# Personalized Learning through AI-Driven Data Pipeline

**Ploy Thajchayapong and Ashok K. Goel**
Georgia Institute of Technology
ploy@gatech.edu, ashok.goel@cc.gatech.edu

**Abstract**

The integration of artificial intelligence (AI) in education holds significant promise for transforming personalized learning. By analyzing student learning data, AI systems can adapt instruction to meet individual needs through tailored content, adaptive learning paths, real-time feedback, and continuous improvement loops. However, effective personalization at scale demands not only access to large volumes of learner data but also robust data architectures to collect, organize, standardize, and analyze that data in a secure and meaningful way. However, note that the ability of AI to personalize learning requires data about the learner and prior learning. Personalization at scale requires data at scale. The Architecture for AI-Augmented Learning (A4L) framework addresses these needs by establishing a comprehensive data pipeline that supports AI-driven personalization. This pipeline introduced capabilities for direct data ingestion, anonymization, and standardization, as well as integrated analytics and visualization pipelines to deliver actionable insights to educators and learners alike.

## Introduction

The integration of artificial intelligence in education has the potential to revolutionize personalized learning (Luckin 2018; Holmes, Bialik, and Fadel 2019). By leveraging student learning data, educational systems can adapt to individual learning needs, offering tailored content and feedback (Liu and Koedinger 2017). Human-AI interaction can enable personalized learning by leveraging AI's ability to analyze vast amounts of learning data and adapt instruction to individual needs including adaptive learning paths, real-time feedback and assessment as well as continuous improvement through AI feedback loops (VanLehn 2011; Kulik 2013; Roll and Wylie 2016).

However, note that the ability of AI to personalize learning requires data about the learner and prior learning. Personalization at scale requires data at scale. The question then becomes not just of simple data but of a data architecture for deploying AI agents for learning and teaching, collecting, storing, standardizing, organizing, and analyzing data, and feeding the results back to teachers, learners, and the AI assistants for personalization of learning.

The A4L framework is an initiative designed to enhance personalized learning experiences through AI-driven insights. Evolving through multiple phases, A4L initially focused on data uploading, analysis, visualization, and sharing to better understand individual learner needs (Goel et al. 2025). Building on this foundation, the next iteration of A4L introduced direct data ingestion, Learning Tools Interoperability (LTI)[1] data integration, data anonymization, data standardization using Caliper Analytics[2], analytics pipeline and visualization pipeline to facilitate seamless personalization, ensuring more secured yet efficient and adaptive learning experiences. A4L facilitates personalized learning by analyzing individual learning behaviors, identifying strengths and weaknesses, and adapting instructional strategies accordingly.

## Personalized Learning through AI-Driven Framework

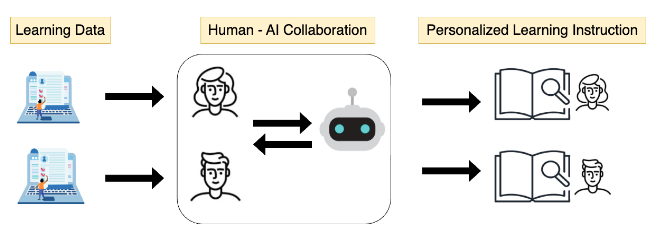


Figure 1. Personalized Learning Framework



[1] https://www.1edtech.org/standards/lti
[2] https://www.1edtech.org/standards/caliper

Figure 1 illustrates a conceptual framework for delivering personalized learning instruction through dynamic human-AI collaboration. The process begins with the continuous collection of learning data from students as they engage with digital platforms such as learning management systems (LMS) or AI-enabled educational tools. These platforms capture a wide array of behavioral signals—including clickstream activity, assessment outcomes, help-seeking behavior, and interaction histories—which serve as the raw input for personalization. This data flows into a human-AI collaborative loop, where both AI agents and human instructors interact and contribute distinct yet complementary strengths. Within this interaction space, the AI system provides predictive insights about future performance. Simultaneously, the human instructor brings pedagogical expertise and contextual understanding to the table. Teachers can interpret the AI-generated insights through the lens of their knowledge of student motivation, engagement levels, or non-observable classroom dynamics. This mutual exchange between the human and the AI agent allows for a more holistic interpretation of the data, enabling richer and more actionable instructional insights. The output of this collaborative process is the generation of personalized learning instruction. Instructors, supported by AI, adapt content delivery to meet the needs of individual learners. This may include adjusting reading materials, sequencing of topics, feedback timing, or providing differentiated assignments. Importantly, personalization is not delivered by the AI alone but is the result of a co-orchestrated effort that ensures both scalability and instructional quality. The final outcome is a learning experience that is not only tailored to each student's current needs but also responsive to their evolving learning trajectory—maximizing both effectiveness and engagement. In summary, Figure 1 represents more than a data flow—it encapsulates a pedagogical model in which technology augments human teaching, forming a continuous feedback loop between learning behavior, instructional insight, and personalized response.

