Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Explainable Course Recommendations for Students Using Knowledge Graphs and Generative AI

Syed Aamer Hashmi

amer.syed20@gmail.com

Amity University

Yashpal Singh

Amity University

Harshit Bhardwaj

Amity university, Noida

Research Article

Keywords: Personalized Learning, Course Recommendation, Knowledge Graph, Computational Modelling, Recommender Systems, Retrieval-Augmented Generation, Explainable AI

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11021419/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

There are more self-paced courses than ever before. For students, course recommendations can become a daunting task. Most recommendation systems rank courses by similarity or popularity, but who knows why? Students are left choosing according to faith and not with an understanding. This paper proposes a recommendation system consisting of a model including a student's academic background, skills, and goals. The model eases the opaqueness of recommendation systems and provides explanations of course recommendations. The model consists of a multi-relational knowledge graph and a generative AI platform. The knowledge graph consists of students, courses, skills, and feedback. The generative AI platform retrieves relationships pertaining to the system's recommendations and explains them to the student. The system reduces query times as catalogues grow using an efficient similarity-search index. The model was evaluated against five recommendation system benchmarks using NPTEL, Coursera, and Udemy. Over 1,200 student profiles and 200 courses were anonymized for testing. Our model significantly increased accuracy, coverage, and overall explainability compared to the benchmarks. We consider our testing to be a proof of concept and do not assume our results reveal performance as a whole. We believe the results show increased confidence in students, as these recommendations are no longer based on similarity but are grounded in actual educational philosophy.

1. Introduction

In this age of the internet, online education has evolved systems such as Udemy, Coursera, NPTEL, and YouTube. Students can choose from numerous resources to aid their learning. However, with so many options, students choose poorly or sometimes don't choose at all. This phenomenon is known as decision fatigue. Other systems have tried to alleviate this problem with traditional algorithmic methods known as content and collaborative filters [1]. For example, if a user likes course A and course B, our algorithm will suggest course C. However, these conventional methods do not account for academic background, learning behavior, or career goals.

Most users find recommendation systems to be black boxes. For example, they suggest a course but do not explain why they think the course is best for the student [2]. In the long run, black box recommendations hurt the overall value of the system and lead to a decrease in user trust. To design a system that is catered to students' needs, we must provide the course recommendation as well as convince a student why that course will be a step towards the student's goal. We must also use heterogeneous data to justify our recommendation. This includes integrating NPTEL, Udemy, Coursera, and other course data into a single, justified recommendation model.

To address these challenges, this paper proposes a Knowledge Graph-enhanced Hybrid Course Recommender System (KG-HCRS). The system integrates knowledge graphs, RAG, and HNSW-based similarity search to recommend courses. It organizes structured and unstructured course information into a knowledge graph that represents the connections between student profiles, subjects, learning styles, and course characteristics with rich contextual semantics. RAG is then used to dynamically

retrieve the most relevant knowledge-graph nodes and generate transparent, personalized explanations [3], while HNSW provides computational scalability for large datasets. This combination is intended to both increase recommendation precision and ensure that every recommendation can be traced, decoded, and aligned with a learner's long-term objectives.

The main contributions of this study are as follows:

1. **1. Hybrid Recommendation Architecture**: A unified framework that combines knowledge-graph reasoning, semantic-embedding similarity search, and retrieval-augmented generation to provide personalized and explainable course recommendations.
2. **2. Student–Course Knowledge Modelling**: A multi-relational knowledge graph built from connections between students, courses, topics, skills, instructors, and feedback drawn from multiple online learning platforms.
3. **3. Efficient Semantic Retrieval with HNSW**: Hierarchical Navigable Small World indexing is used to accelerate similarity search over a high-dimensional embedding space, enabling scalable, near real-time recommendations.
4. **4. Explainability via RAG**: A retrieval-augmented generation pipeline that produces natural-language explanations for every recommendation, grounded in knowledge-graph relationships and course metadata.
5. **5. Hybrid Evaluation Framework**: The system is assessed using standard recommendation metrics (precision, recall, F1, NDCG) alongside novelty and a composite explainability score, with statistical significance verified against each baseline individually.

Collectively, these contributions work toward transparent and intelligent learning recommender systems that combine structured knowledge reasoning with generative AI.

2. Literature review

2.1 Knowledge Graph-Enhanced Recommender Systems

The semantic structure provided by knowledge graphs in recommenders has changed the game in terms of educational technology, offering more contextualization and personalization. Zhang et al. [4] analyzed different ways of developing knowledge graph embedding techniques in the context of learning processes. Some previous work [5] focused on utilizing knowledge graphs to create flexible learning path recommendations to cater to different learning needs and be adaptive.

Knowledge graphs provide plenty of opportunity to evolve learning systems. [6] showed that knowledge graphs can be scaled to e-learning systems that contain big academic archives and can suggest tailored courses. Surveys of explainable knowledge-graph-based recommendation models show their benefit of explainability over black-box deep-learning models [7]. This has created a preference toward adaptive, semantic, interpretable, and flexible systems compared to rule-based systems. Most knowledge-graph-

based techniques lack adaptability and have problems with keeping up with new educational resources, which necessitates integrating them with generative retrieval systems.

2.2 RAG in Education

Large language models have put a spotlight on the deficiencies of static knowledge graphs, especially on their unchanging nature and the challenges that make the representation of shifting educational data impossible. RAG leverages generative reasoning together with real-time retrieval to provide both up-to-date and contextual recommendations. The RAG framework was presented by (Wang et al., 2025) as a mixture of parametric language modeling and retrieval to deliver high performance on knowledge-intensive tasks, and they were the first to explore this paradigm. After them,[9]] studied RAG and its scalability, contextuality, and situational adaptability in the field of education.

[10] proved that RAG-derived systems maintain their relevancy against evolving educational data. [11] and [12] advanced this by combining knowledge graphs in retrieval to reduce hallucinations and improve factual correctness in student-focused recommendation systems. In spite of these improvements, most of the current RAG-based educational systems lack pedagogical substance and rational explanation for their outputs.

