International Journal Of Advance Research, Ideas And Innovations In Technology

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 11, Issue 4 - V11I4-1155)
Available online at: https://www.ijariit.com

Mapping the Learning Path: Integrating Multi-Modal AI, Agentic AI, and MAB Reinforcement Learning to Create a Planner for Self-Paced Learning

Varun Iyer
varuniyer256@gmail.com
Symbiosis Skills and Professional University, Maharashtra

ABSTRACT

We present an intelligent, emotion-aware educational planner designed to deliver a highly personalized, self-paced learning experience. The system integrates task decomposition, graph-based reasoning, utility-driven scheduling, and affective state modeling into a cohesive, adaptive framework. High-level learner goals are broken down into actionable tasks via language models and structured into graph representations, which are analyzed using graph neural networks. A utility function—modulated by affective and cognitive state—guides task prioritization, while a contextual multi-armed bandit model dynamically schedules daily activities. The system also incorporates agentic AI to support user autonomy, mid-session adaptation, and long-term engagement. By merging planning, emotional intelligence, and explainable automation, this work proposes a modular architecture for learner-centric, holistic task management.

Keywords: *Emotion-Aware Scheduling, Utility-Based Task Management, Graph Neural Networks, Affective Computing, Human-In-The-Loop AI, Multi-Armed Bandit, Agentic AI, Personalized Planning Systems, Edge AI, Task Decomposition*

1 INTRODUCTION

Traditional, uniform instructional approaches often fall short in meeting the diverse learning needs, paces, and preferences of individual students, leading to disengage- ment and suboptimal outcomes [1]. Personalized learning, supported by advances in artificial intelligence (AI), offers a path toward educational experiences tailored to individual learners. Self-paced learning systems augmented with AI can move beyond static content delivery by dynamically adapting to learners' cognitive profiles, pro- gress, and challenges.

This paper proposes an intelligent self-paced educational planner that integrates multi-modal AI and agentic AI to provide holistic, adaptive support. The system de- composes learning goals into manageable tasks, prioritizes them through a utility- based scheduler, and personalizes study sessions in response to the learner's progress and affective state. Multi-modal inputs—including textual, visual, and auditory data—enable richer learner modeling, while agentic AI components offer proactive guidance and tool use.

Crucially, the system incorporates affective AI to account for the learner's emo- tional state—a key determinant of motivation and learning effectiveness [4]. By con- tinuously adapting based on both cognitive and emotional feedback, the proposed architecture aims to deliver a more engaging, empathetic, and effective learning expe- rience.

2 RELATED WORK

The proposed system draws on several intersecting lines of research, including per- sonalized learning systems, affective computing, task modeling via graph-based rea- soning, utility-based scheduling, and human-in-the-loop adaptation. Prior works have explored these components individually, but their integration into a cohesive, emo- tionally adaptive scheduling framework remains underexplored.

2.1 Affective Computing and Emotion-Aware Learning Systems

Recent years have seen increased interest in using affective signals to guide learning environments and digital assistants. Systematic reviews highlight the role of emotion- aware feedback in education, cognitive support, and user engagement modeling [2, 3, 6]. Techniques range from multimodal sensor-based emotion detection [8] to affective state modeling through passive, sensor-free methods [7]. Applications span immersive learning environments [3] and emotion-informed content delivery [2]. However, most such systems are either tutor-facing or pedagogically fixed, lacking real-time task adaptation grounded in emotional state.

2.2 Multimodal Learning Analytics and Task Decomposition

LLMs have recently been adapted for planning and analytics across multimodal inputs [4]. Multimodal learning analytics frameworks often focus on observational and post- hoc analysis rather than proactive scheduling [5]. Graph-based reasoning has also been explored to enhance LLM-based planning agents [9], but this work largely re- mains theoretical or sandboxed, without full integration into user-facing task pipe- lines. The proposed system builds on these approaches by directly transforming de- composed tasks into a graph structure for downstream prioritization and flow reasoning.

2.3 Utility-based Scheduling and Bandit Algorithms

Task recommendation and scheduling have frequently drawn from utility theory and reinforcement learning. Foundational work on multi-armed bandit (MAB) strategies [13] has informed numerous scheduling frameworks. Contextual bandits in mobile health and educational interventions have been shown to optimize just-in-time adap- tive decisions [12]. However, these systems often focus on single-domain goals and omit emotional state from reward modeling. The proposed framework extends these models by incorporating affect-modulated utility scores and using MABs for selecting emotionally aligned daily tasks.

