

Recommendation Systems for Personalized Digital Learning Interfaces: Analysis of User Engagement and Learning Progression

Manomay Jain

Abstract

Recommendation systems increasingly determine what users see next in digital products, yet recommendation quality is often evaluated through prediction accuracy rather than its effect on meaningful user outcomes. This study develops a transparent simulation framework to examine how recommendation relevance, personalization strength, novelty, and interaction frequency may relate to user engagement and learning progression in personalized digital learning interfaces. A reproducible synthetic dataset of 500 users was generated to represent plausible variation in prior skill, personalization, recommendation relevance, novelty, interaction frequency, click-through rate, content completion, engagement, and learning progression. The observations are simulated and are not presented as records from real users. Descriptive statistics, Pearson correlations, quartile comparisons, and heteroskedasticity-robust multiple regression were used to test three hypotheses. In the simulated data, recommendation relevance was strongly associated with engagement ($r = .65$) and moderately associated with learning progression ($r = .53$). In multivariable analysis, recommendation relevance remained the strongest predictor of engagement ($b = .426, p < .001$), while personalization and novelty also contributed positively. Recommendation relevance also predicted learning progression ($b = .319, p < .001$), alongside prior skill. The engagement model explained 48.1% of simulated variance, while the learning-progression model explained 34.5%. These findings should be interpreted as a methodological demonstration rather than empirical population evidence. The framework illustrates how recommender-system evaluation can move beyond accuracy to examine whether personalized interface logic improves the quality of user interaction and supports more coherent learning pathways.

Keywords: recommender systems; personalization; digital interfaces; user engagement; learning pathways; learning analytics; synthetic data

I. Introduction

Digital interfaces are no longer passive containers of information. Streaming platforms, e-commerce applications, social networks, productivity tools, and learning platforms continuously rank, filter, and reorder content for individual users. Recommender systems are the computational layer behind much of this personalization. Their purpose is to reduce information overload by estimating which item, resource, action, or pathway is most useful for a particular user at a particular moment. Early collaborative-filtering work such as GroupLens demonstrated how the preferences of similar users could be used to filter large information spaces (Resnick et al., 1994). Subsequent research established collaborative, content-based, knowledge-based, and hybrid recommendation as core strategies for personalized information access (Burke, 2002; Adomavicius & Tuzhilin, 2005).

The modern problem is no longer simply whether an algorithm can predict a rating or rank a relevant item. In software products, recommendation quality is experienced through the interface. A relevant suggestion can reduce search effort, increase discovery, and help a user continue a task. An irrelevant or repetitive suggestion can produce disengagement, choice fatigue, or abandonment. This distinction is important because offline accuracy metrics do not always capture the dimensions of recommendation that users actually value. Research increasingly emphasizes novelty, diversity, serendipity, explanation, and usefulness in addition to predictive accuracy. Recent evidence from large-scale digital platforms also suggests that novelty and serendipity can affect engagement independently of conventional accuracy measures (Ping et al., 2025).

The issue becomes particularly consequential in digital learning environments. Educational platforms do not merely seek clicks; they seek to guide learners toward content sequences that are appropriate to their knowledge, goals, and pace. Learning-path recommendation therefore differs from conventional product recommendation because the immediate preference of a learner may not always align with the next pedagogically useful step. Surveys of learning-path personalization identify learner profile, knowledge state, learning objectives, available resources, and sequencing constraints as recurring inputs to recommendation logic (Nabizadeh et al., 2020). Reviews of educational recommender systems similarly emphasize that personalized learning support can

help learners navigate large collections of resources, but evaluation practices remain inconsistent and often focus more on system performance than on educational outcomes (da Silva et al., 2023; Rahayu et al., 2023).

Recommendation therefore operates at the intersection of data science, human-computer interaction, and behavioral design. A strong personalized interface must infer user preferences from explicit and implicit signals, rank suitable items, expose enough novelty to support discovery, and avoid overwhelming the user with excessive recommendation frequency. Matrix-factorization methods demonstrated the power of latent representations for modeling user-item interactions (Koren et al., 2009), while hybrid systems combine complementary signals to overcome sparsity and cold-start limitations (Burke, 2002). More recent systems incorporate contextual, semantic, deep-learning, and multimodal information. Yet regardless of algorithmic sophistication, the practical question for product teams remains simple: does the recommendation logic make the interface more useful to the user?