## A4L Architecture

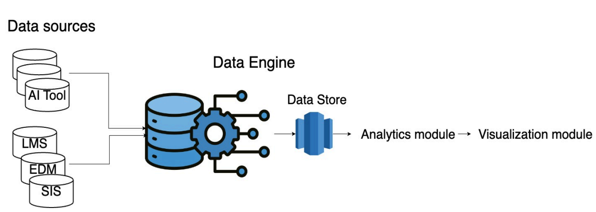


Figure 2. A4L Architecture

Figure 2 presents a high-level view of the A4L framework, depicting its end-to-end design for collecting, processing, and visualizing educational data in order to support personalized learning at scale. The architecture is structured around four interconnected components: Data Sources, Data Engine, the Analytics Pipeline, and the Visualization Pipeline. Together, these components enable a flow of information that drives both AI-powered insights and human-centered decision-making.

The process begins with the ingestion of data from diverse Data Sources, which include learning management system (LMS) logs, AI technologies, and institutional datasets such as Student Information System (SIS) records. These data sources capture a wide spectrum of learner activity and outcomes. Once collected, the data enters a centralized Data Engine, which is responsible for transforming heterogeneous inputs into a consistent, structured format. The engine performs several critical preprocessing tasks: data standardization, anonymization to protect student privacy, and transformation to prepare the data for advanced analysis. After processing, the cleaned and structured data is stored securely in the Data Store, a centralized repository that enables fast retrieval and efficient data access for downstream applications.

Once data is stored, it flows into the Analytics Pipeline, which leverages cloud-based infrastructure to perform real-time and batch analysis. This pipeline is designed for scalability and adaptability, allowing for flexible workflows that respond to different educational contexts and use cases. The pipeline extracts meaningful insights from the processed data, including behavioral trends, performance patterns, and indicators of learner engagement. These insights are tailored for adaptive learning applications and are essential for guiding both automated and human interventions.

The results of the analytics process are delivered through the Visualization Pipeline, which translates complex data outputs into intuitive, role-specific dashboards. These dashboards are accessible to instructors, learners, and researchers, and they serve as a key interface for human-AI collaboration in the classroom. Educators use these visualizations to inform instructional design and pedagogical decisions, while learners use them to reflect on their progress and self-regulate their learning strategies. The dashboard visualizations also serve as a critical mechanism for maintaining an ongoing feedback loop between students and teachers, enabling continuous improvement in teaching and learning.

At its core, the A4L framework is designed to advance several foundational goals. It supports the development of AI teaching assistants that enhance instruction and feedback delivery. It enables the large-scale collection and analysis of data on adult learners, providing empirical foundations for the personalization of learning. Finally, it strengthens the information feedback loop between teachers and students by integrating data-informed insights directly into the teaching process. By weaving together automated intelligence with human oversight, A4L represents a robust infrastructure for enabling equitable, adaptive, and effective learning experiences across diverse educational settings.

## Human-AI Interaction Feedback Loop

Dynamic feedback loop is a core of human-AI interaction in personalized learning environments (Park and Goel 2025). This model highlights the continuous and adaptive collaboration among the learner, the instructor, and the AI agent. At the center of this cycle is learner data, which is passively and actively generated as students engage with digital educational platforms—completing assignments, asking questions, navigating content, and interacting with instructional agents. This data serves as the foundational input for the system and reflects learners' progress, behaviors, challenges, and evolving needs.

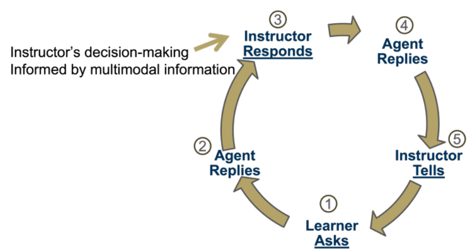


Figure 3. Human-AI interaction feedback loop

Figure 3 illustrates a human-AI collaborative feedback loop designed to support personalized learning through continuous and responsive interaction among learners, AI agents, and instructors. The cycle begins when a learner asks a question or seeks assistance during their engagement with the course. This learner-initiated action represents the entry point of the loop, highlighting the active role of students in driving their own learning process.

In response to the learner's query, the AI agent replies by delivering immediate support. This may include direct answers, recommended resources, or adaptive guidance based on the learner's current context and historical learning data. These AI-generated responses help to scaffold understanding and maintain engagement without delay.

The cycle then moves to the instructor responds phase. Here, the instructor reviews the learner's interaction and the AI agent's reply, along with broader learning analytics. Drawing on their pedagogical expertise, the instructor refines instructional strategies, adjusts learning materials, or intervenes with more targeted support. This phase reflects the human educator's critical role in interpreting AI insights within a broader instructional context.

Following the instructor's adjustments, the AI agent replies again, now using updated information—both from learner behavior and instructor input—to continue supporting the learner. This ensures that the AI agent's guidance remains aligned with both the learner's evolving needs and the instructor's educational intentions.

The cycle culminates in the instructor tells phase. Here, the instructor provides holistic, personalized feedback to the learner, possibly revising course structure, reinforcing key concepts, or guiding next steps based on cumulative insights from both human and AI perspectives. This closing action not only delivers meaningful guidance to the learner but also sets the stage for the next cycle of interaction and inquiry.