2.4 Human-in-the-loop and Explainable Personalization

Human-in-the-loop machine learning has been positioned as a critical feature for trustworthy AI [10]. In assistive scheduling contexts, this ensures user autonomy and prevents over-optimization at the cost of cognitive comfort. Differentiating human-in- the-loop from AI-in-the-loop strategies [11] enables a shift from automation toward collaboration. The proposed system incorporates both passive (affective sensing) and active (user overrides) feedback, offering transparency in task selection through natu- ral language justifications and overrideable recommendations.

2.5 Systemic Gaps and Positioning

While the above domains have seen substantial development independently, few sys- tems have attempted to integrate real-time emotional state, graph-based planning, utility scoring, and contextual scheduling into a unified, deployable architecture. Moreover, existing models rarely consider deployment constraints such as on-device processing or privacy-preserving affect inference. This work addresses those gaps by proposing an edge-compatible, human-aligned task scheduler that learns not only from task success, but from the user's affective trajectory over time.

1 PROPOSED ARCHITECTURE

The proposed system is designed as a modular pipeline, where each component con- tributes to the goal of producing a daily schedule aligned with both the user's long- term objectives and their present emotional state. The system avoids rigid, pre-scripted tutoring and instead focuses on adaptive self-management through intelligent scheduling. Below are the core functional modules:

1.1 Task Decomposition System

The task decomposition engine is responsible for transforming high-level user goals into a structured set of actionable subtasks. This component is implemented using a language model-based agent capable of interpreting natural language goals and recursively generating task trees.

Each task is annotated with estimated effort, prerequisites, deadlines, and semantic tags (e.g., topic, format).

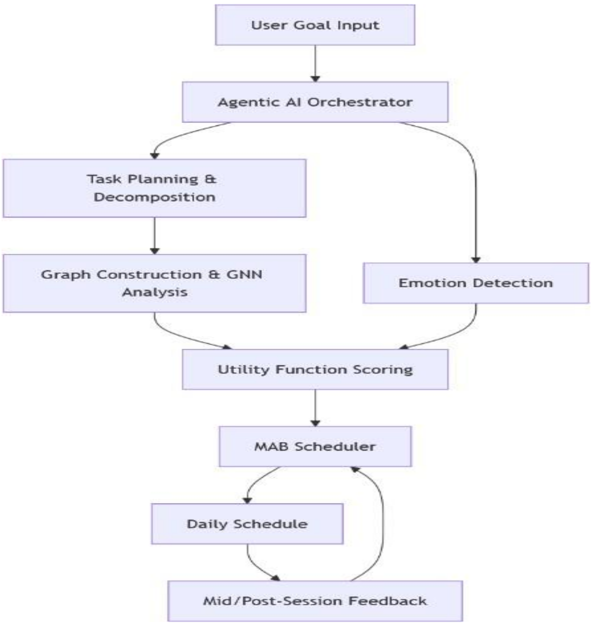

Fig. 1. High-level architecture of the proposed emotion-aware task planning and scheduling system.

1.1 Graph Construction and Analysis

Once tasks are generated, they are converted into a directed graph, where nodes represent tasks and edges denote dependencies, deadlines, or thematic relations. A Graph Neural Network (GNN) processes this task graph to:

i. Identify critical paths and bottlenecks
ii. Infer temporal ordering constraints
iii. Assess task centrality and blocking status

The GNN's output serves as a structured understanding of the workflow required to achieve the user's objective.

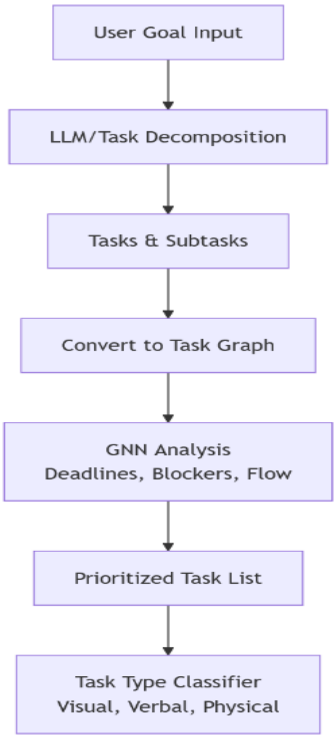

Fig. 2. Task decomposition flow using LLMs and graph neural networks for downstream pri- oritization and dependency reasoning.

