Active Learning Map: A Novel Assimilation Guided Adaptive Learning Platform

Suhan Roy[1], Jaswant Jupalli Prabhas[1], Praseeda[1], Tulika Saha[1], Prasad Ram[2], Sushree Behera[1], Srinath Srinivasa[1]

[1]International Institute of Information Technology Bangalore, India
[2]Gooru Learning, California, USA

suhan.roy@iiitb.ac.in
jupalli.jaswant@iiitb.ac.in
praseeda@iiitb.ac.in
tulika.saha@iiitb.ac.in
pram@goorulearning.com
sushree.behera@iiitb.ac.in
sri@iiitb.ac.in

Abstract: Identifying individualized assimilation patterns of students can play an important role in personalizing learning interventions and improving their learning outcomes. While studying from similar course materials, learners frequently exhibit significant variability in how they internalize and prioritize core concepts. Recognizing this variability helps in designing meaningful personalized interventions, which in turn can lead to greater learner engagement and outcomes. An effective way to represent assimilation patterns is in the form of a high dimensional data structure called a Polyline that captures the differentiated way in which a given student has assimilated knowledge across different topics in the course. In this work, we extend this representational concept to create a model free Deep Reinforcement Learning Framework that can provide personalized recommendations to learners based on their Polylines. This is then visualized on a two-dimensional progression space called an Active Learning Map that enables the assimilation patterns to be visually displayed and interpreted. We cluster learners based on assimilation pattern and learning style using a Gaussian Mixture Model and train a Conditional Generative Adversarial Network (CGAN) as an Assimilation Estimator to simulate how their Polylines evolve. Using this CGAN as an environment, we train a Deep Q-Network (DQN) to discover individualized recommendation policies. Simulations demonstrate that this CGAN based environment enables the DQN to find highly effective policies using a very small sample of real student data. Furthermore, we find our DQN recommendation engine consistently outperforms a random policy, a coherent resource-matching heuristic, and a Proximal Policy Optimization (PPO) algorithm. Our proposed system functions as an autonomous intelligent agent within the Learning Map and acts as a guide, helping learners to visualize exactly where they are among a complex web of topics and lessons, as well as providing the best resources to help the learners close their knowledge gap.

Keywords: Assimilation, Polyline, Adaptive Learning Systems, Deep Reinforcement Learning, Active Learning Map

1. Introduction

One of the key elements of learning that drives outcomes is assimilation or how learners internalize their learning (Ryan, Connell and Grolnick, 1992). Students who have taken the same course often display wide differences in what they consider key learnings and takeaways from the course. When adaptive learning environments factor learner assimilation in the way their learning interventions are designed, it results in meaningful learning and retention (Ausubel, 2000). However, in online learning environments, capturing assimilation patterns and representing them in a way that is amenable to being processed by algorithms is a key challenge. Many online learning environments adopt adaptive learning architectures that orchestrate personalized learning instructions to meet the unique needs of individuals. The crux of a good adaptive learning system is to design a recommendation system that sequentially outputs customized recommendations, leading to the best suitable learning path for every single learner. Some of the early approaches to adaptive learning implemented an ‘’averagian’’ model where the class average was used as a benchmark to calibrate individual learners. However, later work recognizes learner variability as a rule, rather than an exception, and emphasizes the need for understanding individual variability while designing learning interventions. Assimilation or internalization of learning is a latent phenomenon that is not directly observable. In this regard, constructivist learning (Bevevino, Dengel and Adams, 1999), where learners are required to periodically construct their own knowledge and theories based on what they have learned, offers a glimpse into underlying assimilation patterns. Using this approach, a representation called “Polyline” was introduced for representing a learner's individualized assimilation pattern (also called a “jagged profile'') across the different topics representing a course (Praseeda et al., 2025). While Polylines provide a computationally sound representation to reason with assimilation, there