This paper develops a compact, testable framework around that question. It focuses on personalized digital learning interfaces because they provide two distinct outcomes: engagement with the interface and progression through learning content. The study uses a reproducible synthetic dataset to demonstrate how a real product experiment or learning-platform study could test whether recommendation relevance and personalization are associated with better outcomes. Synthetic observations are used only to validate the analytical design; they are not represented as real user behavior.

Three hypotheses guide the analysis. H1: recommendation relevance is positively associated with user engagement. H2: higher personalization is associated with stronger engagement and learning progression, after accounting for recommendation relevance and prior skill. H3: recommendation quality should be evaluated using behavioral and pathway outcomes, not click behavior alone; accordingly, relevance is expected to remain positively associated with learning progression even after controlling for interaction frequency and prior skill. The contribution is a practical empirical framework that links recommender-system quality to outcomes visible at the interface level rather than treating ranking accuracy as the sole endpoint.

II. Literature Review and Conceptual Framework

2.1 Recommendation systems and personalization

Recommender systems emerged as a response to information overload. Collaborative filtering uses patterns of behavior across users to infer preferences, whereas content-based approaches compare item attributes with a user's known interests. Knowledge-based approaches incorporate domain rules or explicit requirements, and hybrid systems combine multiple strategies. Adomavicius and Tuzhilin (2005) described these approaches as a foundation for the next generation of recommender systems and emphasized the importance of contextual and multidimensional information. Burke (2002) showed that hybrid architectures can exploit complementary strengths and reduce the weaknesses of any single recommendation strategy.

Latent-factor methods later became influential because they model users and items within lower-dimensional spaces that capture otherwise hidden preference structure. Koren et al. (2009) documented how matrix factorization became a powerful approach for large recommendation problems and could incorporate implicit feedback, temporal effects, and confidence. Contemporary systems extend these ideas using neural networks, graph models, sequence models, review text, and multimodal content. A recent ACM Computing Surveys review illustrates how review-based recommenders increasingly combine language and behavioral signals to infer fine-grained preferences (Raza et al., 2025). Despite methodological advances, the underlying product objective remains the same: rank a small set of useful options from a much larger candidate space.

2.2 Beyond accuracy: engagement, novelty, and serendipity

Traditional evaluation often uses error metrics, ranking metrics, or top-k relevance. These measures are valuable, but they do not fully capture whether recommendations improve the user experience. Recommender systems can become over-specialized by repeatedly presenting items that are very similar to previous consumption. Research on serendipity argues that useful recommendations should sometimes expose users to relevant but unexpected content, thereby supporting discovery rather than merely reproducing known preferences (Jafari & Ravanmehr, 2021).

This matters for engagement because a recommendation list can be accurate yet dull. Ping et al. (2025), using data from a music-streaming platform, found that novelty and serendipity were positively associated with user engagement, while diversity did not uniformly improve outcomes. The implication is that recommendation quality is multidimensional. Relevance is necessary, but the interface must also balance exploitation of known interests with exploration of new but potentially useful content. For digital learning products, this balance is especially important: overly familiar recommendations may fail to advance the learner, while overly novel material may be poorly matched to readiness.

2.3 Learning-path recommendation

Learning-path recommendation treats content sequence as part of the recommendation problem. A learning path is a structured sequence of activities designed to move a learner toward a particular objective. Nabizadeh et al. (2020) reviewed methods for learning-path personalization and identified user background, knowledge, goals, learning style, and resource characteristics as common personalization parameters. The review also highlighted the difficulty of evaluating path quality because a recommendation can be relevant at the item level without producing an effective sequence over time.