The combination of RAG and knowledge graphs offers an appealing pathway for ensuring trustworthy and rational educational recommendation systems.

2.3 Explainable AI and Student-Centric Justification

[13] and [14] described models that capture learning process data. [15], [16], and [17] invented educational AI systems augmented with retrieval and explanation. [18] integrated explainable learning analytics with automated feedback. Accountability, fairness, and clarity of rationale in recommendation systems have been researched by more recent studies [19], [20], [21], [22], [23], [24]. From these studies, explainable outputs are instrumental in building user trust and enhancing learning engagement. Most of the explanations in the literature are model-centric and lack educational frameworks. Therefore, there is a necessity for frameworks that provide interpretable outputs and student-centric educational explanations.

2.4 Synthesis and Research Gaps

The preceding literature reviewed some of the recent advancements in hybrid recommendation systems that incorporate structured knowledge representation, dynamic retrieval, and explanation. There are other research gaps that remain to be addressed as follows:

1. **1. The gap of integration**: Research on methods to integrate knowledge-graph reasoning with content-based factual grounding of RAG generation remains unexplored.
2. **2. The gap of explanation**: Most of the explanation research is focused on relevance and content, and other educational related factors are being ignored.

3. **3. Assessment Disparity**: Scalable, longitudinal, and student-centered assessment frameworks for measuring interpretability and learning effectiveness continue to be severely lacking.

Personalization in recommender systems has improved, but issues still persist, specifically in the educational domain. Many collaborative-filtering and deep-learning based systems over-prioritize predictive accuracy at the expense of pedagogical and overall system transparency and interpretability. Although some recommenders provide rationales for their recommendations, they often still lack relevant explanations in the context of educational systems and services. This is problematic and limits the adoption of even highly accurate systems.

Furthermore, existing approaches to retrieval-augmented recommendation systems have focused primarily on semantic retrieval, and have largely neglected Prerequisite dependency, domain hierarchies, and progressions of conceptual learning. Knowledge graphs, partially, address this problem by modelling such relations, but still require a significant number of generative rationales that learners can understand. Most existing models assume high recommendation accuracy in an educational repository; however, they become impractical in the field due to high retrieval latency and poor scalability. The KG-HCRS framework addresses all these issues by integrating semantic retrieval, knowledge graph reasoning, large-scale retrieval with HNSW, and RAG explanation within a unified educational recommendation system.

3. Proposed work

The proposed system starts with collecting data from different sources: structured syllabi from NPTEL, student feedback from Udemy and Coursera, and student profiles with grades and learning styles. Data are then cleaned and tokenized to address inconsistency at the semantic level. Once data concerning courses and students are collected, both are embedded using transformer-based sentence embeddings of Hugging Face (Sentence-BERT).

The system develops a Knowledge Graph in which nodes represent students, courses, skills, and feedback. Edges connect nodes based on a relationship of semantics as well as performance (e.g. prerequisites, similarity, interest). To enable construction of the graph and answer similarity queries efficiently, the embedding space is indexed using HNSW.

A RAG pipeline is then executed over the Knowledge Graph. A retriever identifies the most relevant nodes and course material. A large language model is then utilized as a generator to produce personalized contextual explanations. Finally, the system provides recommendations and performance metrics (precision, recall, F1, and NDCG) through a Streamlit front end.

Figures 1 shows the KG-HCRS proposal's end-to-end process. Data are collected from many sources (profiles of students, metadata of the courses, and feedback), are then pre-processed, and finally, semantic embeddings are created using Sentence-BERT. The embeddings are used to build a knowledge graph of students, courses, skills, and topics. HNSW is used to index the embeddings, while a

combination of semantic similarity and reasoning based on the knowledge graph is used to determine the most relevant courses. The results are sent to the RAG-based explanation engine for the generation of relevant subgraphs, along with context, which is used as input to Claude Sonnet 4.5 while the Anthropic API is used for the explanation of the recommendations. The explanations are assessed based on coverage, specificity, and coherence.

3.1 knowledge graph construction

Knowledge graphs capture semantic relationships among educational entities. Students, courses, topics, skills, and instructors are represented as nodes, and relationships between them — such as enrolled in, teaches, requires skill, or related to — are represented as edges. Formally,

KG = (V, E) (1)

where $V = \{v_1, v_2, \ldots, v_n\}$ are nodes and $E \subseteq V \times V$ are edges. Each node v is embedded as a dense vector using a transformer encoder:

$e_v = f_{LLM}(v)$ (2)

where the input to f_{LLM} is the textual attributes of node v (i.e., course description, student profile). To compare the implied meaning or context of different course descriptions, and therefore enable semantic retrieval as opposed to simple keyword matching, Sentence-BERT is employed.

Course nodes c_i are linked to topic nodes t_k with weighted relationships:

$w(c_i, t_k) = \text{frequency}(t_k \in c_i) / \text{total topics in } c_i$ (3)

Where the frequency term addresses the distinction of the more significant over the more general topic. This weighting allows us to measure the alignment of a student's interest and the content of a course.

3.2 Student Profiling and Embedding Generation

Each student s is modeled as a profile vector consisting of the student's academic history, behavioral engagement, and expressed aspirations:

$p_s = \alpha \cdot g_s + \beta \cdot h_s + \gamma \cdot i_s$ (4)

where:

- g_s: academic performance embedding (CGPA, subject strengths)
- h_s: behavioural embedding (learning pace, time spent on topics)
- i_s: interest embedding (career goals, preferred domains), with $\alpha + \beta + \gamma = 1$

Each vector is obtained from normalized survey responses and traces of learning activities. The weights α, β, and γ are optimized by grid search through the correlation with the ground truth relevance labels described in Section 4.2.

