.48550/arXiv.2401.15884

Zhao, S., Yang, Y., Wang, Z., He, Z., Qiu, L. K., & Qiu, L. (2024). Retrieval augmented generation (RAG) and beyond: A comprehensive survey on how to make your LLMs use external data more wisely. *arXiv*. https://doi.org/10.48550/arXiv.2409.14924

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2