1.2 Emotion Detection Module

To support adaptive scheduling, the system incorporates an emotion detection component that processes both objective signals (e.g., facial micro-expressions, voice tone) and subjective self-reports. The model outputs a low-dimensional affective state vector that represents user mood, energy, and stress level.
This emotional context is used to weight the perceived difficulty and desirability of upcoming tasks.

1.3 Task Utility Function

Each task is assigned a utility score that reflects its overall value and feasibility at the current moment. The utility function incorporates multiple features:

i. Task urgency and importance
ii. Estimated effort and duration
iii. Emotional compatibility (derived from the emotion detection module)
iv. Task type (e.g., visual, verbal, physical)
Utility can be represented as a scalar score or a multi-dimensional vector depend- ing on the policy model used.

1.4 Task Type Classifier

A lightweight classifier processes each task description to determine the interaction mode it requires — such as reading, writing, watching videos, or physical activity. This classification influences the utility function, allowing the system to defer high- cognitive-load or emotionally draining tasks if the user is not in an optimal state.

1.5 Multi-Armed Bandit Scheduler

The final selection of tasks is performed by a contextual Multi-Armed Bandit (MAB). The bandit model receives:

i. Utility scores (or vectors)
ii. Task duration estimates
iii. Available user time (provided by the scheduling agent)
iv. Recent reward history

The MAB selects a subset of tasks to recommend for the day, balancing between exploiting known effective scheduling patterns and exploring new ones. It is updated daily using a reward signal based on task completion and emotional outcome.

Each module is designed to operate independently but contribute to a shared feed- back loop. This modularity enables future refinement, replacement, or tuning of indi- vidual components without requiring a full system retraining.

2 UTILITY FUNCTION AND PLANNING MECHANISM

The core novelty of this system lies in its utility-aware, emotionally adaptive schedul- ing model. The system does not merely rank tasks by static priority, but instead uses a personalized utility function that incorporates both cognitive metadata and real-time affective state, enabling dynamic re-prioritization of tasks in response to the user's readiness and emotional bandwidth.

2.1 Utility Function design

Each task is assigned a utility score that reflects its present-day desirability and feasi- bility. This utility score is computed as a weighted function over several features:

i. **Urgency**: How soon the task must be completed.
ii. **Importance**: Its long-term relevance to the user's goals.
iii. **Effort**: Estimated cognitive load required.
iv. **Task Type Compatibility**: Match between the task format (e.g., visual, kinesthet- ic) and the user's current mood or energy state.
v. **Affective Alignment**: Degree to which the task is emotionally congruent with the
- user's pre-session affective vector.

A simplified version of the scalar utility function is as follows:

$$U_i = \alpha \cdot I_i + \beta \cdot R_i - \gamma \cdot L_i - \delta \cdot D_i \quad (1)$$

Where:

i. U_i: utility of task iii
ii. I_i: importance
iii. R_i: urgency or time sensitivity
iv. L_i: estimated effort or load
v. D_i: emotional dissonance or friction score
vi. $\alpha,\beta,\gamma,\delta$: user-defined or learned weight parameters