3.3 Similarity Computation and Course Relevance

Cosine similarity is the main similarity metric used:

$R(s, c) = \cos(\mathbf{p}_s, \mathbf{e}_c)$ (5)

Cosine similarity supports soft, semantic matching between student interests and course content, even if course names differ lexically. Consider courses titled “Introduction to Deep Learning” and “Neural Network Fundamentals” which contain a high level of semantic similarity as their respective embeddings may reside in proximal regions of the latent space. Semantic similarity, however, does not lend itself to straightforward interpretation, which is why we will incorporate knowledge-graph path reasoning.

3.4 Knowledge Graph Path Reasoning

It is possible to provide interpretable paths across students and courses for the knowledge graph. An example of such interpretable paths might be Student → possesses skill → Machine Learning → prerequisite for → Deep Learning for Beginners. Each recommendation is explained with such a reasoning path. To assess the reasoning, we introduce the following notation:

$\text{PathScore}(s, c) = \sum_{p \in P(s,c)} \lambda^{|p|} \cdot w(p) \cdot \prod_{(u,v) \in p} \text{sim}(u, v)$ (6)

where $P(s, c)$ represents all paths of student s to course c and $|p|$ is the number of edges of path p. $\lambda \in (0, 1]$ is a decay factor to discount longer, less direct paths, $w(p)$ is a weight attributed to the sequence of relation types of the path considered in its context, and the term inside the product aggregates the semantic similarity $\text{sim}(u, v)$ of all pairs of nodes along the paths. In the context of course recommendations, textual similarity cannot be the primary basis of recommending a course. Courses should be recommended along paths that are constructed with a clear educational focus to the taught content or required skills.

3.5 Hybrid Scoring Function

To evenly score both similarity and reasoning, a hybrid function incorporates both:

$\text{Hybrid}(s, c) = \lambda \cdot R(s, c) + (1 - \lambda) \cdot \text{PathScore}(s, c)$ (7)

where λ adjusts the balance between numerical similarity and graph-based reasoning and will be optimized using grid-search on the validation data (see Section 4.5 for the sensitivity analysis).

3.6 Efficient Retrieval via HNSW

For small course sets, retrieval time is still manageable, but it is expensive for larger course sets. To address this, an HNSW index is used, reducing the overall retrieval complexity from O(n) to O(log n).

Given a query vector p⃗, the algorithm returns the top-k most relevant course candidates as:

Ĉ⃗ = Top-k{HNSW(p⃗, C)} (8)

where Ĉ⃗ denotes the top-k nearest neighbours among course embeddings C. This significantly improves response time, while still providing high recall, compared to exact search, as explained in Section 4.6.

3.7 RAG for Explainability

When explaining a recommendation, it is insufficient to only provide a similarity score. In order for a student to be able to act on the explanation, it is also necessary to provide reasoning of how a course fits a student's profile. With this in mind, we propose a RAG pipeline that combines a knowledge-graph to retrieve the subgraph that connects a student to a recommended course along the reasoning path and provides this subgraph along with course metadata to the generative model.

The explanation is generated by Claude Sonnet 4.5 using the Anthropic API with a prompt designed to reference only the given evidence. Generating explanations based on graph evidence, as opposed to allowing the model to generate freely from the course title, should reduce instances of hallucinated justifications and enable explanations to be traced back to knowledge graph relationships.

3.8 Explainability Scoring

Generated explanations are scored along the dimensions of Coverage, Specificity, and Coherence. The dimensions are summarized into a single score, Explainability Scoring (ES), using the following equation:

$ES = w_1 \cdot Coverage + w_2 \cdot Specificity + w_3 \cdot Coherence$ (9)

Where w1 = w2 = w3 = 0.33. Equal weighting was used as a simplifying design choice rather than one that is refined for a predefined validation goal, since we assumed that the dimensions are of similar importance to explainability and will elaborate on this in Section 4.8.

Coverage and Specificity are computed using ROUGE-L and BLEU-4 in relation to template-based reference explanations. These references were constructed automatically for each student-course pair by filling in a template constructed from the corresponding path (student interest, matched skill, prerequisite relationship, and topic alignment) in the knowledge graph. This has the advantage of being deterministic and reproducible; however, it means that in this study, ROUGE/BLEU scores measure agreement with a structured, template-based reference as opposed to human authored explanations. We discuss this consideration in detail in Section 4.8. Coherence is measured utilizing an automatic fluency metric on the generated text.

3.9 Continuous Learning and Feedback Integration

Feedback is used to optimize knowledge graphs and model weights: for each recommendation cycle, user ratings and click data are collected and added to the knowledge graph as edges or updated edge

weights:

$w_{new}(c) = w_{old}(c) + \beta \cdot Feedback(s, c)$ (10)

This facilitates a personalization model that adapts and improves automatically over time.

3.10 Implementation and Methodological Details

The KG-HCRS framework is implemented via knowledge graph modeling, semantic embedding, efficient retrieval, and generative reasoning. The knowledge graph is constructed from metadata of course offerings and students, as well as feedback data, all of which are available on multiple online learning platforms. The nodes of the knowledge graph are students, courses, skills, and topics, whereas the instructors are considered as relations. Directed edges are used to model the relationships of interest.

To generate 384-dimensional dense vectors, we used Sentence-BERT (all-MiniLM-L6-v2). Stop words are removed and the text is lemmatized and tokenized prior to embedding. For efficient similarity search, HNSW indexing is applied via a vector database. Index parameters were selected to balance retrieval accuracy and computational efficiency: maximum number of connections M = 16, efConstruction = 200, and efSearch = 200.

The hybrid recommendation system combines embedding-based similarity and knowledge-graph path relevance, controlled by the parameter λ. A sensitivity analysis of λ revealed that λ = 0.6 was the best trade-off point for the combined explanatory score and F1 (Section 4.5).

