RESEARCH ARTICLE

Deployment Challenges of Bayesian Network–Based Learning Path Recommendation in Real Classroom Settings

Muhammad Talha Mansoor [1,*]

[1] Department of Computer Science, Virtual University of Pakistan, Pakistan

Abstract

Bayesian Network (BN)–based Learning Path Recommendation (LPR) systems are widely adopted in personalized education for modeling uncertainty and providing interpretable learner representations. However, existing studies predominantly evaluate these systems under controlled settings that assume balanced data, simplified curricula, and unconstrained resources. Consequently, limited empirical understanding exists regarding their performance in authentic classrooms. This study addresses this gap by examining the real-world deployment of a BN-based LPR system using naturally occurring classroom data. The system is evaluated using 426,004 quiz responses collected across 19 formative assessments, reflecting routine instructional practice. Two curricular configurations were examined: an independent-topic structure with no explicit dependencies, and a prerequisite-based structure where mastery depended on prior knowledge. Rather than focusing on predictive accuracy, the study investigates operational feasibility, including structural complexity and memory consumption. From a software-engineering perspective, the central problem addressed is the scalability and inference feasibility of BN-based LPR systems under curriculum dependencies and real-world resource constraints. Results show that BN inference remains stable and computationally tractable for independent topics. In contrast, prerequisite-based modeling substantially increases network density, memory requirements, and inference variability, in some cases causing deployment failures under standard resources. Sparse assessment coverage at higher cognitive levels further undermines recommendation stability. This work provides a deployment-oriented perspective that complements performance-centric research and emphasizes the need for deployment-sensitive evaluation beyond accuracy-focused assessment in real classrooms.

Keywords: bayesian networks, learning path recommendation, real-world deployment, curriculum dependency, adaptive learning systems, educational data mining, computational feasibility.

Submitted: 23 December 2025
Accepted: 08 February 2026
Published: 16 March 2026

Vol. 2, **No.** 1, 2026.
10.62762/JSE.2025.490810

***Corresponding author:**
Muhammad Talha Mansoor
talha131@gmail.com

Citation
Mansoor, M. T. (2026). Deployment Challenges of Bayesian Network–Based Learning Path Recommendation in Real Classroom Settings. *ICCK Journal of Software Engineering*, 2(1), 71–84.

1 Introduction

Bayesian Network (BN)–based Learning Path Recommendation (LPR) systems are widely used for personalized education due to their ability to model uncertainty and support interpretable reasoning. However, most existing studies evaluate these systems under controlled or simulated conditions, assuming balanced assessment data, simplified curricular structures, and modest computational demands. Consequently, there is limited empirical understanding of how such systems behave when deployed using naturally occurring classroom data, where assessments are uneven, prerequisite dependencies exist, and computational resources are constrained [1].

Learning path recommendation (LPR) systems are adaptive educational systems designed to guide learners through instructional content by suggesting appropriate sequences of topics or activities based on observed learning evidence. Their primary purpose is to support personalized progression by determining whether a learner should advance to new material or revisit prerequisite concepts [2, 3]. Typical LPR systems integrate three core components: (i) a representation of curriculum structure and topic relationships, (ii) a learner model that estimates latent knowledge or mastery states from assessment evidence, and (iii) a decision mechanism that maps inferred learner states to instructional recommendations. Bayesian Networks are commonly adopted for this purpose because they provide a principled probabilistic framework for modeling uncertainty in learner knowledge, handling incomplete assessment data, and encoding expert-defined prerequisite relationships in an interpretable manner [4].

In authentic educational settings, assessment data are often unevenly distributed across topics and cognitive levels. Curricula frequently include prerequisite dependencies that introduce structural coupling between learning concepts. In addition, deployment environments are subject to practical runtime and memory constraints [5, 6]. This gap is particularly relevant for LPR systems that aim to provide actionable instructional guidance, such as suggesting whether a learner should proceed to a new topic or revisit prior material [7, 8]. In practice, such recommendations must be generated reliably and within instructional timeframes. However, the interaction between curriculum structure, assessment design, and probabilistic inference complexity remains underexplored in real deployment contexts [9–11].

The present study addresses this gap by examining the practical challenges encountered during the deployment of a Bayesian Network–based LPR system in a real classroom environment. Rather than proposing a new modeling technique or optimizing predictive performance, this work focuses on operational feasibility, including structural complexity, memory consumption, inference stability, and sensitivity to sparse assessment evidence.

From a software engineering perspective, these challenges reflect design trade-offs related to system scalability, architectural complexity, and runtime feasibility when deploying AI-based educational systems. Using large-scale quiz data collected during routine instructional activities, the study contrasts system behavior under independent topic structures and prerequisite-based topic dependencies. Through empirical analysis supported by descriptive statistics and computational observations, the work documents the dependency modeling impacting the inference cost, recommendation stability, and real-time applicability. This paper makes the following contributions:

1. It provides an empirical deployment study of a Bayesian Network–based LPR system using large-scale, real classroom data.
2. It analyzes the impact of curriculum structure by contrasting independent and prerequisite-based topic configurations, highlighting their effects on structural complexity and inference feasibility.
3. It reports practical deployment constraints observed during real classroom use, including increased memory requirements, unstable recommendation behavior, and system failures that arise when curriculum dependencies are introduced.
4. It offers deployment-oriented insights that complement performance-centric studies and inform the practical design of probabilistic adaptive learning systems.

By foregrounding deployment-level evidence and explicitly reporting limitations, this study contributes a practice-oriented perspective to the literature on adaptive learning technologies. The findings aim to support more realistic design choices and responsible adoption of Bayesian Network–based LPR systems in authentic educational environments. Figure 1 presents a side by side comparison of evaluation conditions.

One side shows commonly assumed settings such as balanced assessments, independent topics, and unrestricted computational resources. The other side represents classroom deployment conditions where assessment coverage is uneven, topic relationships are prerequisite driven, and computational resources are limited.