Altogether, this feedback loop demonstrates a model of shared responsibility between human instructors and AI agents. It places the learner at the center of the process, ensures that AI responses are integrated with expert human judgment, and maintains a continuous, data-informed cycle of adaptation. This dynamic interaction fosters more responsive, effective, and personalized learning experiences.

## Conclusion

The A4L framework provides a holistic and scalable foundation for implementing personalized learning in digitally mediated education, particularly for adult learners in online environments. By thoughtfully integrating artificial intelligence with human instructional expertise, A4L addresses the longstanding challenge of tailoring education to individual needs at scale—a goal that traditional classroom models often struggle to achieve. The core innovation of A4L lies in its data-centric design, which supports not just the technical handling of educational data but also the pedagogical processes required to translate those data into meaningful instructional interventions.

As depicted in Figure 1, the personalized learning framework begins with the continuous collection of rich multimodal learning data. These data are drawn from learners' interactions with educational tools—clickstreams, assessments, help-seeking behaviors, discussion forum activity, and AI tutoring logs. What distinguishes A4L is the way this data is not merely archived or analyzed in isolation but rather fed into a collaborative human-AI decision-making loop**.** Educators and AI agents operate as co-instructors, with the AI surfacing trends and predictive insights while the human educator interprets them through a pedagogical and contextual lens. This partnership allows for learning interventions that are not only timely and tailored but also grounded in an understanding of the learner's motivations, learning history, and goals.

Figure 2 complements this pedagogical vision by outlining the high-level system architecture that supports it. The A4L pipeline comprises multiple components: data sources, the data engine, an analytics pipeline, and a visualization pipeline. Data flows through a structured sequence of ingestion, standardization, anonymization, transformation, and storage. The architecture ensures that data is protected—particularly personally identifiable information—while remaining accessible for analytics and model training. Using cloud-based services, the analytics pipeline performs both near real-time and batch analyses, generating insights that are disseminated through interactive dashboards. These dashboards are role-sensitive, giving educators, learners, and researchers the tools to visualize, reflect, and act upon the data in ways that inform instruction and learning strategies. Crucially, the visualizations act as more than reports—they are interfaces for action, enabling instructors to adapt their teaching in response to continuously evolving learner needs.

Building on this infrastructure, Figure 3 introduces a cyclical model of human-AI collaboration that forms the operational core of A4L's instructional loop. The process begins as learners interact with instructional materials and AI agents, generating data that is analyzed by the AI system. These insights are then passed to instructors who offer targeted feedback and content adjustments. The AI, in turn, responds to learners' direct queries and recommends relevant learning pathways or resources. This ongoing exchange—between learners, AI agents, and instructors—results in a closed feedback loop where instructional decisions are not static or top-down but instead adaptive, context-sensitive, and personalized.

Together, these components reflect a new paradigm for instructional delivery—one in which teaching and learning are driven by data, shaped by algorithms, and guided by humans. Rather than positioning AI as a replacement for educators, A4L emphasizes shared agency**,** where human teachers maintain central authority over pedagogy while AI amplifies their capacity to respond to diverse learner needs in real time. This hybrid model ensures that personalization is not only efficient but also pedagogically sound and ethically grounded.

In sum, A4L demonstrates how a thoughtfully designed human-AI collaborative system can make personalized learning both technically feasible and instructionally meaningful. It advances the field by merging robust data infrastructure with intelligent analytics and responsive visualization, all within a framework. As educational systems continue to evolve in response to digital transformation and lifelong learning demands, the A4L framework offers a powerful blueprint for realizing the promise of AI-enhanced personalized education at scale.

## Acknowledgments

This research has been supported by grants from the US National Science (#2247790 and #2112532) to the National AI Institute for Adult Learning and Online Education (aialoe.org). We thank members of the A4L team for their contributions to this work including our 1EdTech colleagues Suzanne Carbonaro, Blaine Helmick, Tim Couper, and Andrea Deau.

## References

Luckin, R. 2018. *Machine learning and human intelligence: The future of education for the 21st century*. UCL Press.

Holmes, W., Bialik, M., & Fadel, C. 2019. Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Center for Curriculum Redesign.

Liu, C., & Koedinger, K. R. 2017. Data-driven personalized education. Journal of Learning Analytics.

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

Kulik, J. A. 2013. Adaptive learning technology: The promise and potential. Educational Psychology Review, 25(3), 279-289.

Roll, I., & Wylie, R. 2016. Evolution and revolution in artificial intelligence in education. International Journal of Artificial Intelligence in Education, 26(2), 582-599.

Goel, A., Thajchayapong, P., Nandan, V., Sikka, H., and Spencer Rugaber. *AI in Learning and Education: The Need for a Data Architecture*. Online Higher Education Administration: Strategies for Maximizing Returns and Improving Learning Outcomes, Kathleen S. Ives, Ray Schroeder, and Marie Cini (editors), Routledge Press.

Park, J. and Goel, A. Human-Centric Teaching at Scale in Online Education through Bidirectional Feedback in Human-AI Interaction. Paper presented at AAAI 2025 Spring Symposium for Current and Future Varieties of Human-AI Collaboration. San Francisco, CA, March 31-April 2