is still a need to factor this into recommendation engines that can provide personalized learning interventions. This paper addresses the problem of designing personalized pedagogical interventions using high-dimensional learner Polylines to capture the non-linear, highly individualized nature of knowledge assimilation. While traditional predictive neural network models (Li et al., 2023) tend to regress to the mean and discrete frameworks like Bayesian Knowledge Tracing (Subramanian and Mostow, 2021) cannot map to continuous, high-dimensional spaces, we model these trajectories by training a Conditional Generative Adversarial Network (CGAN) as an Assimilation Estimator Model. To ground our work in real world educational dynamics, we introduce ATLAS (Assimilation Tracking from Lecture Assigned Summaries), our dataset consisting of independent, constructivist student summaries collected post-lecture over a graduate level course "Advances in NLP". We construct Polylines from these summaries to provide a high-resolution, longitudinal trace of continuous knowledge growth. To capture shifting assimilation patterns, a Gaussian Mixture Model (GMM) clusters students based on learning styles and academic inclinations. We then develop a model-free Deep Q-Network (DQN) engine to recommend personalized learning resources that maximize learning gains. Simulated evaluations against random and rule-based heuristic baselines demonstrate the policy's efficiency. Finally, we deploy the trained DQN model onto a two-dimensional progression space called an Active Learning Map, enabling visually interpretable tracking of student assimilation patterns and resource recommendations.

2. Related Work

2.1 Assimilation Patterns and Mental Models

Previously researchers have tried to map the assimilation of students in terms of understanding and representing their internalization of knowledge by constructing mental models. Constructing mental models of students has been explored in key disciplines such as mathematics, chemistry (Coll and Treagust, 2001), and in studies on how learners internalize abstract ideas about non-living entities (Stavy, 1988). However, there exist no such studies to understand the assimilation pattern of a large number of learners over an entire structured course and how assimilation of learners can change over the consumption of new topics and diverse learning materials.

2.2 Representation of Learner Profiles

A learner's profile represents their progress in terms of competencies or proficiencies across all the topics included in the course. Several previous works have chosen to represent learner's proficiencies as a vector of binary latent traits. Each latent trait here represents complete mastery in a particular topic as a discrete binary signal. Such signals are often erroneous and provide incomplete context about the learner's proficiency, as competency over a topic or skill can often be in terms of multiple levels of mastery and there is continuous room for improvement. Representation of these learning traits as continuous values provide more detail and granularity in terms of identifying where learners can improve. Item Response Theory (Lord, 1980) and its variants (Ackerman et al., 2003) use a continuous scale vector representation that has been widely used in designing Intelligent Tutoring Systems (Burns and Capps, 2013) that help learners to improve their skills and reach their target outputs (van der Linden, 2005). The continuous vector representation is a powerful idea because it translates abstract, qualitative human learning into a rigorous, m-dimensional space that we can navigate and optimize with Reinforcement Learning.

2.3 Reinforcement Learning in Education

Reinforcement Learning (RL) is a paradigm of Machine Learning that has been rapidly growing over the past few years and has found key applications in real-world problems. In contrast to supervised and unsupervised learning, RL involves an agent that learns through direct contact with its environment. The agent takes consecutive actions and earns rewards from its environment as feedback after every action and thus learns to optimize its path. Thus, in order to maximize its overall cumulative sum of expected rewards, the agent learns by acquiring knowledge and experience rather than learning from instructions. RL has found a wide variety of applications in Education and particularly in e-learning platforms as a framework for personalised education to enhance learning outcomes. Several RL methods, namely Markov Decision Process (MDP), POMDP, Markov Chain Monte Carlo (MCMC), and the widely popular Deep RL Framework have been successfully applied in learning-based scenarios (Fahad Mon et al., 2023).

2.4 Adaptive Learning Systems

Personalised recommendations of educational activities for learners enrolled in a course to improve their learning goals have been deeply explored in literature related to Intelligent Tutoring Systems (Chen et al., 2017). Several works have adopted the Contextual Multi-Armed Bandits framework (Manickam et al., 2017) for personalised resource recommendation owing to their ability to balance exploration (exploring new activities) and exploitation or recommending effective known content (Nguyen et al., 2023). On considering the state of the learner as a dynamic and evolving entity, researchers have chosen to model the problem as a Markov Decision Process (MDP). Researchers have used popular Deep RL algorithms such as DQNs and Actor-Critic to learn Instructional policies that can optimize for long-term rewards in terms of Higher assessment scores or Improved learning output.