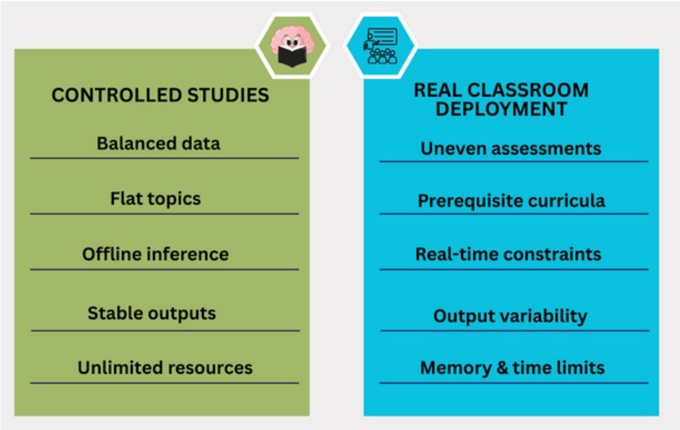

Figure 1. Conceptual comparison between common evaluation assumptions in Bayesian Network–based LPR research and the conditions encountered during real classroom deployment.

The remainder of this paper is organized as follows. Section 2 provides background on LPR systems and reviews related work on BN–based approaches. Section 3 describes the study context, data collection process, and deployment configuration of the BN–based LPR system. Section 4 presents deployment-level empirical results, focusing on computational feasibility, inference behavior, and observed limitations. Section 5 discusses the software engineering implications and design lessons derived from these findings. Finally, Section 6 concludes the paper and outlines directions for future work.

2 Background and Related Work

Personalized LPR has become an active area of research in educational data mining and learning analytics, driven by the increasing availability of fine-grained learner data and advances in probabilistic modeling [12]. Among the proposed approaches, Bayesian Networks (BNs) are frequently adopted due to their ability to represent uncertainty, incorporate expert knowledge, and provide interpretable reasoning about latent learner states [13, 14]. Recent research on LPR has largely concentrated on the development of algorithmic strategies and offline evaluation using curated or benchmark datasets. As summarized in Table 1, most existing studies emphasize personalized path generation under controlled or simulated conditions, with limited attention to real classroom deployment or operational feasibility. Although some works incorporate real educational data at the dataset level, explicit analysis of deployment behavior, computational constraints, and inference stability in authentic classroom settings remains largely absent [15–17]. In contrast, the present study focuses on deployment-level feasibility using naturally occurring classroom data, thereby addressing an underexplored aspect of Bayesian Network–based LPR systems.

Among the various modeling approaches, Bayesian Networks (BNs) are frequently adopted because they can represent uncertainty while encoding expert-defined dependencies about latent learner knowledge. These properties make BNs well suited for educational settings, where assessment evidence is often incomplete and pedagogical relationships between topics are explicitly defined.

Despite their theoretical appeal, most BN–based LPR systems are evaluated under controlled or simulated conditions. These evaluations often assume balanced assessment data, simplified curricular structures, and unconstrained computational resources. Overall, the literature demonstrates strong progress in learning path generation algorithms and modeling techniques, but places limited emphasis on deployment-level evaluation. Prior studies rarely report computational feasibility, memory requirements, or inference stability under realistic classroom constraints. This limitation motivates the present work, which shifts the focus from algorithmic performance to the practical feasibility of deploying Bayesian Network–based LPR systems in authentic educational environments.

3 Methodology

3.1 Study context and deployment setting

This study was situated within a real undergraduate classroom environment, where formative assessment was already part of routine instructional practice. The primary aim was not to design or optimize a new learning system, but to observe the practical behavior and constraints of deploying a Bayesian Network–based LPR approach under realistic educational conditions.

The course content was delivered over multiple instructional units. Topics varied in structure: some were conceptually independent, while others followed explicit prerequisite relationships determined by the

Table 1. Summary of recent LPR studies (2023–2025), highlighting the evaluation settings, data characteristics, and the extent to which real classroom deployment and operational feasibility are addressed.

Ref.	Year	Approach / Method	Primary Task	Data Setting	Real Deployment / Real Data?	Real Classroom Data
[18]	2025	Hybrid AI path recommendation (Nestor)	Personalized LPR	Adaptive learning environments	Not specified	No
[19]	2025	Knowledge Tracing + LLM-enhanced LPR	LPR with multi-step prediction	Public benchmark datasets	No	No
[20]	2023	Knowledge graph + GCN	Personalized learning path generation	Real educational data	Yes (dataset level)	Partial
[21]	2024	Process-oriented LPR	Learner-centric path planning	Experimental / research setting	No	No
[22]	2025	Multi-algorithm collaborative LPR	LPR	Online learning platforms	No	No
[23]	2025	Rule-based personalized LPR system	LPR	Not reported	No	No
[24]	2025	Multimodal intelligent LPR	Personalized adaptive learning paths	Mixed modalities	Not specified	No
[25]	2024	Explainable causal knowledge networks	Knowledge LPR	Research evidence	No	No
[26]	2024	Graph-based personalized LPR	Adaptive learning sequence generation	Educational datasets	No	No
[27]	2024	Deep reinforcement learning–based LPR	Personalized curriculum sequencing	Simulated / curated datasets	No	No
This study	**2025**	**Bayesian Network–based LPR**	**Deployment-level feasibility analysis**	**Naturally occurring classroom data**	**Yes**	**Yes**

course syllabus. Importantly, no changes were made to instructional sequencing or assessment frequency for the purpose of this study. The system was evaluated using naturally occurring educational data, thereby reflecting authentic classroom dynamics rather than controlled experimental conditions. Figure 2 summarizes the deployment workflow of the Bayesian Network based learning path recommendation system. Quiz responses generated during routine classroom assessments are collected and mapped to instructional topics and cognitive levels, forming evidence for learner state estimation. Bayesian inference is then performed incrementally as new evidence is observed, and topic mastery beliefs are updated under practical runtime and memory constraints. Based on these inferred states and the defined curriculum structure, learning path recommendations are produced to guide progression or revision decisions.

3.2 Assessment design and data collection

Student learning evidence was collected through a series of low-stakes multiple-choice quizzes administered during the course. Each quiz item was manually mapped to a specific instructional topic and assigned a cognitive level according to the revised Bloom's Taxonomy [28, 29]. Responses were recorded automatically through an online assessment platform and later exported for analysis. To ensure privacy, all personally identifiable information was removed and replaced with anonymized identifiers prior to processing. Student responses were encoded in binary form (correct or incorrect), mirroring common formative assessment practices used in higher education.