The RAG pipeline of explanation generation integrates subgraphs and contextual information using semantic similarity and graph traversal. This information becomes input context for Claude Sonnet 4.5 and produces explanations grounded in prerequisite alignment, topic alignment, and student interests. As a proof of concept, the main functionality of the pipeline is accessed in a Streamlit interface that captures relevant academic, interest, and goal information from the user. The modular design of the interface easily allows for additional technology and services to be incorporated. The pipeline enhances robustness through validation stages during retrieval and ranking. Before hybrid scoring is applied, the prerequisite compatibility and domain alignment of the course candidates is ensured, thereby limiting the recommendation of courses that are semantically similar but inappropriate from a pedagogical standpoint. In the hybrid scoring, similarity and graph relevance scores are embedded, and both are normalized to ensure equal weight in the scoring function. The design is modular to incorporate new types of entities and supports independent tuning of each unit (e.g. retrieval, explanation generation) in various deployments.

4. Experimental results

4.1 Experimental Setup

The proposed KG-HCRS framework was evaluated using a combination of real-world and curated data sources.

Course Dataset:

Public data from NPTEL, Coursera, and Udemy were used to collect data from 200 publicly available courses spanning 10 major areas, which include data science, machine learning, programming, and artificial intelligence. Data for each course will include the title, description, skills, user rating, and user comments.

Student Dataset:

Student data include 1200 de-identified student records hailing from a private college that is associated with the University of Mumbai. Data for each student include course preference, time spent learning, level of engagement, and career interest. Before data analysis, all Personally Identifiable Information (PII) was removed. Overall, the experimental data includes 1200 de-identified student records and 200 course records from NPTEL, Coursera, and Udemy.

Ethical Considerations:

The student data was recorded within the framework of the institutional policies on data usage. The data was then made anonymous before analysis. No identifiable information was included. The dataset was made available to the researcher by the institution. The dataset is stored securely and is used for the aggregate analysis of the data within this paper.

Data Preprocessing:

Text data such as course descriptions and feedback was processed with tokenization, stop word removal, and lemmatization. Numerical data was procedure with min-max normalization. Categorical data was treated with label encoding.

Cross-Validation:

All experiments were repeated 5 times. The average performance and standard deviation are reported to describe the robustness. Additionally, 5-fold cross-validation was employed to reduce overfitting and assess the performance of the model considering various data splits.

Embedding Generation:

Student profiles and course descriptions were processed by the Sentence-BERT (all-MiniLM-L6-v2) to produce 384-dimensional dense semantic embeddings.

4.2 Ground Truth Construction

Since no data on student-course enrollments or direct ratings of courses were available, a ground-truth relevance set was created utilizing an educational relevance framework based on rules. The aim of this protocol was to ensure that reliability of relevance labels was achieved between student profiles and courses and was used to define the recommendation model.

Four criteria based on educational compatibility were evaluated: alignment of domain interest, satisfaction of prerequisites, compatibility of Bloom's Taxonomy, and student competencies and course learning objectives. Each criterion produces a normalized score between 0 and 1. The overall relevance score is computed as:

Relevance(s,c) = 0.35·Interest + 0.30·Skill + 0.20·Bloom + 0.15·Prerequisite (11)

Courses with a relevance score ≥ 0.70 were considered relevant to that student, comprising the ground-truth set used for the evaluation throughout the assessment.

4.3 Evaluation Methodology

The metrics used in this research for model evaluation are summarized in Table 1.

Table 1
Evaluation metrics used to evaluate the proposed work

Metric	Purpose	Formula	Interpretation
Precision	Fraction of top-k recommended courses relevant to the student	TP / (TP + FP)	Higher = better
Recall	Fraction of relevant courses successfully retrieved	TP / (TP + FN)	Higher = better
F1 Score	Harmonic mean of Precision and Recall	2·(P·R) / (P + R)	Balances precision and recall
NDCG	Ranking quality of the recommendation list	DCG / IDCG	Higher = better
Hit Rate@K	Whether at least one relevant course appears in the top K	$(1/N)\cdot\Sigma\ I(Rel_i > 0)$	Indicates presence of ≥ 1 relevant course
Explainability Score	Interpretability of the generated recommendation explanation	ES = 0.33·Coverage + 0.33·Specificity + 0.33·Coherence	Evaluates generated explanation quality
MAP	Mean Average Precision across queries	MAP = (1/Q)·Σ AP_q	Measures ranking consistency
MRR	Mean Reciprocal Rank of the first relevant item	MRR = (1/Q)·Σ (1/rank_q)	Rewards early placement of relevant courses
Novelty	Degree of non-obvious, less-popular recommendations	Novelty = (1/|R|)·Σ − log P(course)	Higher value indicates less repetition of popular courses

4.3.1 Recommendation Accuracy

For every student, the recommendation system provides the top-five ranked courses using the hybrid ranking method. This set is then examined against the ground-truth relevant courses created in Section 4.2. A recommended course that is present in the student's ground-truth set is considered a true positive (TP); a recommended course that is absent from the student's ground-truth set is a false positive (FP); a relevant course that is not recommended is a false negative (FN).

4.4 Baseline Models for Comparison

To assess the novelty of KG-HCRS, it was compared against five established baselines:

Table 2
List of baseline models

Model	Description
CBF	Content-Based Filtering using TF-IDF embeddings
CF	User-Item Collaborative Filtering
NCF	Neural Collaborative Filtering using an MLP
KGRec	Knowledge-graph-based recommender using TransE embeddings
LLM-RAG	Plain RAG-based recommender without HNSW or knowledge-graph structure

4.5 Quantitative Results

Performance across the models is summarized in Fig. 2 below.

When combining semantic embeddings, knowledge-graph reasoning, and generative explanation, the improved performance in KG-HCRS can be attributed to the integration of the knowledge base. Higher NDCG means improved ranking quality – the system is ranking relevant courses higher - whereas the increase in novelty indicates a greater tendency to recommend new courses that are not commonly selected.

The RAG module has shown significant improvement in explainability. Because KG-HCRS is based on structured semantic relationships rather than similarity or interaction types, it is ideally placed to produce more interpretable, pedagogically relevant suggestions than baselines, although, as we say in Section 4.8, the explainability metric itself is based on references derived from templates and should be approached with caution.