3. Methodology

3.1 Jagged Learner Profiles and Active Learning Map

In our research, we extend the Polyline model to represent the current assimilation state of learners as a Jagged profile, which captures learning outcomes across multiple dimensions (Praseeda et al., 2025). A Learning Entity refers to any entity involved in the course, including learners, learning resources, quizzes or assessments, instructors, and Teaching Assistants (TA). An *LE* also includes a set of basis topics ***T*** as:

$$\boldsymbol{T} = \{T_1, T_2, T_3, \ldots, T_n\}$$

Every learning entity in the course is represented as a polyline across n fundamental syllabus topics, as illustrated in Figure 1.

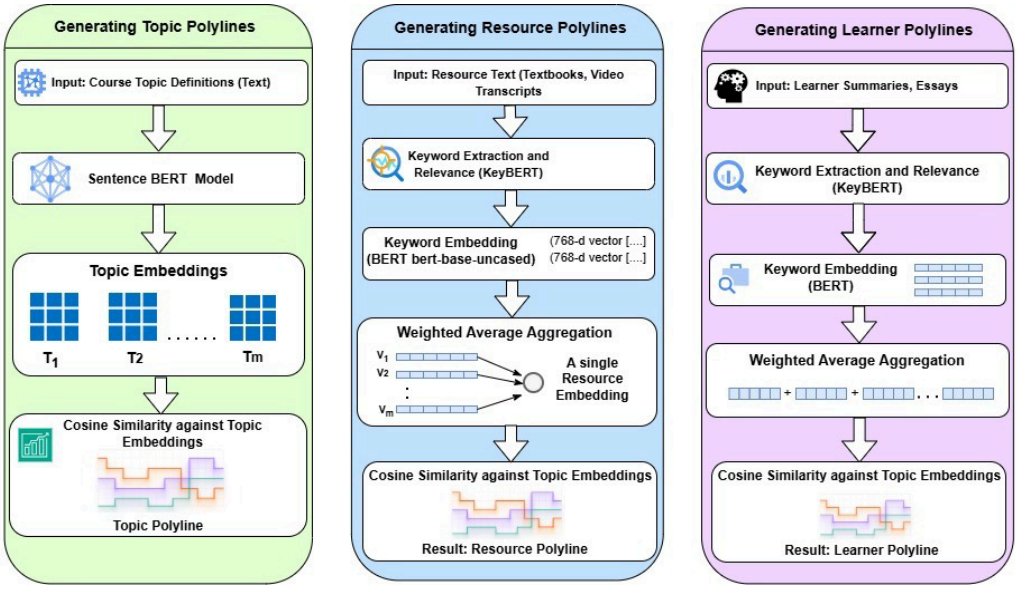


Figure 1: Creating Topic, Resource and Learner Polyline

Polylines in the Progression space follow the covering property. The covering property dictates that a specific Learning Entity *A* covers a distinct Learning Entity *B* provided that each individual dimensional coordinate within the Polyline of *A* maintains a value at least as high as its respective counterpart in the Polyline of *B*. The Highline for a group of students represents the highest achievable upper bound over the set of Basis Topics included in the course. The Highline of a course serves as a target vector that represents the highest proficiencies achievable over the Basis Topics for a new student who is enrolled in the course. The Highline of a course can be traced from the historical polylines of the learners enrolled in the course, or it can be set by the instructor or Teaching Assistant. The Active Learning Map functions as an interpretable two-dimensional space designed to visually track and navigate student progression throughout a course. At its core, the framework relies on high-dimensional vector representations called polylines, which quantify the mastery level of learning resources and individual learners across a fixed set of foundational concepts. To project these multi-dimensional vectors onto the two-dimensional map, we implement the radial mapping algorithm (Praseeda et al., 2025). In this geometric layout, each fundamental course topic is represented as a distinct radial axis emanating from a central origin.