No artificial balancing of item difficulty, response frequency, or cognitive-level distribution was performed. This decision was intentional. The

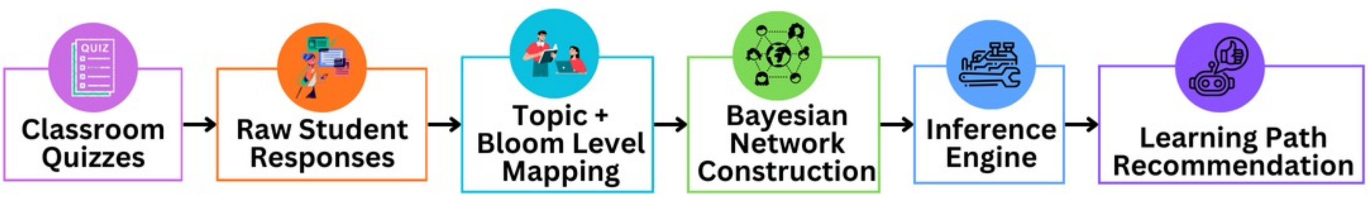

Figure 2. An Overview of the deployment workflow for the Bayesian Network–based LPR system.

objective was to examine deployment-level challenges that arise when probabilistic learning models operate on imperfect, uneven, and pedagogically driven data, as is typical in real instructional settings. Topics were categorized into two groups: independent topics, which could be learned without prior conceptual dependencies, and dependent topics, which required prerequisite knowledge from earlier instructional units. Prerequisite relationships were defined using official course documentation and instructor specified learning objectives. A directed prerequisite link was assigned only when prior topic mastery was explicitly required for subsequent instruction, as stated in the syllabus. These relationships were encoded as fixed directed edges between topic nodes in the Bayesian Network and applied uniformly across all learners, ensuring reproducibility across deployments.

3.3 Experimental Setup

The Bayesian Network–based LPR system was implemented in Python using the BN Learn library for structure learning, parameter estimation, and inference. Experiments were executed in both cloud-based and local computing environments to assess deployment feasibility under differing memory constraints BN nodes represented topic-level mastery and assessment evidence, with directed edges encoding prerequisite relationships where applicable. Inference was performed incrementally as new quiz evidence became available. No alternative baseline models were evaluated, as the focus of this study was deployment feasibility rather than comparative predictive performance.

3.4 BN deployment configuration

A BN framework was employed to model latent learner knowledge states and observable assessment outcomes. Nodes represented student mastery indicators and quiz evidence, while directed edges encoded conditional dependencies derived from topic structure. Building on the experimental setup, the following subsection details the structural configuration of the deployed BN The BN used in this study adopts a topic-centric learner modeling structure that is commonly employed in LPR systems. Network nodes represent two elements: latent topic-level mastery states and observable assessment evidence. Each topic node models the probability that a learner has achieved mastery of a specific instructional topic. Evidence nodes correspond to quiz outcomes that are mapped to both the relevant topic and its associated cognitive level. Directed edges encode conditional dependencies derived from curriculum structure.

Two network configurations were examined. In the independent-topic configuration, topic nodes are not connected, producing a sparse network in which inference remains localized. In the prerequisite-based configuration, directed edges are introduced between topic nodes to represent syllabus-defined prerequisite relationships. This design allows mastery beliefs to propagate across related topics during inference. The configuration was intentionally chosen to reflect authentic instructional design rather than to maximize model expressiveness. Dependency depth was limited, and prerequisite links were defined by instructors, to preserve interpretability and reproducibility. These design choices also ensured alignment with real classroom deployment constraints, where computational resources and system transparency are critical.

The focus of this study was not the proposal of a novel Bayesian architecture. Instead, an existing BN implementation was deployed using the BN Learn library [30]. Network structure learning, parameter estimation, and inference were executed within this BN Learn-based configuration. The system was examined through an observation-driven methodology, focusing on changes in computational cost, stability of recommendation outputs under limited evidence, behavioral differences between independent and dependent topic structures, and feasibility of generating recommendations within instructional timeframes. Traditional predictive accuracy metrics were intentionally de-emphasized. The purpose of this study was not to demonstrate superior prediction, but to identify operational limitations and design trade-offs encountered during real classroom deployment.

3.5 Methodological scope, constraints, and ethical considerations

To maintain methodological clarity and avoid overclaiming, several explicit constraints were imposed. Baseline recommender systems were not included. They are commonly evaluated under offline and performance centered settings that abstract runtime limits and memory use. In this study, the outcomes of interest were system level behaviors, including inference stability, execution failure, and resource consumption under incremental evidence. Under such conditions, baseline comparisons would not provide meaningful contextual insight,