Figure 3 shows the Sensitivity analysis of the hybrid parameter λ, it is absolved that KG-HCRS consistently outperformed all baselines, with the largest relative gains in Explainability (+ 16 percentage points over the strongest baseline, LLM-RAG), Novelty (+ 13 percentage points), and NDCG (+ 6 percentage points). This supports the premise that integrating knowledge-graph semantics with RAG reasoning and HNSW retrieval improves precision as well as interpretability and diversity of recommendations.

4.6 Ablation Study

To assess the contribution of each component, ablation experiments were performed by removing one module at a time. HNSW is an approximate nearest-neighbour index computed over the same embedding space used by exact search; removing it changes the retrieval mechanism from approximate to exact but does not change the underlying embeddings, the knowledge graph, or the RAG explanation module. Accordingly, its ablation is evaluated separately as shown in Table 3, on retrieval speed and recall against exact search, rather than being combined with the accuracy-bearing KG and RAG ablations below.

Table 3
Ablation study results of knowledge graph and RAG components

Configuration	Removed Component	F1	Explainability Score
Model A	Without Knowledge Graph	0.79	0.66
Model B	Without RAG	0.81	0.58
Full Model (KG + RAG + HNSW)	—	0.86	0.91

Interpretation

Removing the Knowledge Graph reduces both F1 and explainability, since path-based reasoning contributes to relevance ranking as well as to the evidence used for explanation. Removing RAG has a smaller effect on F1 but a larger effect on explainability, consistent with RAG's role being primarily to generate the natural-language justification rather than to influence the ranking itself.

Table 3
b: HNSW retrieval ablation — approximate vs. exact search

Configuration	Retrieval Time (ms/query)	Recall@10	F1
Exact (brute-force) search	55	0.92	0.86
HNSW (approximate)	42	0.90	0.83

Interpretation

As shown in Table 3b, Exact search gives slightly better results than HNSW

4.7 Computational Efficiency

As shown in Fig. 4, The average retrieval time for HNSW is 49 ms, while for CF it is 192 ms, resulting in a 3.9x improvement. This shows significant difference especially for real-time educational recommendation systems.

In education, this computational efficiency is useful for large course repositories that students interact with over time. With an increasing size of repository and dataset, exhaustive similarity search becomes expensive. Fortunately, HNSW efficiently allows nearest neighbours retrieval, resulting in recommendations that approach real-time. This improves education that uses adaptive systems, where recommendations have to be made based on a student’s new interactions, assessment results, and changing personal goals.

This system provides both efficient retrieval and explanatory reasoning. This provides a suitable trade-off between scalability and interpretability, which is usually a shortcoming of existing explainable recommender systems. The positive outcomes of the architecture indicate that the designed system is both academically and practically viable for use in digital learning environments.

4.8 Statistical Significance

To determine the statistical significance, a paired t-test was conducted for five independent runs with unique random seeds for each of the baselines relative to KG-HCRS. Given that this is a case of five pairwise comparisons for each metric, the family-wise error rate was adjusted by the Bonferroni method. Therefore, the significance threshold for the individual tests was set to $\alpha = 0.05 / 5 = 0.01$.

Table 4
Pairwise significance comparison (KG-HCRS vs. each baseline)

Metric	vs. CBF	vs. CF	vs. NCF	vs. KGRec	vs. LLM-RAG
Precision	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p = 0.006$	$p = 0.008$
NDCG	$p < 0.001$	$p < 0.001$	$p = 0.002$	$p = 0.009$	$p = 0.009$
Explainability	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p = 0.003$	$p = 0.005$

From these comparisons, the amount of improvement in KG-HCRS is the largest when compared against the weaker, non-graph-based baselines (CBF, CF, and NCF), and is still improved yet more favorable when compared with the strong baseline (LLM-RAG) that includes an RAG-based explanation mechanism but lacks the knowledge graph and HNSW retrieval.

These results place KG-HCRS in a favorable position when compared with other explainable educational recommender systems. Almost all of the collaborative and content-based filtering models are optimized for predictive accuracy, and treat explainability as secondary. However, KG-HCRS focuses on recommendation explanations, that are both generated and justified.

By embedding structured syllabi and student feedback in addition to course metadata in a knowledge graph, and by using RAG for reasoned retrieval, the system achieves both retrieval accuracy and explanatory capability. Additionally, the deployment of KG-HCRS at scale is supported by the HNSW based retrieval, which is often not discussed in the literature of explainable recommenders. The hybrid explainability measure proposed here, ES(s, c), offers the beginning for a repeatable, quantitative framework that can be enhanced by evaluating it against human developed explanations, so that it can be improvements upon the explanations generated by templates.

Limitations:

Although the results are positive, there are some constraints to be mindful of. First, the evaluation dataset is of a single, prototype-scale academic environment, and therefore may not reflect more complex and larger educational environments. Second, the quality of recommendations is dependent on the completeness of external course metadata and feedback data. Third, because ground-truth relevance was constructed using a rule-based formula (Section 4.2) rather than observed student behaviour such as enrolments or ratings, the reported precision, recall, F1, and NDCG values should be interpreted as measuring agreement with this defined relevance construct rather than with actual

student choices; validating the rule-based relevance framework against real enrolment or satisfaction data is an important direction for future work. Fourth, because the Explainability Score's component weights are equal by design rather than empirically tuned, and its ROUGE/BLEU components are computed against template-derived rather than human-written references, the absolute explainability values reported here should be treated as internally comparable across models within this study rather than as a benchmark comparable to explainability scores reported elsewhere. Finally, RAG-based explanation improves interpretability but does not guarantee good performance when the retrieved contextual information is itself of low quality.

5. Discussion

As defined in Section 4.5, it is clear that KG-HCRS outperformed both traditional and neural recommender systems. According to these metrics, we can see the largest gains in explainability and novelty. As mentioned earlier, human empirical knowledge can be integrated into knowledge graphs, and can amplify the strengths of expert systems. Knowledge graphs provide reasoning paths supported by empirical knowledge and can be easily explained. In contrast, RAG systems transform those reasoning paths into operating explanations.