The magnitude of a specific topic's mastery in the polyline vector determines the distance out along its corresponding axis and the algorithm computes a weighted sum of these directional components to compress the high-dimensional knowledge profile into a single, interpretable coordinate point. An intelligent recommendation engine then analyses a student's current position on this map to dynamically prescribe the next most effective learning resource, optimizing the trajectory toward comprehensive topic mastery.

3.2 An Assimilation Learning Dataset

Traditionally, Intelligent tutoring systems have relied on discrete scoring metrics such as lesson wise assessments and binary skill labels to estimate a learner's proficiency profile. While these models successfully use scalar scores to make recommendations, they inherently reduce the continuous, multidimensional aspect of knowledge assimilation into a single quantity. To address this gap, we introduce ATLAS (Assimilation Tracking from Lecture Assigned Summaries), a dataset curated from students enrolled in the Advances in NLP course that provides a granular, longitudinal trace of acquired knowledge. By capturing individual summaries after every lecture and resource interaction, our dataset records the progressive assimilation of technical concepts across the entire student cohort. The course curriculum includes 19 different Basis Topics and there is one learning resource for every topic. From the student summaries, we construct 19-dimensional Learner Polylines which serve as a backbone for understanding how the students internalize the curriculum over time.

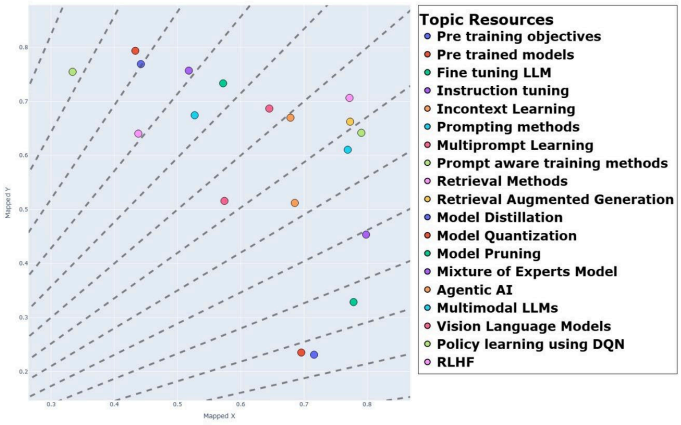


Figure 2 : Active Learning Map: High-dimensional Polyline representation projected into 2D progression space using the Radial Mapping Algorithm.

3.3 Formulating Adaptive Learning as a Markov Decision Process

The typical defining feature of DRL is to use deep learning to approximate reinforcement learning's value function and solve high-dimensional MDPs. The key components of our MDP are as follows:

1. State *s :* We define the current Polyline *PL* describing the learner's assimilation profile as the state.
2. Action *a :* Let the learning materials available in the adaptive learning system be indexed by 1,2,....*L*. The action space is $A = \{1,2,...,L\}$ and the action *a* in the adaptive learning system is the index of recommended learning resource which is a discrete value.
3. *Reward R :* To minimize the pedagogical steps required to reach course mastery, our reward function is strictly defined as the negative L_1 distance between the current Learner Polyline and the target Course Highline.

$$R(PL,a,HL) = - ||HL-PL||_1$$

By penalizing the agent at each step until the polyline aligns with the target highline, the environment frames routing as a shortest-path problem thus forcing the agent to minimize the knowledge gap in the fewest possible steps.

3.4 Clustering Learners based on Assimilation Patterns

To dynamically categorize and update learner profiles based on their Polyline, we use a GMM model. The objective of the clustering algorithm is to group the learners based on their interests and style of learning, based on information from their Polyline.

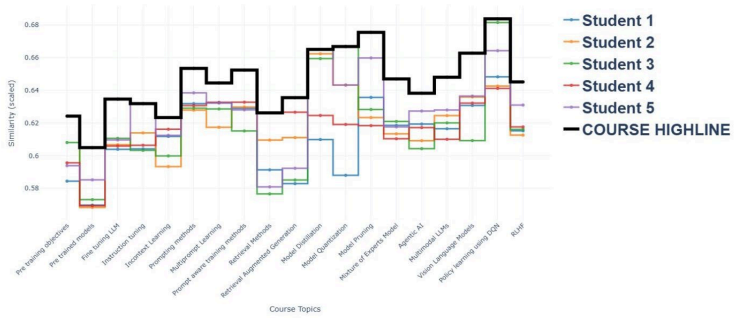


Figure 3: The Polyline representing the Highline for the course "Advances in NLP".