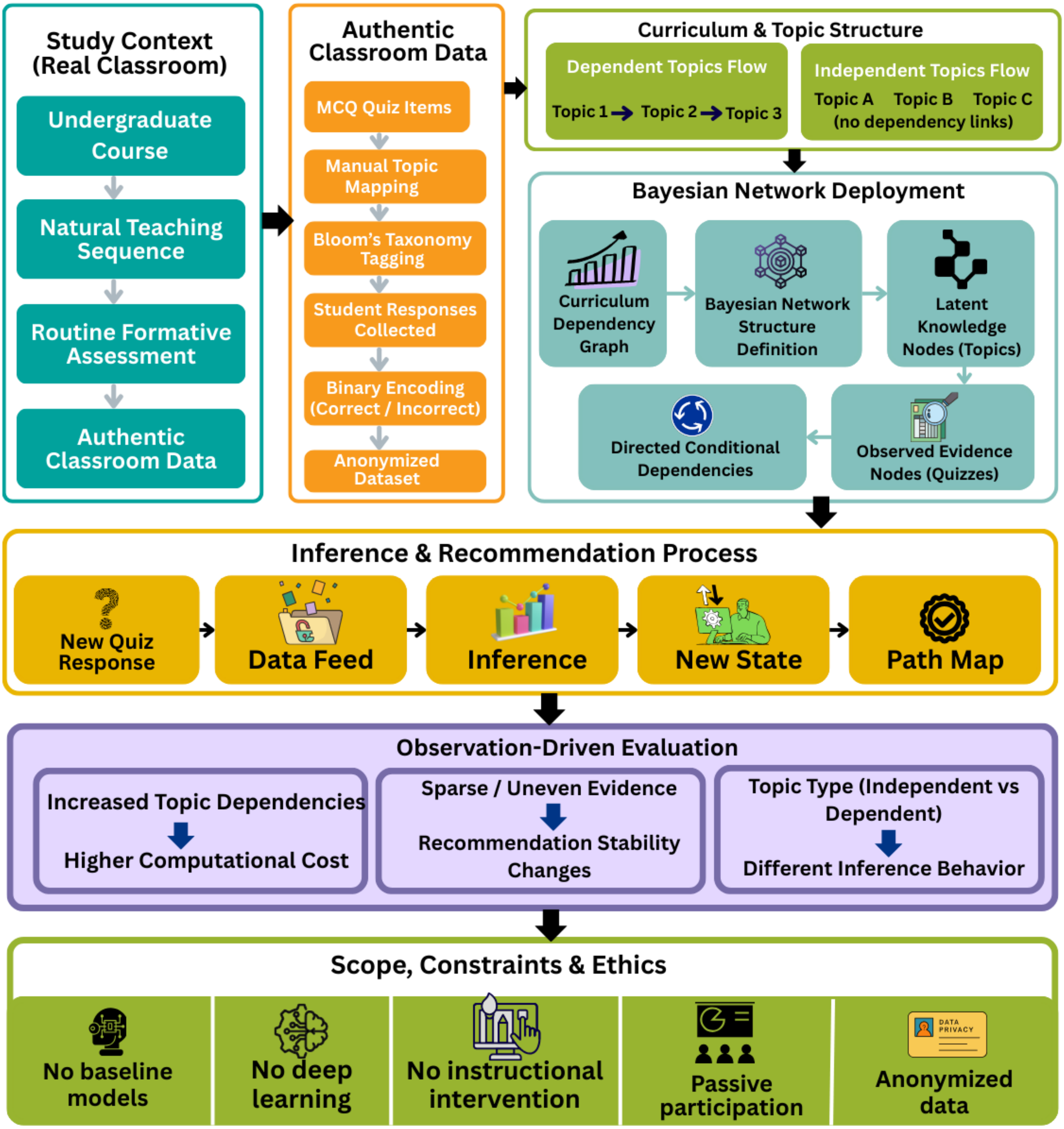

Figure 3. A Comprehensive overview of the real-world deployment and evaluation framework for the BN–based LPR system.

since they are not designed to expose deployment related constraints [31]. All data used in this study originated from routine educational activities. Student participation was passive, and instructional decisions were not influenced by system outputs. Data were anonymized prior to analysis, and findings were reported at the system level rather than the individual learner level. Figure 3 illustrates the end-to-end deployment and evaluation pipeline, showing classroom assessment data, curriculum structure, BN inference, and computational constraints interacting during real-world system operation.

4 Results

This section presents deployment-level empirical results obtained from implementing a BN–based LPR system in a real classroom environment. The results emphasize data scale, structural complexity, computational feasibility, and operational stability, rather than predictive accuracy.

4.1 Dataset scale and assessment characteristics

The system was deployed on a large, naturally occurring educational dataset comprising 426,004 individual quiz observations collected from 19 quizzes, each containing 40 multiple-choice questions. Assessments spanned six Bloom's Taxonomy levels and covered both independent topics and topics with explicit prerequisites. Table 2 highlights two important deployment characteristics. First, the large number of quiz observations combined with uneven coverage across Bloom's Taxonomy levels reflects realistic instructional practice rather than experimental control. Second, the presence of both independent and prerequisite-based topic structures shows that the system was evaluated under structurally heterogeneous conditions. These differences directly affected inference behavior during deployment.

Table 2. Dataset and deployment scale characteristics.

Aspect	Observed Value
Number of quizzes	19
Questions per quiz	40
Total quiz observations	426,004
Assessment type	MCQs (practice and graded)
Bloom's Taxonomy levels	6 (Remember → Create)
Topic structure	Independent and dependent
Maximum topic depth	2 levels (single prerequisite)
Cross-validation strategy	Stratified group k-fold (k = 5)

4.2 Behavior under independent and prerequisite-based topic structures

When instructional topics were modeled as independent entities, the system demonstrated stable and predictable behavior. Bayesian inference could be performed efficiently, even as new quiz evidence was incrementally incorporated. In this study, consistent recommendations were defined operationally. A recommendation was considered consistent if the suggested instructional action for a learner remained unchanged across successive inference steps when small amounts of new assessment evidence were added. This criterion was applied uniformly across learners and topic structures during incremental inference. Even with sparse responses, the probabilistic framework was able to maintain reasonable confidence in learner state estimation. This stability suggests that BN–based recommendation is operationally feasible for curricula where topics are weakly coupled or conceptually isolated.

A Conceptual comparison of BN inference behavior under independent and prerequisite-based topic structures is given in Figure 4. It shows that independent topic structures lead to flat BN construction with localized inference and stable recommendations, whereas prerequisite-based topic structures require dependency-aware networks and multi-path inference, resulting in variable recommendation behavior.

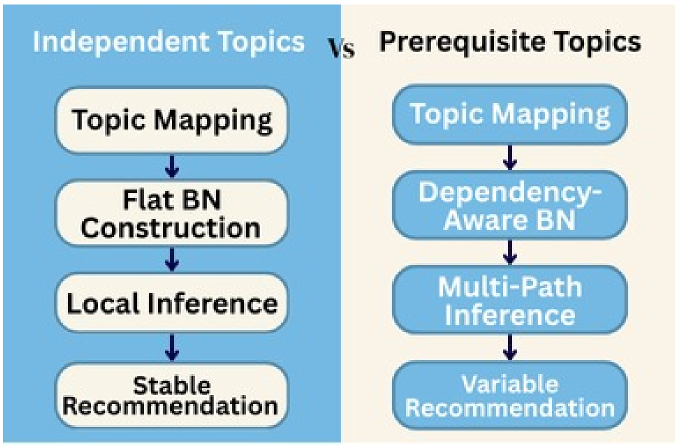

Figure 4. A Conceptual comparison of BN inference behavior under independent and prerequisite-based topic structures (Independent VS Prerequisite Topics).

4.3 Impact of prerequisite dependencies on inference complexity

A markedly different behavior was observed when prerequisite relationships were introduced between topics. As topic dependencies increased, the BN structure became denser, leading to a rapid growth in conditional relationships that needed to be evaluated during inference.

This increase in structural complexity resulted in higher computational overhead. Inference time increased noticeably as prerequisite evidence from multiple topics and cognitive levels had to be aggregated. In some cases, recommendation generation was no longer suitable for near real-time instructional use.