Due to the scoring function that combines embedding-based semantic similarity and path-based relevance in a graph, the evaluated system has more precision and NDCG. Because the system evaluates courses that are related and accessible via an educational path, it ranks courses at a more appropriate level than courses which are related in a surface-based manner. This is of great importance for work or study related decisions in the field of higher education as the selected course can affect the career of a student.

The gains in explainability support research that exists in the literature on educational data mining. The trust and engagement of learners in a recommender system is supported by the inclusion of explanations to the learners, and not the accuracy of the system. LLMs can produce explanations that are fluent but are also likely to contain unsubstantiated claims. Grounding generation in retrieved knowledge-graph evidence as done in this project, was aimed at explanations that are related to evidenced graph relations e.g., prerequisites, skills, and topics.

Overall, these results suggest that KG-HCRS is one example of a growing class of educational decision-support systems that integrate symbolic reasoning, semantic retrieval, and generative explanation. As mentioned in Section 4.8, the dataset used in this study is from one institution, and both the relevance assessments and the explainability metric are based on constructed definitions (a rule-based scoring formula and template-based references, respectively) rather than on student behaviour or written explanations. Therefore, the next step is to validate the framework with multiple-institution data and with human-generated explanations.

6. Conclusion and future scope

The research introduced KG-HCRS, a hybrid course recommender that leverages knowledge graph path reasoning, semantic (Sentence-BERT) embeddings, HNSW, and RAG for explanation to deliver personalized course recommendations that are both accurate and explainable. KG-HCRS outperformed five benchmarks on a dataset containing 1200 student profiles and 200 courses, achieving higher precision, recall, F1, NDCG, novelty, and explainability scores. These results were confirmed to be statistically significantly in pairwise comparisons against each baseline (Section 4.8).

The main focus of the framework is architecture. This is particularly significant in the area of educational decision-support systems, where both accuracy and explainability tend to be addressed in a separate manner. From the ablation study in Section 4.6, both the KG and RAG components of the system make significant contributions to the explanation elements of ranking, with RAG bringing a greater impact to explanation quality, while the HNSW index brings significant efficiency to search and ranking without affecting the ranking logic.

Several directions for future work follow directly from the limitations noted in Section 4.8:

- Multi-institutional and larger-scale validation, replacing the single-institution, prototype-scale dataset used here with data spanning multiple institutions and, where possible, real enrolment or rating signals rather than rule-based relevance labels.
- Human-evaluated explainability, supplementing the template-referenced ROUGE/BLEU protocol used in this study with human judgments of explanation quality, and empirically tuning the Explainability Score's component weights rather than treating them as equal by design.
- Temporal and adaptive personalization, incorporating dynamic knowledge-graph updates and temporal graph neural networks (potentially combined with reinforcement learning) so that recommendations adapt as a student's interests and progress evolve.
- Multilingual and cross-domain extension, since the current framework relies on structured course descriptions in English within a specific academic domain.
- Fairness-aware recommendation, to identify and mitigate potential bias arising from academic performance, prior exposure, or domain popularity.
- Integration with institutional learning management systems, to support real deployment and continuous optimization based on live feedback.

Taken together, these results suggest that KG-HCRS is a promising step toward intelligent, transparent, and scalable educational recommender systems, while the limitations identified above define a concrete path for strengthening the evaluation methodology in future iterations of this work.

Declarations

Ethics Approval and Consent to Participate

The study protocol was reviewed and approved by the Ethics Committee of Anjuman-I-Islam Kalsekar Technical Campus, Navi Mumbai, India. The research was conducted in accordance with the institutional ethical guidelines for research involving human participants. Informed consent was obtained from all participants. Participation was voluntary, and all data were anonymized to ensure confidentiality and privacy.

Consent for Publication

Not applicable.

Informed Consent

Informed consent was obtained from all participants. All personal identifiers were removed, and the data were used solely for academic and research purposes.

Competing Interests

The authors declare that they have no competing interests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Author Contributions (CRediT)

Syed Aamer Hashmi: Conceptualisation, Methodology, Software, Formal analysis, Writing — original draft. Yashpal Singh: Supervision, Validation, Writing — review and editing. Harshit Bhardwaj: Investigation, Data curation, Visualisation, Writing — review and editing.

Use of Artificial Intelligence (AI) Tools

The authors used Claude Sonnet 4.5 (Anthropic) as the generative component of the proposed system itself, as described in Sections 3.7 and 3.10. Separately, the authors used AI-based writing tools, including ChatGPT and QuillBot, to improve language clarity and readability during manuscript preparation. All scientific content, methodology, experimental design, and conclusions were developed, reviewed, and validated by the authors.