Educational recommender systems have grown rapidly as online platforms generate richer traces of learner behavior. A 2023 review by da Silva et al. reported continued growth in systems that recommend learning resources and activities, while also identifying limitations in evaluation methods and the need for stronger evidence about teaching and learning outcomes. Rahayu et al. (2023) similarly reviewed learning-path recommender systems and highlighted student models, learning objects, learning activities, and external context as contributing factors. A broader survey of AI-based e-learning recommenders notes that personalized content and pathways can help address differences in learner requirements and abilities (Bourkhoukhou et al., 2023).

The key distinction is therefore between engagement and progression. A system can increase clicks by recommending entertaining or easy content, but a learning interface should also move users through an appropriate sequence. This motivates the use of a learning-progression score as a separate outcome in the present framework. In a real implementation, progression could be measured using mastery estimates, completion of prerequisite sequences, assessment gains, or movement toward curriculum objectives.

2.4 Research gap

The literature provides strong technical foundations for recommendation and a growing body of work on personalized learning, yet three practical gaps remain. First, many studies evaluate algorithmic performance without directly connecting recommendation quality to interface-level engagement. Second, engagement is often measured using clicks or consumption alone, even though repeated interaction does not necessarily indicate a high-quality pathway. Third, educational recommendation research frequently evaluates a particular algorithm or domain, making it difficult for product teams to translate findings into a compact set of measurable design variables.

This study addresses those gaps by modeling recommendation relevance, personalization strength, novelty, and interaction frequency alongside two distinct outcomes: engagement and learning progression. The framework is deliberately algorithm-agnostic. It can be implemented with collaborative filtering, content-based recommendation, a hybrid model, or a modern sequence model. The empirical question is not whether one specific model is best, but whether the quality of the recommendations exposed through the interface is associated with better user outcomes.

III. Methodology

3.1 Research design

The study uses a cross-sectional synthetic-data design. A reproducible dataset of 500 simulated users was created to represent plausible behavior in a personalized digital learning interface. The simulation is explicitly methodological. No real participants were recruited, and the generated observations must not be interpreted as empirical estimates of user behavior. The purpose is to demonstrate a transparent analysis pipeline that can later be applied to platform telemetry, an A/B test, or an appropriately consented learner study.

Each simulated user was assigned values for prior skill, personalization score, recommendation relevance, interaction frequency, novelty score, user engagement, content-completion percentage, learning-progression score, and click-through rate. Scores for prior skill, personalization, recommendation relevance, novelty, engagement, and progression were bounded on a 1-10 scale. Interaction frequency represented the approximate number of recommendation interactions in a standardized period. Completion and click-through outcomes were represented as percentages.

3.2 Variable operationalization

Recommendation relevance represents the extent to which surfaced content matches the user's current interests, goals, or inferred needs. In a real system this construct could be estimated through explicit ratings, dwell time, skip behavior, accepted recommendations, or offline relevance judgments. Personalization score represents how strongly the interface adapts the selection or ordering of content to the individual user. Novelty captures how different the recommended content is from the user's recent history while remaining potentially useful. Interaction frequency captures exposure intensity rather than recommendation quality.

User engagement is the primary behavioral outcome and represents active, sustained interaction with recommended content rather than simple page exposure. In a real platform it could combine session depth, dwell time, accepted recommendations, saves, meaningful clicks, and return behavior. Learning progression is the

second primary outcome and represents movement through an appropriate learning pathway. It is intentionally distinct from engagement: a user may be highly engaged without making meaningful progress, or may progress efficiently with relatively few interactions. Prior skill is included as a control because more knowledgeable learners may progress faster even when recommendation quality is held constant.

3.3 Synthetic data generation and analysis

The synthetic data-generating process was designed so that recommendation relevance and personalization were positively related but not interchangeable. Relevance was influenced partly by personalization and prior skill, with random variation added to preserve heterogeneity. Engagement was generated as a function of recommendation relevance, personalization, novelty, and a small fatigue penalty for very high interaction frequency. Learning progression was generated as a function of recommendation relevance, engagement, and prior skill. Random noise was included in all outcome equations to avoid deterministic relationships.