These observations indicate that dependency-aware

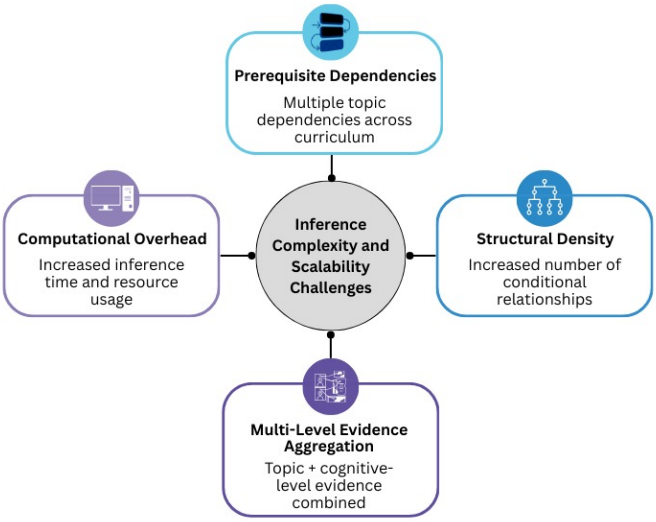

Figure 5. Radial illustration showing prerequisite dependencies, increased BN structural density, multi-level evidence aggregation, and computational overhead collectively contributing to inference complexity and scalability challenges during deployment.

topic modeling introduces a scalability challenge, particularly when multiple prerequisites and cognitive dimensions are involved. Figure 5 illustrates that prerequisite dependencies introduce multiple topic relationships and multi-level evidence aggregation, which increase structural density and computational overhead, ultimately leading to inference complexity and scalability challenges during deployment.

4.4 Structural complexity and computational constraints

Two BN configurations were constructed and evaluated: one modeling independent topics and other incorporating prerequisite-based dependencies. The dependent-topic network required additional nodes and conditional probability relationships to represent prerequisite ability propagation.

Association tests used for structure verification revealed denser dependency patterns in prerequisite-based networks compared to independent-topic networks. This increase in structural density resulted in a more complex Directed Acyclic Graph (DAG), increasing both parameter learning effort and inference cost. Figure 6 demonstrates that independent topic modeling results in sparse directed acyclic graph structures with lower parameter learning and inference costs, whereas prerequisite-based topic modeling produces denser graphs that substantially increase parameter learning complexity and inference overhead.

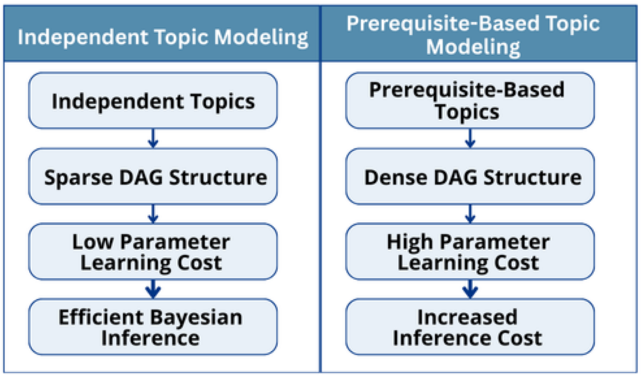

Figure 6. Illustration of prerequisite dependencies increasing BN structural density, parameter learning cost, and inference overhead compared to independent topic modeling in classroom deployment settings.

4.5 Computational constraints and memory consumption

A major empirical finding concerns the computational feasibility of discretization and inference, summarized in Table 2. The discretization failure occurred under clearly defined computational environments. On a cloud-based execution using Google Colab, the discretization process exhausted the available 12 GB of RAM and terminated. Subsequent execution on a local system with 32 GB of physical RAM and paging support resulted in memory consumption exceeding 200 GB, leading to execution failure after approximately one hour. These failures were observed when applying principled Bayesian discretization directly to prerequisite ability variables. These observations indicate that memory limitations were the primary factor constraining deployment feasibility. To complete the discretization process, a manual scaling and integer-conversion workaround was introduced. With this adjustment, the process completed successfully but still required approximately 40 GB of RAM, indicating substantial computational overhead. Under the execution environments described in Section 3.3, significant computational and memory constraints emerged during discretization and inference as shown in Table 3.

These results demonstrate that topic dependency, rather than dataset size alone, is the dominant factor driving computational limitations in real-world deployment.

Table 3. Computational and structural constraints observed during deployment.

Aspect	Independent Topics	Dependent Topics
BN structural density	Low	High
Prerequisite modeling	Not required	Required
Discretization complexity	Quadratic (manageable)	Exceeded practical limits
Peak RAM usage	Not problematic	12 GB (cloud crash); 200 GB (local crash); 40 GB (with workaround)
Additional preprocessing	No	Yes (manual scaling + discretization)
Inference feasibility	Stable	Constrained

4.6 Inference stability, evidence sparsity, and recommendation variability

The system was sensitive to the uneven distribution of assessment evidence across topics and cognitive levels. Certain Bloom's Taxonomy levels were underrepresented in the quiz data, which reflect natural instructional emphasis rather than balanced experimental design. As a consequence, learner state estimation for higher-order cognitive levels often relied on limited evidence. In this study, uncertainty and confidence were assessed using observable quantitative indicators derived from inference behavior. Increased uncertainty was reflected by higher variability in probabilistic outputs and elevated log loss values under incremental evidence updates. Reduced confidence was indicated by frequent changes in recommended instructional actions in response to small additions of assessment evidence, particularly in dependency rich topic structures.

Quantitatively, this behavior was supported by higher log-loss values and greater variability observed for prerequisite-based configurations, as reported in the comparative log-loss analysis. While the probabilistic framework could accommodate missing data, the stability of outputs declined as evidence became increasingly sparse.

This finding highlights a practical limitation: probabilistic learning models remain constrained by assessment design and data availability, regardless of their theoretical robustness. Another observed challenge was variability in recommendation outputs when new evidence was introduced incrementally. For some learners, small changes in quiz performance resulted in noticeable shifts in recommended actions, particularly in dependency-rich topic structures.

Such fluctuations were less pronounced for independent topics but became more frequent as prerequisite dependencies and multiple cognitive levels were incorporated. From a deployment perspective, this instability may reduce interpretability and trust, especially when recommendations change without substantial new learning evidence. This behavior suggests a trade-off between model responsiveness and recommendation stability, which must be carefully managed in practical systems.

