Journal of Inclusive Methodology and Technology in Learning and Teaching
ISSN 2785-5104
Anno5 n.3 (2025)

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

Reimagining Personalization through Explainable Decision Support for Learners

Neda Afreen [1], Ludovico Boratto [1], Gianni Fenu[1], Mirko Marras [1,*]

[1] Department of Mathematics and Computer Science, University of Cagliari
{neda.afreen, ludovico.boratto, fenu, mirko.marras}@unica.it,
* Correspondence: mirko.marras@unica.it;

Abstract: The expansion of digital education has enabled flexibility but also introduced cognitive overload as learners navigate vast ecosystems of resources. Intelligent decision support systems have emerged to mediate this complexity, yet most rely on behavioral correlations rather than pedagogical reasoning, lacking transparency and fostering dependence on opaque models. This article reports findings from a large-scale audit of multiple recommendation approaches applied to diverse educational datasets, analyzing their capacity to guide learners effectively and transparently. Results show that collaborative and autoencoder-based methods generalize well, while knowledge-aware models outperform when curricular structures are explicit. Path-reasoning methods, though not top performers in ranking accuracy, deliver valuable pedagogical explanations that promote reflection, positioning intelligent systems as transparent and pedagogically grounded mediators.

Keywords: educational decision support; explainable artificial intelligence; personalized learning; recommender systems; recommendation methods; course recommendation.

1. Introduction

The exponential growth of online education has created an abundance of opportunities for learners to access content flexibly, asynchronously, and across institutional boundaries. From massive open online courses to micro-credentials and self-paced tutorials, learners can now construct unique trajectories through diverse resources. This diversification reflects a broader transformation of the educational landscape, one in which learning increasingly occurs within hybrid ecosystems blending formal and informal contexts. Yet, while the democratization of access represents a major step forward, it also presents new challenges. Faced with extensive catalogs and multiple pathways, learners must continuously decide what to study next, which prerequisites to complete, and how to align learning goals with evolving skills and career aspirations. This decision-making process, once guided by structured curricula and instructors, is now largely delegated to individual learners, creating the risk of cognitive overload, inefficient sequencing, and disengagement (Guo et al., 2022).

To address this complexity, intelligent decision support systems have emerged as mediators of the modern learning experience (Afreen et al., 2025; Urdaneta-Ponte et al., 2021). These systems promise to recommend relevant content, helping learners navigate large repositories more effectively. However, their current implementations often inherit assumptions from e-commerce or entertainment recommender systems, where behavioral similarity is used as a proxy for relevance (Zhang & Chen, 2020). In education, such logic is limited. Learning is not only about preference but also about progression, dependency, and conceptual readiness (Khosravi et al., 2022). The assumption that "learners like you also took this course" fails to capture the pedagogical logic underlying curricular design. Moreover, when recommendations are delivered with no explanatory rationale, they risk transforming personalization into dependence, as learners may follow algorithmic guidance blindly (Conati et al., 2018).

In this paper, we argue for a paradigm shift from behavioral to causal and explainable recommendation in education. We explore how knowledge graph-based systems, enhanced by causal reasoning and natural language explanations, can serve as pedagogically coherent companions that guide, explain, and empower learners rather than merely predict their next course. Through an extensive comparative study of recommendation models across, we seek to identify not only which models perform best but also which designs best align with beyond-utility and explainability objectives.

2. Context

Contemporary education increasingly unfolds within complex socio-technical ecosystems in which learners, educators, and algorithms co-shape learning trajectories. Within these ecosystems, decision support systems act as intermediaries, by structuring access to knowledge, sequencing content, and mediating between personal interests and pedagogical coherence. Their role, therefore, extends beyond technical facilitation to epistemological shaping: they help define how knowledge is encountered and understood. Yet the majority of existing systems operate on correlation-based logics. They infer that if many learners transition from Course A to Course B, such a sequence must be appropriate for others. While such correlations capture patterns of behavior, they ignore the why behind those transitions , i.e., the causal or pedagogical relationships that justify moving from one concept to another (Cavenaghi et al., 2024).