3.5 Conditional GAN as an Assimilation Estimator

Extensive research has demonstrated that neural networks are highly effective transition model estimators, capturing the complex, non-linear mappings required to precisely track and predict state transitions across learning trajectories (Li et al., 2023). To capture the non-deterministic nature of this progression, we employ a Conditional Generative Adversarial Network (CGAN) to estimate a probability distribution over potential future states rather than a single deterministic outcome (Mirza and Osindero, 2014). Trained on sequential logs from the ATLAS dataset, the network utilizes a supervised triplet format: an input polyline (current state), a condition (resource ID), and a label polyline (next state). By using the resource ID as a conditioning variable, the generator (G) learns to recognize underlying progression patterns to synthesize a targeted next polyline directly from the previous polyline based on the specific resource visited. Meanwhile, the discriminator (D) acts as a critic, comparing the generated paths against real student data from the ATLAS dataset to ensure the predicted steps are realistic and accurate.

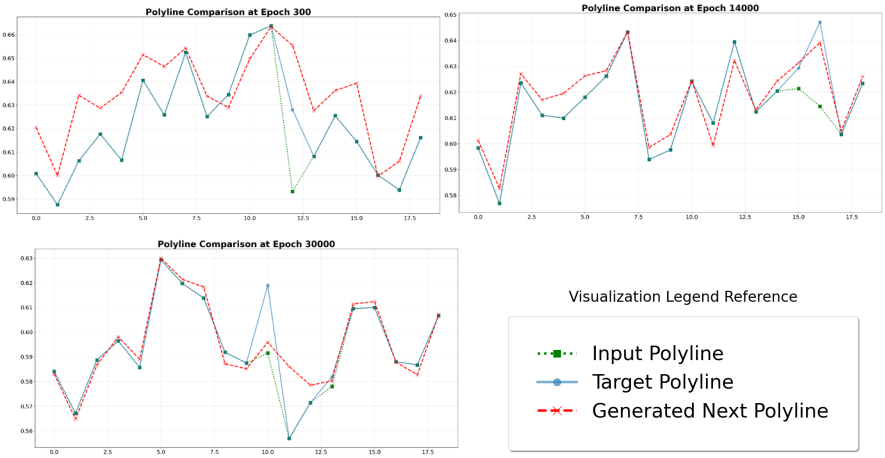


Figure 4: Our CGAN Model effectively learnt to understand and accurately predict assimilated knowledge states from our learner dataset.

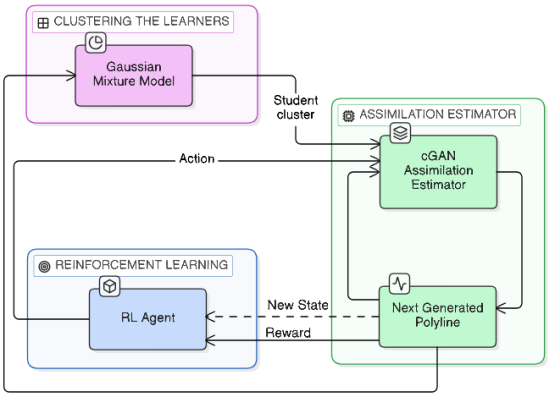


Figure 5: Diagram illustrates the proposed setup and Methodology to train our RL models.

3.6 Proposed Formulation to Train RL Algorithm

To discover novel pedagogical paths rather than relying purely on familiar sequences, the framework leverages distinct exploration mechanics that mirror instructional experimentation. The DQN engine employs an ϵ-greedy strategy, starting with high randomized exploration (ϵ = 0.9) to aggressively test diverse resource combinations before systematically decaying to a minimum floor (ϵ = 0.05) as it converges on a highly targeted, effective recommendations. Conversely, PPO explores inherently through a stochastic policy, sampling actions from a broad probability distribution over available resource. Early in training, this distribution is high-entropy, forcing the model to unroll widely divergent 50-step student journeys. As the engine accumulates compounding rewards, the policy sharpens around the optimal sequences. This robust exploration prevents both models from falling into local optimum traps such as repeating a predictable, basic sequence and forces them to discover hidden structural shortcuts across the continuous polyline space to find accelerated trajectories that minimize the knowledge gap in fewer steps.