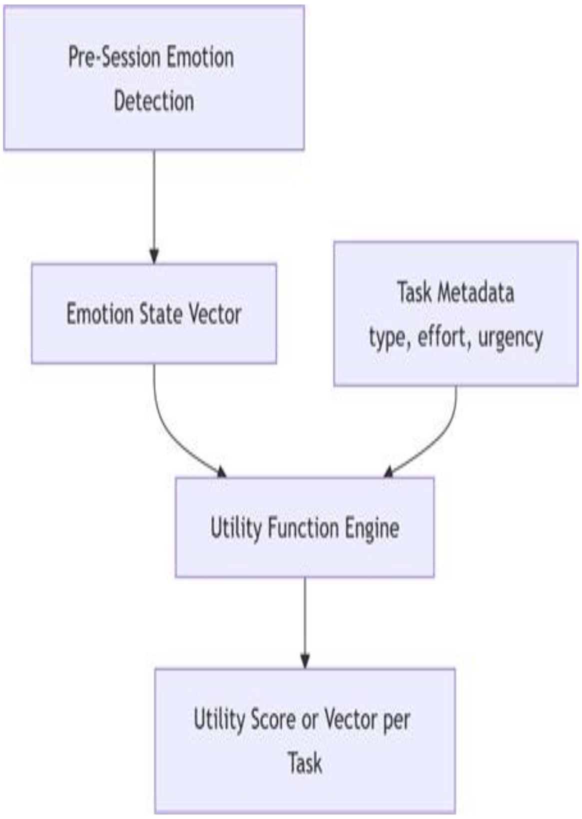

Fig. 3. Utility function inputs combining cognitive and affective factors to generate scalar or vector-valued task scores.

In practice, utility may be represented as a vector of multiple features and passed to a downstream policy model (e.g., Multi-Armed Bandit) that interprets the values in context. This allows the system to balance high-utility but emotionally taxing tasks with lower-stakes, low-friction ones, producing sustainable daily schedules.

2.2 Task Type Classification and Emotion Modulation

To support personalized adaptation, each task is classified according to its dominant modality — e.g., reading, writing, video watching, problem solving, or physical activ- ity. Based on user preference profiles and detected affective state (e.g., fatigue, anxie- ty, engagement), the system adjusts utility scores to deprioritize tasks that may cause friction under current conditions.

For example, if the user is cognitively depleted but emotionally stable, reading tasks may be deprioritized in favor of passive video-based tasks or light procedural activities.

2.3 Multi-Armed Bandit Scheduling

Final task selection is conducted by a contextual Multi-Armed Bandit (MAB) algo- rithm. The MAB receives, for each task:

i. The computed utility score or vector
ii. Estimated time to completion
iii. Available user time for the day (queried from the scheduler module)

The MAB selects a subset of tasks to populate the daily schedule, balancing between exploitation of high-utility knowns and exploration of less frequently chosen options to improve future policy accuracy.
After the day's session, emotional trajectory and task completion data are used to compute a reward signal, which is fed back to the bandit model to adjust task- selection probabilities over time.

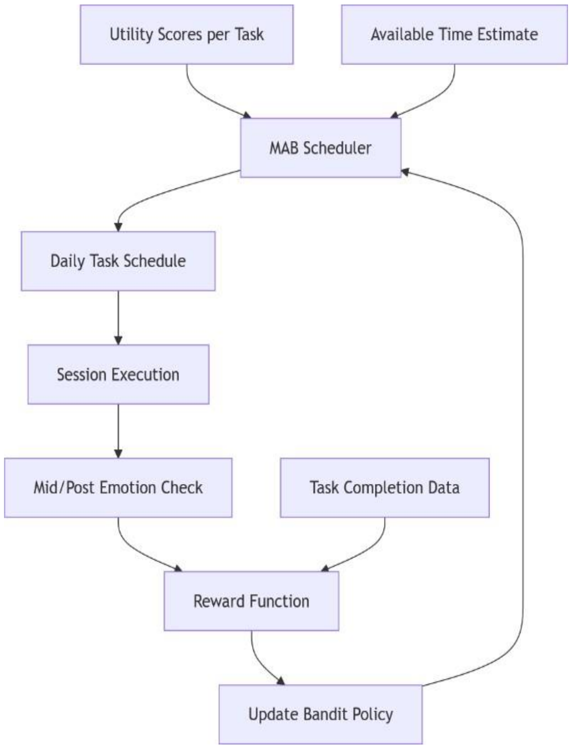

Fig. 4. Contextual multi-armed bandit scheduler selecting daily tasks based on utility scores and resource availability.

Adaptive Feedback and Policy Refinement

The system conducts mid-session and post-session emotional check-ins, along with task tracking. If negative affect is detected (e.g., frustration, fatigue), or if task non- completion is high, the utility function and bandit policy are adaptively adjusted. This ensures the system does not repeatedly recommend emotionally mismatched or cogni- tively unrealistic schedules.
Over time, this loop enables the system to converge toward task selections that bal- ance progress, cognitive load, and user emotional state, reinforcing not just productiv- ity, but sustainable engagement.