This limitation has several consequences. First, it risks recommending content that learners are not yet ready for, leading to frustration or superficial engagement. Second, it may reinforce popularity bias, favoring frequently taken courses regardless of their pedagogical appropriateness. Third, it deprives learners of transparency. When algorithms suggest a course without clear justification, learners cannot develop metacognitive awareness of their progress or understand the logic guiding their learning path. Such opacity undermines trust and limits the learner's epistemic agency, i.e., the ability to reason critically about knowledge and make informed educational choices.

From a pedagogical perspective, these shortcomings are nontrivial. Learning is a cumulative process characterized by dependencies between concepts. Effective sequencing must consider what has been mastered and what foundational knowledge remains missing. Without such causal modeling, recommender systems risk mimicking surface-level engagement rather than supporting meaningful learning progressions (Afreen et al., 2025). Furthermore, lack of explainability disconnects learners from the rationale underlying their educational path, weakening self-regulation and reflection. For inclusive education, where diverse learners require equitable and transparent guidance, these limitations can amplify existing disparities. Systems that are opaque or biased tend to benefit already advantaged learners who can interpret or compensate for the missing logic, leaving others at risk of confusion or disengagement.

3. Methodology

To examine how different intelligent systems can best support learners in choosing their next educational steps, we designed a comprehensive evaluation framework that allows us to compare several types of recommendation methods within a common structure. The aim was to understand not only which models perform well, but also how their behavior aligns with real learning processes.

Our study drew on four publicly available educational datasets, each representing a distinctive kind of learning environment. One came from a marketplace offering short, stand-alone courses on a wide variety of topics (Dessì, Fenu, Marras, & Reforgiato Recupero, 2018). Another described a massive open online course platform, where courses are organized around clearly defined prerequisites and conceptual progressions (Yu et al., 2020). A third dataset represented a repository of video lectures, typically used for informal learning (Bulathwela, Pérez-Ortiz, Novak, Yilmaz, & Shawe-Taylor, 2021). The fourth came from a practice-based platform in which learners complete short exercises and receive feedback (Liu et al., 2021).

Because these datasets differed in structure and purpose, we first translated them into a common format that would allow meaningful comparison. We represented each as a knowledge graph, a kind of map that links all the main elements of a learning ecosystem. In this map, learners are connected to the courses or resources they have used, courses are linked to the concepts they teach, instructors are associated with the courses they deliver, and institutions are tied to the courses they offer. Relationships such as "covers," "is prerequisite of," or "taught by" make it possible for algorithms to understand how different pieces of knowledge relate to one another.

Within this framework, we compared three broad approaches to recommendation. The first category includes methods that learn directly from patterns of learner activity - for example, identifying that people who took one course often went on to take another. These systems are powerful because they can discover useful regularities even when little descriptive information about the courses is available. However, they focus only on what people did, not on why courses belong together pedagogically. The second category builds on the first by also considering the structure of the knowledge graph. These models can take into account, for example, that two courses share similar topics, or that one is the natural continuation of another. By combining behavioral data with semantic and relational information, they can offer suggestions that are not only popular but also educationally coherent. The third family of approaches goes a step further by explicitly reasoning through the connections in the graph. Instead of relying on general similarity, these systems can trace a path that explains how a new course is connected to a learner's previous experiences. For instance, they might identify that a learner completed a course on data analysis, which covered statistics, and that another course on machine learning builds directly on that foundation. In this way, the system can provide a recommendation that is not just relevant but also accompanied by a clear, human-understandable rationale.

To make sure our comparisons reflected realistic learning progressions, we evaluated all models using time-based splits of the data. This means that we trained the systems on earlier learner activities and tested them on later ones, simulating how recommendations would work in a real setting where future choices are unknown.

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

We evaluated the performance of all models using a comprehensive set of metrics covering three complementary dimensions: utility, beyond-utility, and explainability. For utility, we adopted three ranking-oriented indicators commonly used in recommender system research. Precision measures the proportion of recommended courses that are actually relevant to the learner, while Recall quantifies the fraction of all relevant courses successfully retrieved by the system. Normalized Discounted Cumulative Gain (NDCG) extends these by accounting for the position of relevant courses in the ranked list, giving higher weight to those appearing near the top, thus reflecting not only what is recommended but also how effectively it is prioritized.