References

1. Slim A, Hush D. An Automated Framework to Recommend A Suitable Academic Program, Course and Instructor, 2019 *IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*, pp. 145–150, 2019. 10.1109/BigDataService.2019.00026
2. Channarukul S, Saejiem N, Bhumichitr K. Social-Aware Automated Course Planner: An Integrated Recommender System for University Registration System, pp. 545–548, 2017.
3. Wei C, Duan K, Zhuo S, Wang H, Huang S, Liu J. Enhanced Recommendation Systems with Retrieval-Augmented Large Language Model. J Artif Intell Res. Feb. 2025;82:1147–73. 10.1613/jair.1.17809.
4. Zhang JC, Zain AM, Zhou KQ, Chen X, Zhang RM. A review of recommender systems based on knowledge graph embedding. Expert Syst Appl. Sep. 2024;250:123876. 10.1016/J.ESWA.2024.123876.
5. Wang S et al. Knowledge Graph Retrieval-Augmented Generation for LLM-based Recommendation, May 2025, [Online]. Available: http://arxiv.org/abs/2501.02226
6. Baig D, Nurbakova D, Mbaye B, Calabretto S. Knowledge Graph-Based Recommendation System for Personalized E-Learning, in *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*, New York, NY, USA: ACM, Jun. 2024, pp. 561–566. 10.1145/3631700.3665229
7. Choi W-C, Lam C-T, Pang PC-I, Mendes AJ. A Systematic Literature Review of Explainable Artificial Intelligence (XAI) for Interpreting Student Performance Prediction in Computer Science and STEM Education, in *Proceedings of the 30th ACM Conference on Innovation and Technology in Computer Science Education V. 1*, New York, NY, USA: ACM, Jun. 2025, pp. 221–227. 10.1145/3724363.3729027
8. Wang S et al. Knowledge Graph Retrieval-Augmented Generation for LLM-based Recommendation, May 2025, [Online]. Available: http://arxiv.org/abs/2501.02226
9. Huang J et al. A Comprehensive Survey on Retrieval Methods in Recommender Systems. Jul. 2024.
10. Shi D, Wang T, Xing H, Xu H. A learning path recommendation model based on a multidimensional knowledge graph framework for e-learning. Knowl Based Syst. May 2020;195:105618. 10.1016/j.knosys.2020.105618.
11. Song W, Duan Z, Yang Z, Zhu H, Zhang M, Tang J. Ekar: An Explainable Method for Knowledge Aware Recommendation, Jun. 2019, Accessed: Oct. 12, 2025. [Online]. Available: http://arxiv.org/abs/1906.09506
12. Meng Z, Yi Z, Ounis I. KERAG_R: Knowledge-Enhanced Retrieval-Augmented Generation for Recommendation, Jul. 2025.
13. Rong Z, Yuan L, Yang L. Enhanced knowledge graph recommendation algorithm based on multi-level contrastive learning. Sci Rep. Oct. 2024;14(1):23051. 10.1038/s41598-024-74516-z.
14. Kwon J, Ahn S, Seo Y-D. RecKG: Knowledge Graph for Recommender Systems, in *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*, New York, NY, USA: ACM, Apr. 2024, pp. 600–607. 10.1145/3605098.3636009

15. Li Z, Wang Z, Wang W, Hung K, Xie H, Wang FL. Retrieval-augmented generation for educational application: A systematic survey. Computers Education: Artif Intell. Jun. 2025;8:100417. 10.1016/j.caeai.2025.100417.
16. Ye Y. An overview of knowledge graph-based recommendation systems. Appl Comput Eng. Jul. 2024;73(1):57–68. 10.54254/2755-2721/73/20240363.
17. Johora FT, Hasan MN, Rajbongshi A, Ashrafuzzaman M, Akter F. An explainable AI-based approach for predicting undergraduate students academic performance. Array. Jul. 2025;26:100384. 10.1016/j.array.2025.100384.
18. Hooda M, Rana C, Dahiya O, Rizwan A, Hossain MS. Artificial Intelligence for Assessment and Feedback to Enhance Student Success in Higher Education, *Math. Probl. Eng.*, vol. 2022, pp. 1–19, May 2022. 10.1155/2022/5215722
19. Islam MM, Sojib FH, Mihad MFH, Hasan M, Rahman M. The integration of explainable AI in Educational Data Mining for student academic performance prediction and support system. Telematics Inf Rep. Jun. 2025;18:100203. 10.1016/j.teler.2025.100203.
20. Li J, Zhao Y, Gao X, Zhao W. Feature Extraction and SVM Parameters Optimization Based on ICQPSO, in *Proceedings – 2020 International Conference on Computer Network, Electronic and Automation, ICCNEA 2020*, Institute of Electrical and Electronics Engineers Inc., Sep. 2020, pp. 329–332. 10.1109/ICCNEA50255.2020.00074
21. Li J, Lyu Q, Qiu W, Khong AWH. Improving Course Recommendation Systems with Explainable AI: LLM-Based Frameworks and Evaluations, 2025, 10.5281/zenodo.15870185
22. Swathi K, Yuvatez R, Kovid S. Enhancing Trust in AI-Powered Learning: The Role of Explainable AI in the Education Field.
23. zhao N, Long Z, Zhao Z-D, Wang J. URIR: Recommendation algorithm of user RNN encoder and item encoder based on knowledge graph, Nov. 2021, Accessed: Oct. 12, 2025. [Online]. Available: http://arxiv.org/abs/2111.00739
24. Yang Y, Peng X, Chen M, Liu S. An explainable graph-based course recommendation model based on multiple interest factors. Expert Syst Appl. Mar. 2025;264:125889. 10.1016/j.eswa.2024.125889.