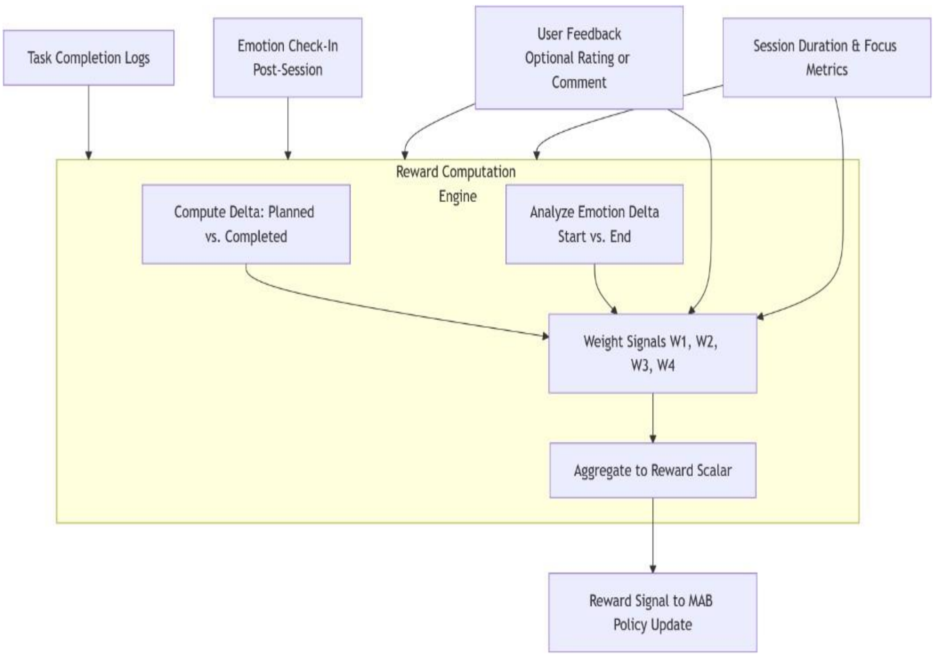

Fig. 5. Reward computation pipeline combining task completion, emotion feedback, and ses- sion metadata to update bandit policy.

1 HUMAN-IN-THE-LOOP AND EXPLAINABILITY

Despite the system's reliance on algorithmic scheduling and affective inference, user agency remains central to its design. Rather than fully automating daily planning, the system engages in a human-in-the-loop (HIL) feedback loop, allowing the user to guide, correct, and shape its behavior over time.

1.1 User-Driven Initialization and Overrides

The system begins each session by accepting a high-level goal or task list from the user, ensuring that the broader intent remains human-defined. At every scheduling stage, users retain the ability to:

i. Edit task metadata (e.g., perceived effort or priority)
ii. Remove or postpone tasks
iii. Override the system's proposed schedule, either partially or fully This ensures that the planner supports rather than dictates daily activity.

1.2 Affective Check-ins and Soft Feedback

Affective data is captured not only through sensor input (e.g., facial expression or voice) but also through periodic explicit user check-ins, such as mood sliders, energy ratings, or short reflective prompts. These inputs can:

i. Influence task selection directly
ii. Modify weightings in the utility function (e.g., discounting cognitive-heavy tasks when user reports fatigue)
iii. Be used to retroactively evaluate the quality of a schedule

This blend of passive sensing and active reflection supports more accurate affect modeling and preserves user autonomy in defining their state.

1.3 Adaptive Learning from Completion and Feedback

At the end of each session, users are prompted to provide light feedback on their ex- perience — e.g., whether they felt the session was productive, overwhelming, or mismatched with their energy levels. Alongside task completion logs and emotional check-ins, this feedback is used to:

i. Compute a scalar **reward or penalty**
ii. Update the Multi-Armed Bandit (MAB) policy for future task selection
iii. Calibrate the emotional compatibility model

This feedback loop encourages alignment between system recommendations and user preferences over time, improving personalization without requiring extensive manual configuration.

1.4 Explainability and Trust

All utility scores and scheduling decisions are made explainable to the user through natural language summaries. For example, a task might be marked as "deferred due to emotional mismatch" or "prioritized for its urgency and low effort."
This transparent reasoning supports user trust and allows for informed overrides — a critical component of ethical human-centered AI design.