4. Simulations Overview

To evaluate the performance of our trained models, we conducted 1000 independent simulation episodes per policy over a fixed 50 step horizon. At the beginning of each episode, the learner's initial knowledge state is represented as an all zero Polyline paired with a randomly selected starting cluster ID. To test the agents' capacity to optimize paths from arbitrary configurations, the first action at step (t=0) is forced to a random selection from the initial resource subset [0, 1, 2]. For all subsequent time steps, the Polyline state evolves dynamically based on policy recommendations. The PPO agent samples actions directly from its optimized stochastic policy, while the DQN agent follows an ϵ-greedy policy with a minimal exploratory baseline (ϵ = 0.1). We retained this nonzero exploration baseline during simulations to understand the performance of the engine to prepare for real world deployment on an e-learning platform where it must handle new learners. In production, real human behaviour and cognitive profiles inevitably deviate from synthetic simulations. A purely deterministic deployment risks faltering when encountering out-of-distribution (OOD) student states, potentially trapping the engine in suboptimal loops. Throughout these test trajectories, the agents interacted continuously with the independent CGAN environment simulator, which updated cluster profiles via the GMM and strictly enforced the unidirectional covering property to prevent knowledge regression. To measure policy efficiency, we tracked Total sum of cumulative negative L_1 distance rewards across each complete trajectory, directly quantifying how effectively each engine minimized the remaining knowledge gap against the target Course Highline.

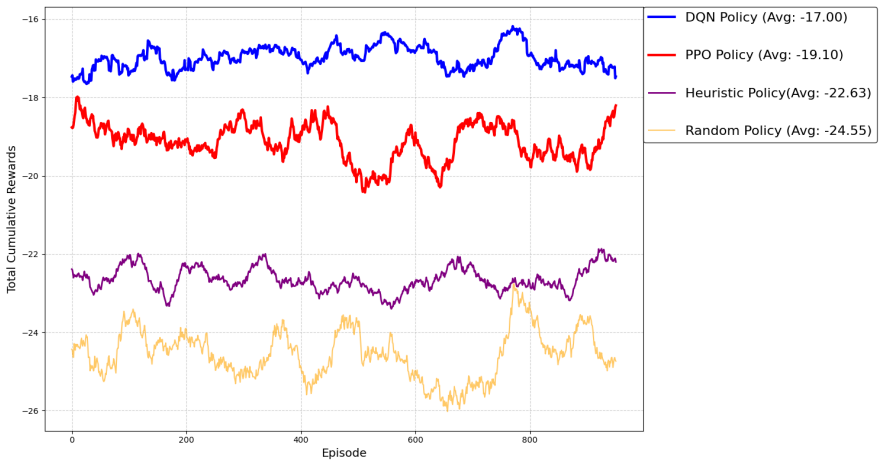


Figure 6: Smoothed rewards as learning gains over DQN, PPO, Heuristic and Random policies.

5. Results and Discussion

We evaluate our trained reinforcement learning policies against two baselines, an L_1 distance based heuristic and a Random recommendation policy. The heuristic policy selects semantically proximate resources to maintain coherence and prevent cognitive jumps, and the random recommendation policy serves as a robust baseline in Intelligent Tutoring Systems (ITS) where any curricular exposure can trigger baseline gains. Performance is

quantified using the cumulative sum of negative rewards, which measures the aggregate knowledge gap between the evolving Polyline and the target Course Highline. As this reward moves closer to zero (becomes less negative), it represents a policy that successfully maximizes learning gains by routing the student toward the Highline within the fixed horizon. A reward closer to zero (less negative) indicates an optimal, accelerated learning path that minimizes instructional friction and reaches the Highline in fewer steps. The DQN emerged as the best-performing policy, navigating these dynamics with the highest efficiency.