The beyond-utility dimension evaluates the ability of a system to promote exploration, equity, and exposure to diverse learning opportunities. Coverage measures the proportion of the entire course catalog that appears in recommendations, indicating how broadly the system draws from available resources. Diversity assesses the variety among the recommended items, ensuring that learners are exposed to multiple perspectives rather than repetitive content. Novelty captures how often a learner encounters previously unseen or less familiar materials, while Serendipity reflects the ability to surface unexpected yet relevant learning options that encourage curiosity.

Finally, the explainability dimension examines how clearly and pedagogically meaningfully the system communicates the reasoning behind its recommendations. Three knowledge-graph–based indicators were used. Linking Interaction Diversity (LID) quantifies the variety of learner interactions that contribute to forming an explanation, capturing how multifaceted the reasoning is. Shared Entity Diversity (SED) measures the range of conceptual entities or topics appearing in explanatory paths, indicating the semantic richness of the explanation. Path Type Diversity (PTD) evaluates the structural variety of reasoning paths within the knowledge graph, showing how many distinct relational patterns the system can leverage to justify a recommendation.

This methodology, by combining diverse datasets, a unified representation of educational relationships, and consistent evaluation procedures, allowed us to compare recommendation methods on meaningful grounds. More importantly, it created the conditions to study not only how well these systems predict what learners might do next, but how coherently they can support actual learning journeys.

4. Findings and Interpretation

4.1. RQ1 Analysis of utility objectives

In the first research question, we examined how different classes of recommendation models perform in terms of ranking utility, i.e., their ability to prioritize courses most relevant to a learner's current progress and interests. Results in Table 1 show that model effectiveness depends strongly on data structure and domain organization. In the sparse, heterogeneous environment of COCO, the autoencoder-based collaborative model (MacridVAE) achieved the best ranking accuracy (NDCG 0.15; MRR 0.13; Recall 0.24), demonstrating the strength of latent representations in capturing complex learner-item patterns. In the structured MOOCube dataset, where explicit relations among topics, instructors, and prerequisites exist, the knowledge-aware model (CFKG) performed best (NDCG 0.25; Recall 0.40), suggesting that incorporating curricular semantics enhances recommendation precision. In the denser datasets (PEEK and MOOPer) traditional matrix factorization (ENMF) remained highly competitive, with the highest overall accuracy (NDCG 0.87; MRR 0.84; Recall 0.96) in practice-based environments. While collaborative and knowledge-aware approaches dominated in terms of utility, path-reasoning models contributed unique advantages. Although their ranking metrics were generally lower, their ability to provide transparent, structured explanations enhanced interpretability and trust.

Table 1. Utility-based performance for the best-performing models across four educational contexts. Metrics show average values under consistent splits across runs.

Educational Context	Best Performer	Precision	Recall	NDCG
Course Marketplace (COCO)	MacridVAE (Autoencoder)	0.15	0.13	0.24
MOOC Platform (MOOCube)	CFKG (Knowledge-aware)	0.25	0.21	0.40
Video Lecture Repository (PEEK)	MacridVAE (Autoencoder)	0.36	0.29	0.60
Practice-oriented Platform (MOOPer)	ENMF (Matrix Factorization)	0.87	0.84	0.96

4.2. RQ2 Analysis of beyond-utility objectives

The second research question explored how recommendation systems perform in terms of beyond-utility objectives, which measure not only accuracy but also how well a system supports exploration, equity, and exposure to diverse learning opportunities. Metrics such as coverage, novelty, and serendipity were analyzed across the same four datasets to assess whether the models encourage discovery or inadvertently reinforce popularity bias. As shown in Table 2, the reinforcement learning–based PGPR exhibited the highest serendipity and novelty in open and less structured contexts like COCO (serendipity 0.99; novelty 0.71) and MOOPer (serendipity 0.95; novelty 0.73), effectively surfacing relevant but unexpected content. In contrast, the generative model PEARLM achieved the widest coverage in more structured contexts, such as MOOCube (0.99), demonstrating an ability to connect learners to underexplored areas of the catalog. Meanwhile, matrix factorization (ENMF) maintained balanced novelty and recall in denser environments like PEEK, reflecting its strength in stable, behavior-rich datasets. These results emphasize that educational value extends beyond ranking correctness. Systems optimizing only accuracy tend to converge on popular resources, potentially narrowing learners' horizons. Conversely, models with high serendipity and coverage promote engagement and curiosity, offering learners opportunities to explore material just beyond their comfort zone. Beyond-utility optimization therefore represents a step toward more inclusive, reflective learning ecosystems, where personalization fosters discovery rather than repetition.