Ethical safeguards are embedded in the system's architecture. Users maintain control over all scheduling outcomes, affective inputs are handled with privacy in mind, and no diagnostic inference is attempted. Emotion-based adaptation is fully optional, and reinforcement learning components are aligned to user-defined success criteria, not engagement metrics. These decisions are aimed at supporting trust, agency, and re- sponsible personalization.

2 CONCLUSION AND FUTURE WORK

This paper presented a modular, emotion-aware scheduling system designed to sup- port user-driven goal execution through adaptive task planning. By integrating task decomposition, graph-based prioritization, affective modeling, and utility-driven se- lection via a Multi-Armed Bandit framework, the system provides a personalized, explainable, and human-centered alternative to rigid productivity tools.
Unlike instructional agents or static planners, this architecture emphasizes cogni- tive and emotional alignment, adapting schedules in real time based on user readiness and response. The system remains fully overrideable, prioritizing autonomy and trust, and is explicitly designed for sustained engagement rather than task maximization.
While portions of this architecture may intersect with ongoing research in educa- tional and assistive AI, the overall system design—particularly its integration of emo- tional state, agentic scheduling, and utility-based reasoning—is original to the author. It is reserved for formal expansion in doctoral work, and should not be reproduced without attribution.
We believe systems like this signal a shift toward AI that scaffolds human agency rather than substituting for it—an architecture not of automation, but of alignment.

ACKNOWLEDGEMENTS

This paper is submitted as part of the author's academic work toward the completion of an M.Sc. degree at Symbiosis Skills and Professional University. The author grate- fully acknowledges the support of their institution in facilitating this research and publication process.

REFERENCES

[1] Auer, P., Cesa-Bianchi, N., Fischer, P.: Finite-time analysis of the multi-armed bandit problem. *Machine Learning* **47**, 235–256 (2002)

[2] Yuvaraj, R., Mittal, R., Prince, A.A., Huang, J.S.: Affective computing for learning in ed- ucation: A systematic review and bibliometric analysis. *Education Sciences* **15**(1), 65 (2025)

[3] Lampropoulos, G., Fernández-Arias, P., Antón-Sancho, Á., Vergara, D.: Affective compu- ting in augmented reality, virtual reality, and immersive learning environments. *Electron- ics* **13**(15), 2917 (2024)

[4] Davalos, E., et al.: LLMs as educational analysts: Transforming multimodal data into ac- tionable reading assessment reports. *arXiv preprint* arXiv:2503.02099v1 (2025)

[5] Martínez-Maldonado, R., Echeverría, V., Fernández-Nieto, G., et al.: Lessons learnt from a multimodal learning analytics deployment in-the-wild. *ACM Trans. Comput.-Hum. Inter- act.* **31**(1), 8 (2023)

[6] Vistorte, A.O.R., Deroncele-Acosta, A., Martín Ayala, J.L., et al.: Integrating artificial in- telligence to assess emotions in learning environments: A systematic literature review. *Front. Psychol.* **15**, 1387089 (2024)

[7] de Morais, F., Goldoni, D., Kautzmann, T., da Silva, R., Jaques, P.A.: Automatic sensor- free affect detection: A systematic literature review. *arXiv preprint* arXiv:2310.13711 (2023)

[8] Rouast, P.V., Adam, M.T.P., Chiong, R.: Deep learning for human affect recognition: In- sights and new developments. *arXiv preprint* arXiv:1901.02884 (2019)

[9] Wu, X., Shen, Y., Shan, C., et al.: Can graph learning improve planning in LLM-based agents? In: *NeurIPS Workshop on Graph Representation Learning* (2024)

[10] Munro, T., Holmberg, C.: Human-in-the-loop machine learning: A state of the art. *Soft Computing* **26**, 675–700 (2022)

[11] Wang, H., Moore, J.: Human-in-the-loop or AI-in-the-loop? Automate or collaborate?
i. *arXiv preprint* arXiv:2412.14232 (2024)

[12] Murphy, S.A., Tewari, A., Nahum-Shani, I.: Just-in-time adaptive interventions via con- textual bandits in mobile health. *Ann. Behav. Med.* (2016)

[13] Laak, K.–J., Aru, J.: AI and personalized learning: Bridging the gap with modern educa- tional goals. *arXiv preprint* arXiv:2404.02798 (2024)