The analysis proceeds in four stages. First, descriptive statistics summarize the simulated user population. Second, Pearson correlations examine bivariate relationships among recommendation variables and outcomes. Third, personalization quartiles are compared graphically to inspect outcome gradients. Fourth, heteroskedasticity-robust ordinary least squares regression estimates independent associations with engagement and learning progression. Robust standard errors are used because user-level behavioral outcomes commonly exhibit non-constant variance. Statistical significance is reported for completeness, but p-values do not have their conventional inferential interpretation because the data were generated rather than sampled from a real population.

IV. Results

4.1 Descriptive patterns

Across 500 simulated users, the mean recommendation relevance score was 6.01 on a 10-point scale. Mean personalization was approximately 6.0, mean engagement was approximately 5.9, and mean learning progression was approximately 5.8. Recommendation relevance showed the clearest bivariate association with both outcomes. Its correlation with engagement was $r = .65$, while its correlation with learning progression was $r = .53$. Personalization was also positively related to engagement ($r = .57$) and progression ($r = .36$). Engagement itself was moderately associated with learning progression ($r = .51$).

Click-through rate was positively associated with relevance and engagement, but it was not treated as a substitute for either outcome. This distinction is intentional. A recommendation can generate clicks because it is visually prominent or curiosity-inducing, yet fail to support a coherent learning path. For that reason, click-through rate is retained as a supplementary interface metric rather than the primary criterion of recommendation success.

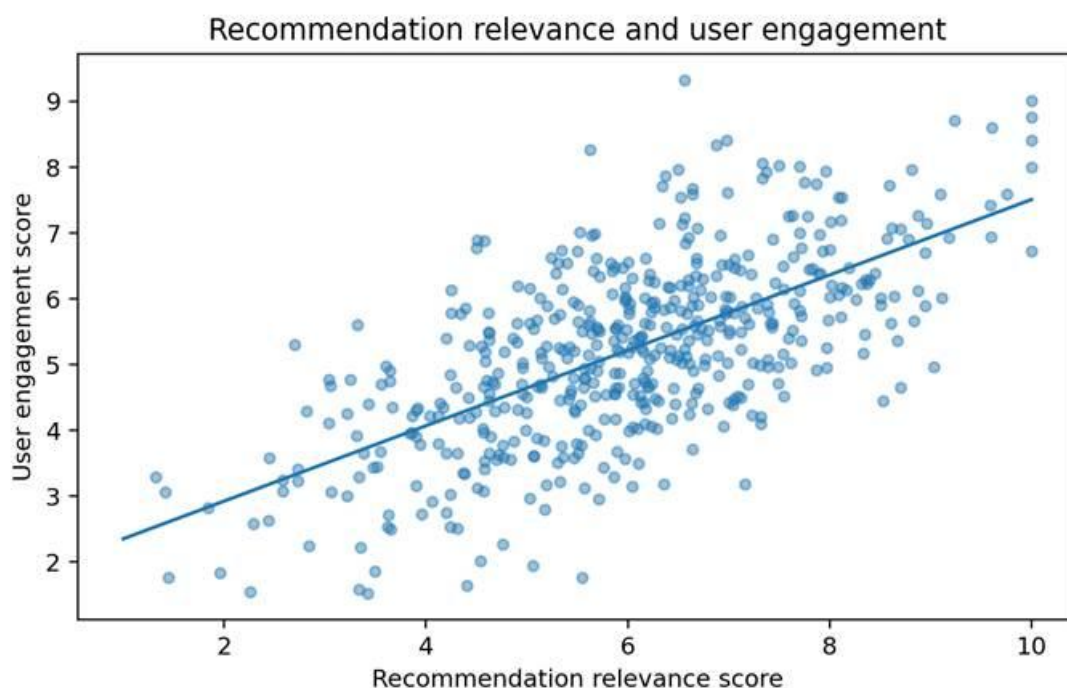


Figure 1. Recommendation relevance and simulated user engagement ($N = 500$).