Table 2. Comparative analysis of beyond-utility objectives across educational contexts. Metrics show average values under consistent splits across runs.

Educational Context	Best Performer	Coverage	Diversity	Novelty	Serendipity
Course Market-place (COCO)	PGPR (RL-based)	0.75	0.68	0.71	0.99
MOOC Platform (MOOCube)	PEARLM (Genera-tive)	0.99	0.73	0.69	0.83
Video Lecture Repository (PEEK)	ENMF (Matrix Factorization)	0.65	0.55	0.52	0.47
Practice-oriented Platform (MOOPer)	PGPR (RL-based)	0.88	0.79	0.73	0.95

4.2. RQ3 Analysis of explainability objectives

The explainability analysis summarized in Table 3 highlights distinct strengths among reasoning paradigms. Reinforcement learning–based models such as PGPR achieved the best results in open and unstructured environments (COCO and PEEK), combining high Linking Interaction Diversity (LID) with wide Shared Entity Diversity (SED). This means that PGPR explanations draw on multiple learner behaviors and connect them to a broad range of concepts and entities in the knowledge graph, making recommendations easier to contextualize. In contrast, generative models such as PEARLM performed best in structured environments (MOOCube and MOOPer), offering high Path Type Diversity (PTD) that captures multi-hop relationships among courses, prerequisites, and topics. Their explanations often reflect the logical flow of learning progressions rather than isolated correlations. Taken together, these results suggest that explainability is domain-dependent: in loosely organized learning ecosystems, diversity of links and entities fosters comprehension and trust, whereas in structured curricula, path diversity enhances pedagogical coherence. Both forms of reasoning contribute to interpretability, transparency, and ultimately to the learner's sense of guidance within complex digital environments.

Table 3. Explainability performance for the best-performing models across four educational contexts. Metrics show average values under consistent splits across runs.

Educational Context	Best Performer	LID	SED	PTD
Course Marketplace (COCO)	PGPR (RL-based)	0.68	0.81	0.60
MOOC Platform (MOOCube)	PEARLM (Generative)	0.44	0.69	0.50
Video Lecture Repository (PEEK)	PGPR (RL-based)	0.45	0.98	0.48
Practice-oriented Platform (MOOPer)	PEARLM (Generative)	0.60	0.78	0.68

5. Discussion and Implications

Our findings point toward a redefinition of personalization in education. Personalization should not be reduced to behavioral imitation; it must involve pedagogical reasoning and learner empowerment. Systems that simply mirror peer behavior risk reproducing biases and encouraging shallow engagement. By contrast, systems that incorporate causal and conceptual reasoning can adapt recommendations to the learner's cognitive state and curricular context, enhancing effectiveness.

Explainability emerges as a key ingredient for inclusive learning. When learners understand why a recommendation appears, they can engage metacognitively with their learning process, reflecting on prerequisites, alternatives, and long-term goals. Explanations transform recommendation systems into conversational partners, fostering co-construction of knowledge rather than passive consumption. For educators, this transparency provides diagnostic insight: they can inspect recommendations, identify misalignments, and intervene when learners diverge from intended pathways. In this sense, explainable decision support systems are boundary objects between pedagogy and technology, facilitating dialogue between human and algorithmic reasoning.

From a design perspective, we advocate a two-stage approach. In the first stage, a high-performing collaborative or knowledge-aware model identifies relevant candidates. In the second stage, a path-reasoning or language-model-based component generates personalized explanations grounded in the knowledge graph. This structure balances accuracy with intelligibility, ensuring that learners receive both relevant and understandable guidance. Importantly, explanations should not merely justify recommendations post hoc but function as pedagogical narratives that situate each learning object within the learner's evolving trajectory.

The practical implications extend beyond individual learners. Institutions can leverage explainable systems to enhance advising processes, support differentiated instruction, and monitor equity of access. By tracking which learners receive which recommendations and why educators can identify patterns of overrepresentation and adjust accordingly. For policy makers and learning platform designers, explainable artificial intelligence for decision-making support offers a pathway toward accountability and transparency in educational technologies.