Inference for independent-topic networks remained stable across stratified group k-fold cross-validation. Recommendation generation was repeatable and computationally tractable under sparse but independent evidence. In contrast, dependent-topic inference required additional preprocessing, structural simplification, and discretization adjustments to remain feasible. Log-loss comparisons reported in this study indicate greater instability and variance for dependent-topic configurations, reinforcing the observation that prerequisite modeling introduces operational challenges beyond those observed in independent-topic settings. Inference stability was assessed using quantitative indicators derived from system behavior during incremental inference. Stability was reflected by the variance of probabilistic outputs across successive inference steps and by changes in log loss values under stratified group cross validation. Additional evidence of instability was observed through the frequency of recommendation changes following small updates in assessment evidence, particularly in dependency rich topic configurations.

Assessment coverage across Bloom's Taxonomy levels was uneven, with lower-order cognitive levels dominating quiz design. Higher-order levels (Analyze, Evaluate, Create) frequently lacked sufficient direct evidence for stable ability estimation. As a result, learner state inference for higher-order cognition relied heavily on probabilistic propagation from lower levels. This dependency increased uncertainty and reduced recommendation stability, particularly in prerequisite-based topic structures.

4.7 Deployment trade-offs and observed limitations

The system was designed to generate actionable instructional recommendations rather than abstract learner scores. However, producing these recommendations required repeated Bayesian inference, particularly when dependencies and cognitive hierarchies were present. In practice, a clear trade-off emerged. More detailed modeling of learner knowledge increased computational cost and reduced responsiveness. Simpler topic structures enabled timely recommendations but at the expense of finer-grained diagnostic insight. These results underscore an important deployment consideration: increasing model expressiveness may compromise real-time usability in classroom settings. From a software engineering standpoint, these observations expose a clear trade-off between model expressiveness and system responsiveness. Architectures that encode richer curriculum dependencies increase inference cost and memory usage, reducing scalability and real-time suitability.

Although the analysis is primarily deployment-oriented, the results are supported by quantitative evidence reflecting system scale, computational behavior, and probabilistic output variability. Quantitative indicators include dataset size, BN structural density, memory consumption, execution failures under defined hardware limits, and variability in log-loss values across independent and prerequisite-based configurations. Formal statistical hypothesis testing was not applied because the primary outcomes of interest; computational feasibility, inference failures, and recommendation instability; manifest as system-level behaviors rather than stochastic performance distributions. In this context, descriptive quantitative evidence provides a more appropriate basis for evaluating deployment constraints than traditional performance metrics.

Collectively, the results indicate that while BN–based LPR is viable under constrained conditions, several limitations emerge in real-world deployment. The system performs reliably for independent topic structures but faces scalability and stability challenges when prerequisite dependencies and sparse assessment evidence are introduced.

5 Discussion: Software engineering implications and design lessons

The deployment results expose a clear architectural trade off. Introducing prerequisite relationships increases representational fidelity but also raises structural coupling, which expands inference cost and memory usage as evidence propagates across topics. This coupling reduces modularity, since local updates depend on global network state, and system responsiveness is consequently constrained under classroom scale conditions. Independent topic configurations preserve architectural separation and predictable resource consumption, yet they limit cross topic reasoning and reduce the scope of inferred learner dependencies. From a design perspective, these findings indicate that deployable learning path recommendation systems must balance dependency expressiveness against computational feasibility, with selective encoding of curriculum structure acting as a key mechanism for managing scalability and runtime stability.

This study set out to examine the practical implications of deploying BN–based LPR systems in real classroom environments, rather than to demonstrate predictive superiority. The results provide several important lessons for researchers and practitioners seeking to apply probabilistic learner models beyond controlled or simulated settings.

5.1 Curriculum structure as a primary driver of feasibility

As illustrated by the independent versus prerequisite inference behavior and the dependency-driven escalation in structural density and computational overhead, independent-topic configurations yield sparse BNs that support efficient and stable inference. In contrast, even modest prerequisite modeling introduces a noticeable increase in structural density. This observation helps explain the computational behavior reported in Table 2. The dependent-topic configuration required substantially more memory and preprocessing effort, ultimately leading to system failures under standard computational resources. These findings suggest that BN-based recommendation systems are best suited to curricula with limited dependency depth, unless significant simplification strategies are applied [32, 33]. These findings highlight an architectural consideration for software engineers: dependency-rich learner models require careful decomposition and simplification to remain deployable under classroom-scale constraints.

5.2 Structural density and inference stability

The side-by-side comparison of BN structures (Figure 2 vs. Figure 3) provides a visual explanation for the observed instability and computational cost. Additional dependency paths increase the number of conditional relationships that must be evaluated during inference, thereby amplifying both memory usage and sensitivity to sparse evidence. This structural effect is reflected in the observed recommendation variability, where dependency-rich configurations showed more frequent shifts in recommended actions under incremental evidence updates. Rather, it indicates that small changes in evidence can propagate more aggressively through dense dependency graphs, leading to less stable recommendations. Figure 7 synthesizes the key factors influencing the practical deployability of BN–based LPR systems.

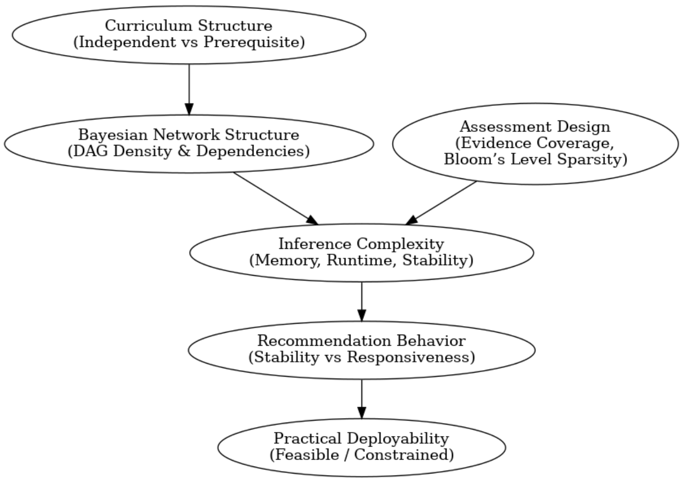

Figure 7. A synthesis of factors influencing the practical deployability of BN–based LPR systems.