6. Active Learning Map

The student interface of the 2D Active Learning Map provides a comprehensive, interactive visualization of the learner's educational trajectory. The environment integrates a diverse array of

learning resources and instructional materials, which are spatially distributed across the map. The platform actively displays the learner's current positional coordinates and historical activity log, dynamically updating their state in real time as they engage with the curriculum through actions such as summarizing their learnings. Throughout their progression, students can continuously monitor their evolving Polyline, a visual representation that captures their continuous knowledge assimilation over time.

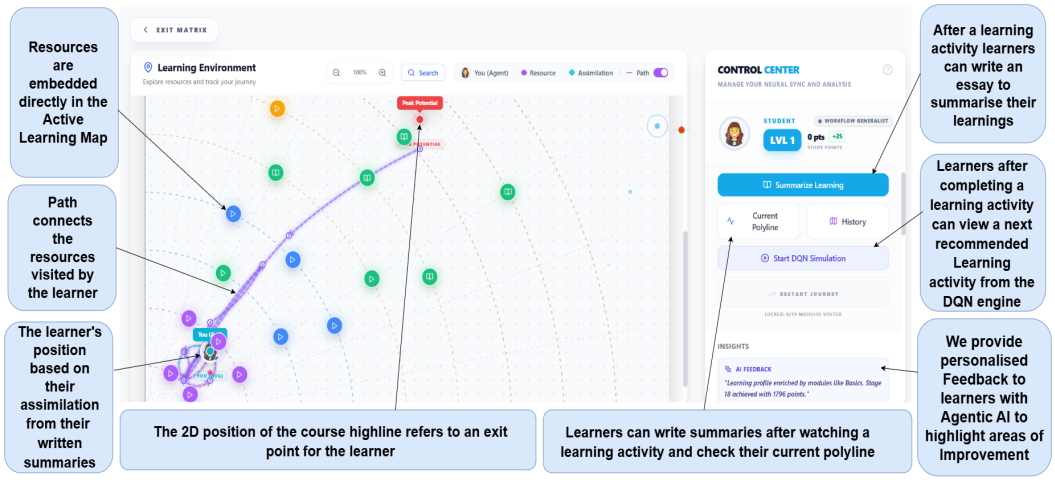


Figure 7: Our DQN Recommendation Engine can visualise and perform recommendations as a learning plan to learners on the Active Learning Map space.

7. Conclusion and Future Work

Understanding and representing assimilation is fundamental to fostering holistic learner transformation. In this work, we introduced a data driven framework for personalized adaptive learning centred on the Active Learning Map. By departing from rigid, traditional structures, our approach prioritizes holistic learning outcomes and captures the complex nature of knowledge assimilation that Adaptive Learning Systems overlook. However, while simulations validate the framework's technical soundness, it has not yet been tested with real learners. Consequently, our immediate future work focuses on integrating the Active Learning Map into a live classroom environment to empirically measure its impact on student engagement and learning outcomes. To capture the full cognitive reality of the student experience, we aim to expand the framework's capabilities by incorporating diverse learning artifacts such as coding assignments and quizzes and scaling the system across multiple courses. Ultimately, accurately modelling real-world assimilation dynamics will rely on collecting and analysing a larger corpus of student data to build increasingly robust machine learning models.

Ethics Declaration

This study was conducted in accordance with the ethical standards of research. All participants were informed about the purpose, and all the data was collected under the consent and supervision of the Course Instructor and Teaching Assistant.

AI Declaration

We used Gemini Pro to improve and refine sentence structure and to suggest alternative phrasings. All AI-generated suggestions were critically reviewed and edited by the authors, and the final content reflects the authors' own analysis and conclusions.

References

Ackerman, T.A., Gierl, M.J. and Walker, C.M., 2003. Using Multidimensional Item Response Theory to Evaluate Educational and Psychological Tests. *Educational Measurement: Issues and Practice*, 22(3), pp. 37–51. doi: 10.1111/j.1745-3992.2003.tb00136.x.

Ausubel, D.P., 2000. Assimilation theory in meaningful learning and retention processes. In: D.P. Ausubel, ed. *The acquisition and retention of knowledge: A cognitive view*. Dordrecht: Springer, pp. 101–145.