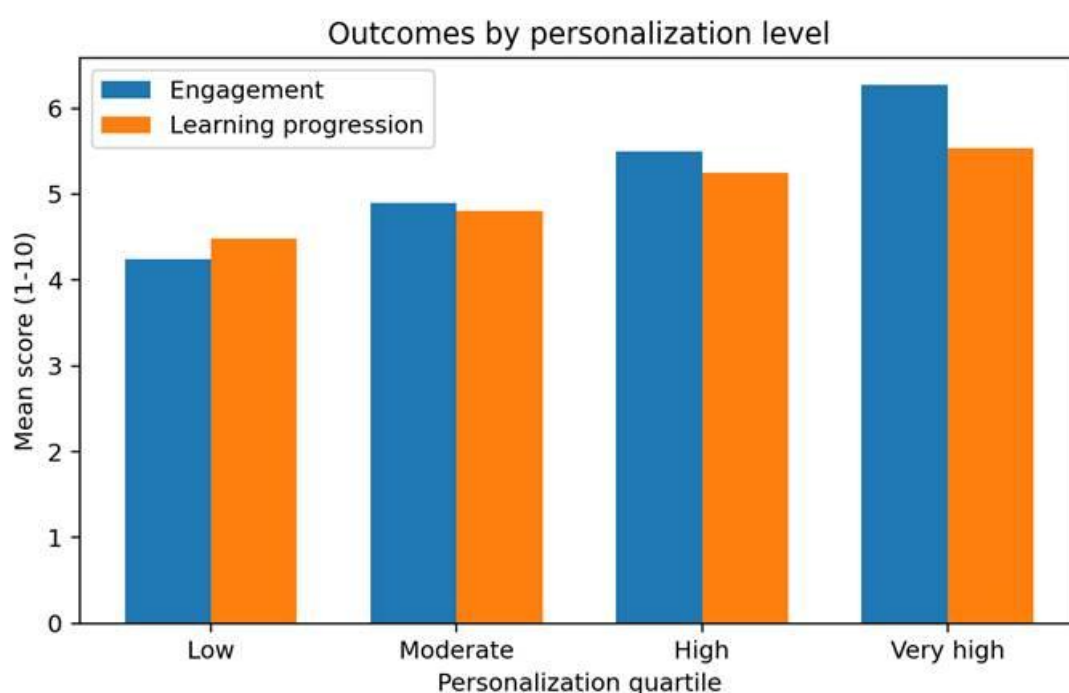


Figure 2. Mean engagement and learning progression by personalization quartile.

4.2 Multivariable models

In the engagement regression, recommendation relevance was the largest independent predictor ($b = .426$, $p < .001$). Personalization was also positively associated with engagement ($b = .209$, $p < .001$), as was novelty ($b = .088$, $p = .001$). Interaction frequency and prior skill were not independently significant once recommendation quality variables were included. The model explained 48.1% of the variance in simulated engagement. This pattern is consistent with H1 and supports the idea that the quality of surfaced recommendations matters more than simple exposure frequency.

In the learning-progression regression, recommendation relevance again remained significant ($b = .319$, $p < .001$). Prior skill was also an important predictor ($b = .176$, $p < .001$), while personalization had a smaller borderline association ($b = .063$, $p = .050$). Novelty did not reach conventional significance and interaction frequency was not independently associated with progression. The model explained 34.5% of the variance. These results support H3: relevance remained associated with progression even after controlling for prior skill and frequency of interaction.

Predictor	Engagement b	p	Progression b	p
Recommendation relevance	.426	< .001	.319	< .001
Personalization	.209	< .001	.063	.050
Novelty	.088	.001	.047	.062
Interaction frequency	-.012	.684	-.035	.285
Prior skill	.018	.548	.176	< .001

Table 1. Heteroskedasticity-robust regression coefficients for simulated engagement and learning progression. $R^2 = .481$ for engagement and $.345$ for progression.