5.3 The role of assessment design and evidence sparsity

Beyond structural effects, assessment design further amplifies the stability challenges. Another critical insight concerns the interaction between assessment design and probabilistic modeling. The uneven distribution of assessment items across Bloom's Taxonomy levels, summarized in Table 1, is pedagogically realistic but statistically consequential. Higher-order cognitive levels often lacked sufficient direct evidence, forcing the model to rely on probabilistic propagation from lower levels.

While Bayesian inference can accommodate missing data in principle, the results show that sustained sparsity reduces recommendation robustness, particularly in dependency-based topic structures. This finding underscores a practical constraint: sophisticated learner models remain bounded by the quality and coverage of formative assessment data.

5.4 Actionability versus computational responsiveness

A central design tension observed in this study lies between model expressiveness and real-time usability. Richer representations of learner knowledge; incorporating prerequisites and cognitive hierarchies produce theoretically appealing models [34]. However, as shown by the memory consumption and inference delays reported in Table 2, such expressiveness may come at the cost of responsiveness. From a software engineering perspective, delayed or unstable recommendations indicate scalability limits and poor runtime performance, which can undermine system reliability in production educational environments [35].

The results therefore suggest that simpler, more constrained models may offer greater practical value, particularly in classroom settings where timely feedback is essential.

5.5 Implications for adaptive learning system design

The findings point to several design implications:

- BN–based recommendation is most viable for independent or weakly coupled topics.
- Prerequisite modeling should be applied selectively and conservatively.
- Assessment coverage across cognitive levels must be considered a first-class design concern.
- Computational constraints should be evaluated early, rather than treated as an implementation detail.

These insights highlight the need for deployment-aware evaluation of adaptive learning systems, complementing performance-centric studies.

5.6 Limitations and scope of interpretation

The findings of this study should be interpreted within its defined scope. The analysis is limited to MCQ-based formative assessment and single-level prerequisite structures. No claims are made regarding learning outcomes, pedagogical effectiveness, or scalability to large curricula. Nevertheless, by documenting concrete computational failures, stability issues, and structural trade-offs, this work contributes empirical evidence that is often absent from adaptive learning literature.

This study demonstrates that theoretical feasibility does not guarantee practical deployability. BNs remain a powerful and interpretable tool for learner modeling, but their real-world application requires careful alignment with curricular structure, assessment design, and computational constraints. By foregrounding these limitations, this work aims to support more responsible and realistic adoption of probabilistic LPR systems in educational practice.

The findings reported in this study open several avenues for future work focused on improving the practical deployability of probabilistic LPR systems, rather than enhancing predictive performance. First, future studies should investigate structural simplification strategies for BNs in dependency-rich curricula. Techniques such as pruning weak prerequisite links, grouping closely related topics, or limiting dependency depth may reduce structural density and inference cost while preserving pedagogical meaning. Systematic evaluation of these trade-offs would help bridge the gap between theoretical modeling and operational feasibility.

Second, resource-aware inference mechanisms warrant further exploration. The memory and computation constraints observed in this deployment suggest the need for approximate or incremental inference approaches that are explicitly designed for classroom-scale systems. Evaluating such methods under real instructional timelines would provide valuable guidance for practitioners. Third, future work should examine assessment-aware model design. Since evidence sparsity across Bloom's Taxonomy levels significantly affected stability, adaptive assessment strategies that ensure minimal coverage of higher-order cognitive levels could be studied as a complementary solution. This direction emphasizes co-design of assessment and learner modeling, rather than treating data availability as a fixed constraint. Future work should also examine longitudinal deployments to assess system stability, computational cost, and recommendation consistency evolving over extended instructional periods.

6 Conclusion

This work shows that the practical success of BN–based LPR systems depends primarily on deployment conditions rather than theoretical modeling strength. Curriculum structure, assessment design, and computational constraints were found to directly shape system feasibility, scalability, and recommendation stability. From a software engineering perspective, the findings highlight that deployment constraints must be treated as first-class design considerations. Rich dependency modeling increases architectural complexity and memory demands. In classroom settings, this added complexity can undermine responsiveness and real-time usability. These trade-offs must therefore be managed deliberately during system design. More broadly, the results suggest that adaptive learning systems should prioritize scalable and deployment-aware architectures. Selective dependency encoding and simplified structural representations may be necessary to balance interpretability with operational feasibility. By shifting attention away from performance-centric evaluation and toward real-world behavior, this study supports more responsible and sustainable integration of probabilistic AI systems into authentic educational environments.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Li, J., Chang, Y., Liu, S., Cai, C., Zhou, Q., Cai, X., ... & Liu, Y. (2024). Higher Education in China during the Pandemic: Analyzing Online Self-Learning Motivation Using Bayesian Networks. *Sustainability, 16*(17), 7330. [CrossRef]

[2] Talha. (2025, April 19). AI-Ready Large-Scale Student Performance Dataset: Quiz Assessments, Bloom's

Taxonomy & ML Applications [Data set]. *Zenodo*. [CrossRef]

[3] Akhtar, A., & Akhtar, S. (2025). Requirements Elicitation in Transition: A Review of Conventional and Contemporary Approaches. *ICCK Journal of Software Engineering*, *1*(1), 32-45. [CrossRef]

[4] Conati, C., Barral, O., Putnam, V., & Rieger, L. (2021). Toward personalized XAI: A case study in intelligent tutoring systems. *Artificial intelligence, 298*, 103503. [CrossRef]

[5] Hooshyar, D., & Druzdzel, M. J. (2024). Memory-Based Dynamic Bayesian Networks for Learner Modeling: Towards Early Prediction of Learners' Performance in Computational Thinking. *Education Sciences*, *14*(8), 917. [CrossRef]

[6] Paleyes, A., Urma, R. G., & Lawrence, N. D. (2022). Challenges in deploying machine learning: a survey of case studies. *ACM computing surveys, 55*(6), 1-29. [CrossRef]

[7] Holstein, K., & Aleven, V. (2022). Designing for human–AI complementarity in K-12 education. *Ai Magazine, 43*(2), 239-248. [CrossRef]

[8] Luo, G., Gu, H., Dong, X., & Zhou, D. (2025). HA-LPR: A highly adaptive learning path recommendation. *Education and Information Technologies, 30*(10), 14597-14627. [CrossRef]

[9] Mavrikis, M., & Margeti, M. (2025). Review of mathematics education in the age of artificial intelligence. [CrossRef]