Bevevino, M.M., Dengel, J. and Adams, K., 1999. Constructivist theory in the classroom internalizing: Concepts through inquiry learning. *The Clearing House*, 72(5), pp. 275–278.

Burns, H.L. and Capps, C.G., 2013. Foundations of intelligent tutoring systems: An introduction. In: H.L. Burns and J.W. Parlett, eds. *Foundations of intelligent tutoring systems*. New Orleans: Psychology Press, pp. 1–19.

Chen, Y., Li, X., Liu, J. and Ying, Z., 2017. Recommendation System for Adaptive Learning. *Applied Psychological Measurement*, 42(4), pp. 243–257. doi: 10.1177/0146621617697959.

Coll, R.K. and Treagust, D.F., 2001. Learners' Mental Models of Chemical Bonding. *Research in Science Education*, 31(3), pp. 357–382. doi: 10.1023/A:1013159927352.

Fahad Mon, B., Wasfi, A., Hayajneh, M., Slim, A. and Abu Ali, N., 2023. Reinforcement Learning in Education: A Literature Review. *Informatics*, 10(3), p. 74. doi: 10.3390/informatics10030074.

Li, X., Xu, H., Zhang, J. and Chang, H.H., 2023. Deep Reinforcement Learning for Adaptive Learning Systems. *Journal of Educational and Behavioral Statistics*, 48(2), pp. 220–243. doi: 10.3102/10769986221129847.

Lord, F.M., 1980. *Applications of Item Response Theory to Practical Testing Problems*. 1st ed. New York: Routledge. doi: 10.4324/9780203056615.

Manickam, I., Lan, A.S. and Baraniuk, R.G., 2017. Contextual multi-armed bandit algorithms for personalized learning action selection. In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. New Orleans, LA, USA, 5-9 March 2017. IEEE Press, pp. 6344–6348. doi: 10.1109/ICASSP.2017.7953377.

Mirza, M. and Osindero, S., 2014. *Conditional Generative Adversarial Nets*. arXiv preprint arXiv:1411.1784. Available at: https://arxiv.org/abs/1411.1784.

Nguyen, D., Howard, L., Yan, H. and Lin, O., 2023. Adaptive Sequence Learning: Contextual Multi-armed Bandit Approach. In: *2023 IEEE International Conference on Dependable, Autonomic and Secure Computing, Pervasive Intelligence and Computing, Cloud and Big Data Computing, Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*. Abu Dhabi, United Arab Emirates, 14-17 November 2023. IEEE, pp. 0751–0756. doi: 10.1109/DASC/PiCom/CBDCom/Cy59711.2023.10361498.

Praseeda, Srinivasa, S. and Ram, P., 2025. Learning map: Representing and reasoning over diverse assimilation patterns of learners. *Next Research*, 2(4), p. 100804. doi: 10.1016/j.nexres.2025.100804.

Ryan, R.M., Connell, J.P. and Grolnick, W.S., 1992. When achievement is not intrinsically motivated: A theory of internalization and self-regulation in school. In: A.K. Boggiano and T.S. Pittman, eds. *Achievement and motivation: A social-developmental perspective*. Cambridge: Cambridge University Press, pp. 167–188.

Subramanian, J. and Mostow, J., 2021. Deep Reinforcement Learning to Simulate, Train, and Evaluate Instructional Sequencing Policies. In: *Proceedings of the 14th International Conference on Educational Data Mining (EDM 2021)*. Paris, France, 29 June - 2 July 2021. International Educational Data Mining Society, pp. 696–700. Available at: https://api.semanticscholar.org/CorpusID:236439662 [Accessed 21 May 2026].

tavy, R., 1988. Children's conception of gas. *International Journal of Science Education*, 10(5), pp. 553–560. doi: 10.1080/0950069880100508.

van der Linden, W.J., 2005. Item Response Theory. In: K. Kempf-Leonard, ed. *Encyclopedia of Social Measurement*. New York: Elsevier, pp. 379–387. doi: 10.1016/B0-12-369398-5/00452-7.