V. Discussion

The synthetic results illustrate a simple but important principle for personalized digital interfaces: recommendation quantity is not recommendation quality. In both models, interaction frequency was weak once relevance and personalization were considered. By contrast, relevance consistently predicted the two outcomes that matter most to a learning interface: whether users remain meaningfully engaged and whether they progress through content. This is directionally consistent with the broader recommender-system literature, which has increasingly moved beyond pure prediction accuracy toward user-centered qualities such as usefulness, novelty, diversity, and serendipity (Jafari & Ravanmehr, 2021; Ping et al., 2025).

The engagement model also suggests that personalization and novelty can contribute independently of relevance. This is plausible because a personalized interface does more than identify relevant content. It determines how recommendations are surfaced, ordered, timed, and explained. A system may know that several resources are relevant, yet the interface still has to choose which one to expose first and whether to introduce something new. Novelty can support exploration, but only when it remains sufficiently aligned with user goals. The finding that novelty contributes more clearly to engagement than to progression reflects this distinction. New content can make an interface more interesting without necessarily improving learning outcomes.

The learning-progression model places prior skill alongside recommendation relevance. This is an important design implication. Educational recommenders cannot treat preference alone as the target. A learner may prefer advanced content that is currently unsuitable, or may repeatedly consume familiar material that produces little additional learning. Strong recommendation logic should therefore incorporate knowledge state, prerequisites, and goal structure rather than optimize only for short-term engagement. This conclusion is consistent with learning-path research that emphasizes learner models and content sequencing as central to personalization (Nabizadeh et al., 2020; Rahayu et al., 2023).

The paper also highlights why product teams should separate offline and online evaluation. Offline ranking metrics are useful during model development because they permit rapid comparison of algorithms. However, a model that performs better on precision or ranking loss may not improve the actual interface if recommendations are repetitive, poorly timed, or difficult to interpret. Online evaluation should therefore combine behavioral outcomes with pathway outcomes. For a learning platform, an A/B test might measure recommendation acceptance, session completion, return rate, mastery gain, and pathway completion simultaneously. This creates a more complete picture of whether personalization adds value.

A further implication concerns the design of recommendation explanations. Although explanation was not explicitly modeled in the synthetic dataset, personalized interfaces may benefit when users understand why a resource is being recommended. Explanation can increase transparency, provide orientation within a learning path, and allow users to judge whether the system's inferred goal matches their own. Future extensions could therefore include explanation quality as a moderator between recommendation relevance and engagement.

From a data-science perspective, the framework is deliberately compatible with multiple recommendation architectures. A collaborative model could estimate relevance from similar users, a content model could estimate semantic similarity between learner profiles and resources, and a hybrid model could combine both. Matrix factorization, nearest-neighbor methods, sequence models, graph-based models, or large-language-model representations could all feed the same interface-level evaluation framework. The product question remains whether those predictions improve user outcomes after they are translated into interface decisions.

VI. Practical Implications for Software Products

The analysis suggests four practical design principles. First, product teams should optimize for relevance before increasing recommendation volume. Showing more recommendations does not automatically produce better outcomes and can create interface noise. Second, personalization should be evaluated as an end-to-end experience, including ranking, timing, presentation, and pathway continuity. Third, learning products should include progression metrics alongside engagement metrics so that recommendation logic does not drift toward maximizing easy clicks. Fourth, exploration should be controlled rather than eliminated. Novel recommendations can increase engagement, but the level of novelty should be calibrated to user readiness and current goals.

These principles can be operationalized through a simple experiment. Users can be randomly assigned to a baseline interface and a personalized interface. The baseline can use popularity or static sequencing, while the treatment uses user-specific ranking. Primary outcomes can include accepted recommendations, meaningful dwell time, content completion, next-session return, and learning progression. Secondary measures can include novelty, diversity, satisfaction, and perceived usefulness. Such a design would provide substantially stronger evidence than a synthetic study and could directly test whether recommendation logic creates product value.

VII. Limitations and Future Research

The principal limitation is that all user-level observations are synthetic. The coefficients and p-values do not estimate real-world effects and should not be used to claim that a specific increase in recommendation relevance will produce a specific increase in engagement. The simulation demonstrates recoverability of the proposed analytical relationships, not empirical discovery. A submission that seeks to make population claims should replace the synthetic sample with real telemetry, experimental data, or an appropriately consented user study.