[10] Aftab, S. (2025). Software Engineering in the Era of Intelligence, Security, and Automation. *ICCK Journal of Software Engineering*, *1*(1), 1. [CrossRef]

[11] Akhtar, S., & Aftab, S. (2025). Towards AI-Augmented Software Engineering: A Theoretical Framework. *ICCK Journal of Software Engineering*, *1*(2), 124. [CrossRef]

[12] Zheng, Y., Wang, D., Zhang, J., Li, Y., Xu, Y., Zhao, Y., & Zheng, Y. (2025). A unified framework for personalized learning pathway recommendation in e-learning contexts. *Education and Information Technologies*, *30*(6), 7911-7948. [CrossRef]

[13] Bin, Q., Zuhairi, M. F., & Morcos, J. (2024). A comprehensive study on personalized learning recommendation in e-learning system. *IEEE Access, 12*, 100446-100482. [CrossRef]

[14] Pelánek, R. (2017). Bayesian knowledge tracing, logistic models, and beyond: an overview of learner modeling techniques. *User modeling and user-adapted interaction*, *27*(3), 313-350. [CrossRef]

[15] Martínez-Martínez, A., Gómez-Cambronero, Á., Montoliu, R., & Remolar, I. (2025). Towards the Adoption of Recommender Systems in Online Education: A Framework and Implementation. *Big Data and Cognitive Computing*, *9*(10), 259. [CrossRef]

[16] Mohammadi, M., Tajik, E., Martinez-Maldonado, R., Sadiq, S., Tomaszewski, W., & Khosravi, H. (2025). Artificial intelligence in multimodal learning analytics: A systematic literature review. *Computers and Education: Artificial Intelligence, 8*, 100426. [CrossRef]

[17] Wiley, K., Dimitriadis, Y., & Linn, M. (2024). A human-centred learning analytics approach for developing contextually scalable K-12 teacher dashboards. *British Journal of Educational Technology, 55*(3), 845-885. [CrossRef]

[18] Haldar, S., Sengupta, S., & Das, A. K. (2025). Personalized Learning Path Recommendation using Graph Reinforcement Learning. *Procedia Computer Science*, *258*, 3480–3489. [CrossRef]

[19] Yang, Q., & Liang, C. (2025). A Second-Classroom Personalized Learning Path Recommendation System Based on Large Language Model Technology. *Applied Sciences*, *15*(14), 7655. [CrossRef]

[20] Ma, Y., Wang, L., Zhang, J., Liu, F., & Jiang, Q. (2023). A Personalized Learning Path Recommendation Method Incorporating Multi-Algorithm. *Applied Sciences*, *13*(10), 5946. [CrossRef]

[21] Zhang, F., Feng, X., & Wang, Y. (2024). Personalized process–type learning path recommendation based on process mining and deep knowledge tracing. *Knowledge-Based Systems*, *303*, 112431. [CrossRef]

[22] Zhou, L. Y., & Wang, Y. Y. (2025). Simulation of personalized english learning path recommendation system based on knowledge graph and deep reinforcement learning. *Scientific Reports*, *15*(1), 34554. [CrossRef]

[23] Nguyen, D. T., & Nguyen, T. M. T. (2024, November). A Knowledge Map Mining-Based Personalized Learning Path Recommendation Solution for English Learning. In *KDIR* (pp. 48-59). [CrossRef]

[24] Sun, C., Huang, S., Sun, B., & Chu, S. (2025). Personalized learning path planning for higher education based on deep generative models and quantum machine learning: a multimodal learning analysis method integrating transformer, adversarial training and quantum state classification. *Discover Artificial Intelligence*, *5*(1), 29. [CrossRef]

[25] Li, X., & Zhang, B. (2024). Personalized Learning Path Recommendation Algorithm for English Listening Learning. *Journal of Electrical Systems*, *20*(6s), 2188–2199. [CrossRef]

[26] Cheng, X., Zhang, Z., Wang, J., Fang, L., He, C., Guan, Q., ... & Luo, W. (2025). GraphRAG-Induced Dual Knowledge Structure Graphs for Personalized Learning Path Recommendation. *arXiv preprint arXiv:2506.22303*.

[27] Abu-Rasheed, H., Jumbo, C., Al Amin, R., Weber, C., Wiese, V., Obermaisser, R., & Fathi, M. (2025, April). LLM-assisted knowledge graph completion for curriculum and domain modelling in personalized higher education recommendations. In *2025 ieee global*

engineering education conference (educon) (pp. 1-5). IEEE. [CrossRef]

[28] Gul, R., Kanwal, S., & Khan, S. S. (2020). Preferences of the teachers in employing revised blooms taxonomy in their instructions. *sjesr, 3*(2), 258-266. [CrossRef]

[29] Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, *41*(4), 212–218. [CrossRef]

[30] bnlearn documentation. (n.d.). BNLearn's Documentation. Retrieved January 14, 2026, from https://erdogant.github.io/bnlearn/pages/html/index.html

[31] Kuhlmane, S., Grasevic, D., Bakery, R. S., Scotte, G., & Siemens, G. (2025). Learning Analytics and Impact on Personalized Student Learning. *Journal of Teaching and Learning, 2*(1), 27-50. [CrossRef]

[32] Baker, R. S. (2019). Challenges for the future of educational data mining: The Baker learning analytics prizes. *Journal of educational data mining, 11*(1), 1-17. [CrossRef]

[33] Käser, T., & Schwartz, D. L. (2020). Modeling and analyzing inquiry strategies in open-ended learning environments. *International Journal of Artificial Intelligence in Education, 30*(3), 504-535. [CrossRef]

[34] Zhang, N., Biswas, G., & Hutchins, N. (2022). Measuring and analyzing students' strategic learning behaviors in open-ended learning environments. *International Journal of Artificial Intelligence in Education, 32*(4), 931-970. [CrossRef]

[35] Cukurova, M. (2025). The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *British Journal of Educational Technology, 56*(2), 469-488. [CrossRef]

Muhammad Talha Mansoor is a research scholar who has completed MS in Computer Science from the Virtual University of Pakistan, Pakistan. He is currently working as a Lecturer at the Government College of Education, Karachi, Pakistan. His research interests center on the application of Artificial Intelligence in education, with a focus on enhancing teaching, learning, and assessment processes. He is particularly interested in intelligent tutoring systems, learning analytics, and AI-driven personalized learning to improve student engagement and educational outcomes. (Email: talha131@gmail.com)