6. Conclusions

The evolution of digital education demands recommender systems that are not only accurate but also explainable, diverse, and pedagogically grounded. This study demonstrates that while collaborative and knowledge-aware models achieve strong relevance scores, the highest educational value arises when systems also cultivate exploration and transparency. Reinforcement learning and generative reasoning models enhance interpretability by tracing causal and conceptual links between learning objects, producing explanations that mirror the cognitive logic of instruction.

Explainability metrics reveal that the richness and diversity of explanatory paths are as important as ranking precision for fostering learner engagement and trust. When learners can see why a recommendation appears - whether through recent interactions, shared conceptual entities, or diverse reasoning chains - they gain agency over their learning process and develop meta-cognitive awareness. For educators and designers, these explanations serve as analytical tools, offering visibility into how algorithms map dependencies and where interventions might be needed.

Future work should extend these insights through longitudinal studies of learner behavior, measuring how exposure to explainable and diverse recommendations influences persistence, curiosity, and self-regulation. Advancements in large language models open the door to natural-language explanations grounded in structured graphs, provided that factual alignment is maintained. Ultimately, the path forward lies in hybrid educational intelligence, i.e., systems that combine the precision of statistical models with the reasoning clarity of knowledge graphs and the narrative fluency of language models. Such systems will not merely recommend; they will teach, explain, and co-construct learning journeys, embodying a vision of artificial intelligence as an ethical, transparent, and reflective partner in education.

References

Afreen, N., Balloccu, G., Boratto, L., Fenu, G., Malloci, F. M., Marras, M., & Martis, A. G. (2025). Can path-based explainable recommendation methods based on knowledge graphs generalize for personalized education? In Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization (pp. 273–278). ACM.

Afreen, N., Boratto, L., Fenu, G., Marras, M., & Soccol, A. (2025). Effective and Transparent Course Recommendation through Causal Reasoning with Language Models. In *CEUR WORKSHOP PROCEEDINGS* (Vol. 4026, pp. 84-87). CEUR-WS.

Bulathwela, S., Pérez-Ortiz, M., Novak, E., Yilmaz, E., & Shawe-Taylor, J. (2021). PEEK: A large dataset of learner engagement with educational videos. arXiv preprint arXiv:2109.03154.

Cavenaghi, E., Zanga, A., Stella, F., & Zanker, M. (2024). Towards a causal decision-making framework for recommender systems. *ACM Transactions on Recommender Systems*, *2*(2), 1-34.

Conati, C., Porayska-Pomsta, K., & Mavrikis, M. (2018). AI in education needs interpretable machine learning: Lessons from open learner modelling. arXiv preprint arXiv:1807.00154.

Dessì, D., Fenu, G., Marras, M., & Reforgiato Recupero, D. (2018). COCO: Semantic-enriched collection of online courses at scale with experimental use cases. In Proceedings of the World Conference on Information Systems and Technologies (WorldCIST 2018) (Vol. 746, pp. 1386–1396). Springer.

Guo, G., Zhang, J., Zhu, F., Qin, S., Zhou, M., Xu, Y., & He, X. (2022). A survey on knowledge graph-based recommender systems. IEEE Transactions on Knowledge and Data Engineering, 34(8), 3549–3578.

Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. Computers and Education: Artificial Intelligence, 3, 100074.

Liu, K., Zhao, X., Tang, J., Zeng, W., Liao, J., Tian, F., Zheng, Q., Huang, J., & Dai, A. (2021). MOOPer: A large-scale dataset of practice-oriented online learning. In Proceedings of the China Conference on Knowledge Graph and Semantic Computing (CCKS 2021) (Vol. 1466, pp. 281–287). Springer.

Urdaneta-Ponte, M. C., Mendez-Zorrilla, A., & Oleagordia-Ruiz, I. (2021). Recommendation systems for education: Systematic review. Electronics, 10(14), 1611.

Yu, J., Luo, G., Xiao, T., Zhong, Q., Wang, Y., Feng, W., Luo, J., Wang, C., Hou, L., Li, J., Liu, Z., & Tang, J. (2020). MOOCube: A large-scale data repository for NLP applications in MOOCs. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2020) (pp. 3135–3142).

Zhang, Y., & Chen, X. (2020). Explainable recommendation: A survey and new perspectives. Foundations and Trends® in Information Retrieval, 14(1), 1–101..