Second, complex constructs such as relevance, personalization, engagement, and progression are represented as simplified scores. Real systems should operationalize these constructs from observable behavior or validated user-reported measures. Relevance can vary by context and task, engagement can be active or passive, and progression should ideally reflect domain-specific learning objectives.

Third, the framework is cross-sectional. Recommender systems are dynamic, and users adapt to them over time. Longitudinal data would allow researchers to study preference drift, novelty decay, filter bubbles, changing knowledge state, and the long-term effect of repeated personalization.

Future work should therefore test the framework in a live digital learning environment. A stronger design would use a randomized controlled interface experiment, log recommendation exposures and accepted items, track pre- and post-learning performance, and compare algorithmic variants. Additional variables could include explanation quality, diversity, fairness, cold-start status, trust, and learner autonomy. This would connect recommender-system engineering more directly with human-computer interaction and educational outcome research.

VIII. Conclusion

Recommendation systems are increasingly embedded in the structure of digital interfaces, shaping not only what users discover but also the sequence in which they act. This paper proposes a practical evaluation framework for studying whether recommendation logic improves user engagement and learning progression within personalized digital learning products. The synthetic analysis indicates that recommendation relevance is the strongest common predictor of both outcomes, while personalization and novelty contribute more clearly to engagement and prior skill remains important for progression.

Because the dataset is simulated, the numerical results are not evidence about a real population. Their value lies in converting a broad product question into a testable data-science design. The next step is empirical: compare personalized and non-personalized interfaces using real user behavior, measure both engagement and pathway outcomes, and evaluate recommendation quality beyond accuracy alone. For software products, the central lesson is that successful personalization should not merely predict what a user is likely to click. It should reduce friction, improve discovery, and help the user move toward a meaningful next step.

References

- [1]. Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734-749. <https://doi.org/10.1109/TKDE.2005.99>
- [2]. Bourkhouk, O., Bachari, E. E., & El Adnani, M. (2023). The use of AI in e-learning recommender systems: A comprehensive survey. *Procedia Computer Science*, 224, 437-442. <https://doi.org/10.1016/j.procs.2023.09.061>
- [3]. Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331-370. <https://doi.org/10.1023/A:1021240730564>
- [4]. da Silva, F. L., Slodkowski, B. K., da Silva, K. K. A., & Cazella, S. C. (2023). A systematic literature review on educational recommender systems for teaching and learning: Research trends, limitations and opportunities. *Education and Information Technologies*, 28, 3289-3328.
- [5]. Jafari, R., & Ravanmehr, R. (2021). Serendipity in recommender systems: A systematic literature review. *Journal of Computer Science and Technology*, 36, 375-396. <https://doi.org/10.1007/s11390-020-0135-9>
- [6]. Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30-37. <https://doi.org/10.1109/MC.2009.263>
- [7]. Lampropoulos, G. (2023). Recommender systems in education: A literature review and bibliometric analysis. *Advances in Mobile Learning Educational Research*, 3(2). <https://doi.org/10.25082/AMLER.2023.02.011>
- [8]. Nabizadeh, A. H., Gonçalves, D., Gama, S., Jorge, J., & Rafsanjani, H. N. (2020). Learning path personalization and recommendation methods: A survey of the state-of-the-art. *Expert Systems with Applications*, 159, 113596.
- [9]. Ping, Y., Li, Y., & Zhu, J. (2025). Beyond accuracy measures: The effect of diversity, novelty and serendipity in recommender systems on user engagement. *Electronic Commerce Research*, 25(3).
- [10]. Rahayu, N. W., Ferdiana, R., & Kusumawardani, S. S. (2023). A systematic review of learning path recommender systems. *Education and Information Technologies*, 28, 7437-7460.
- [11]. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of netnews. *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, 175-186. <https://doi.org/10.1145/192844.192905>
- [12]. Raza, S., et al. (2025). Review-based recommender systems: A survey of approaches, challenges and future perspectives. *ACM Computing Surveys*, 58(1), Article 25. <https://doi.org/10.1145/3742421>