Figures

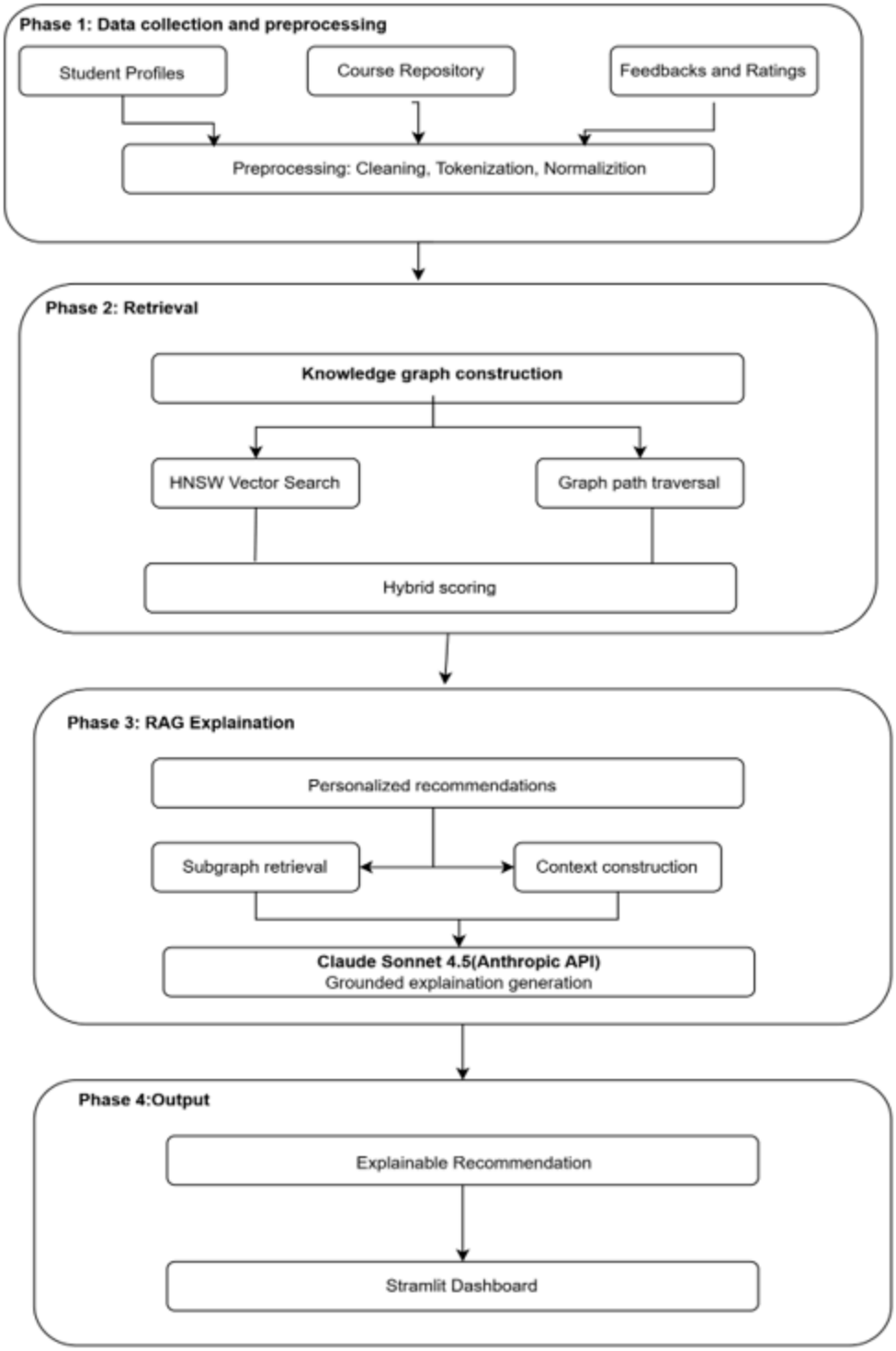


Figure 1

Architecture of KG-HCRS

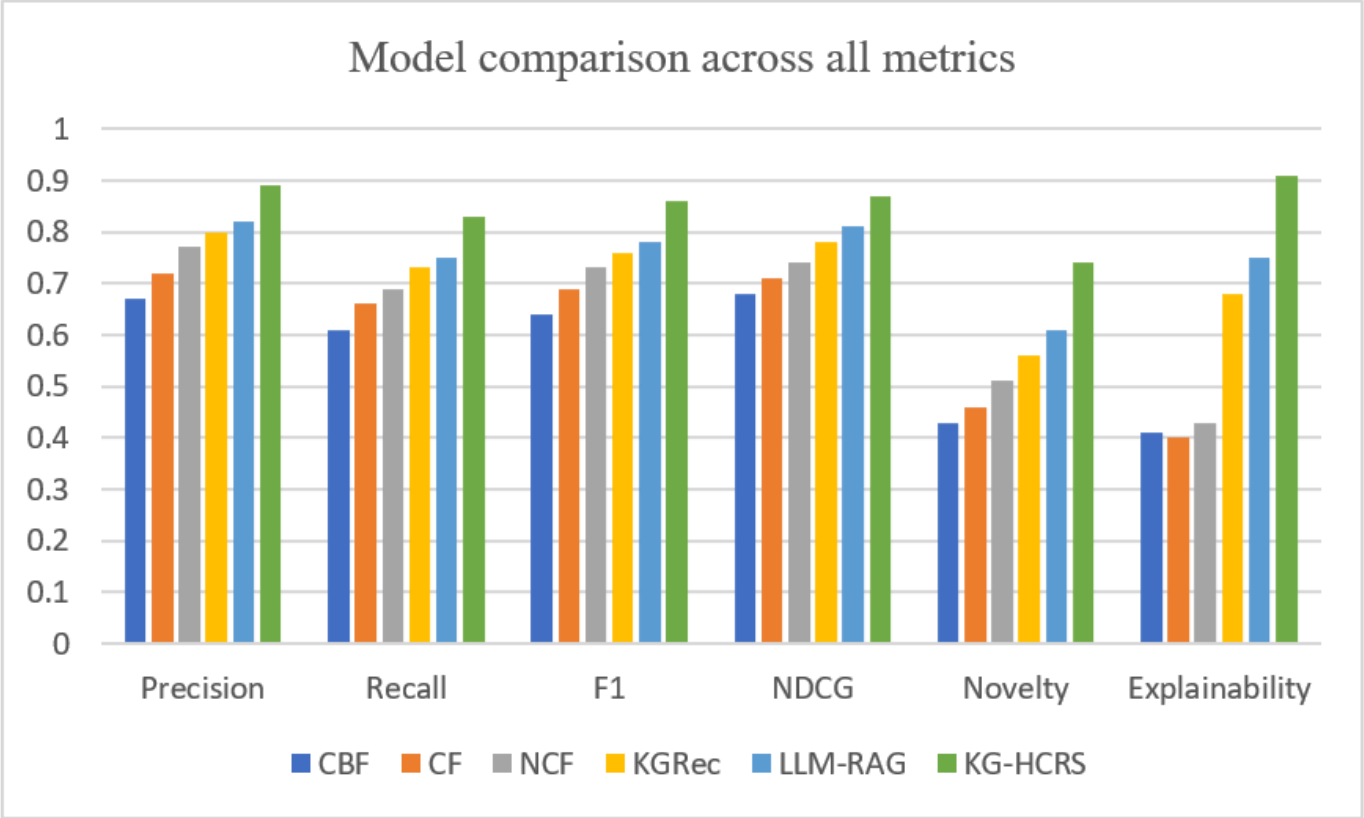


Figure 2

Evaluation results across all models



Figure 3

Sensitivity analysis of the hybrid parameter λ

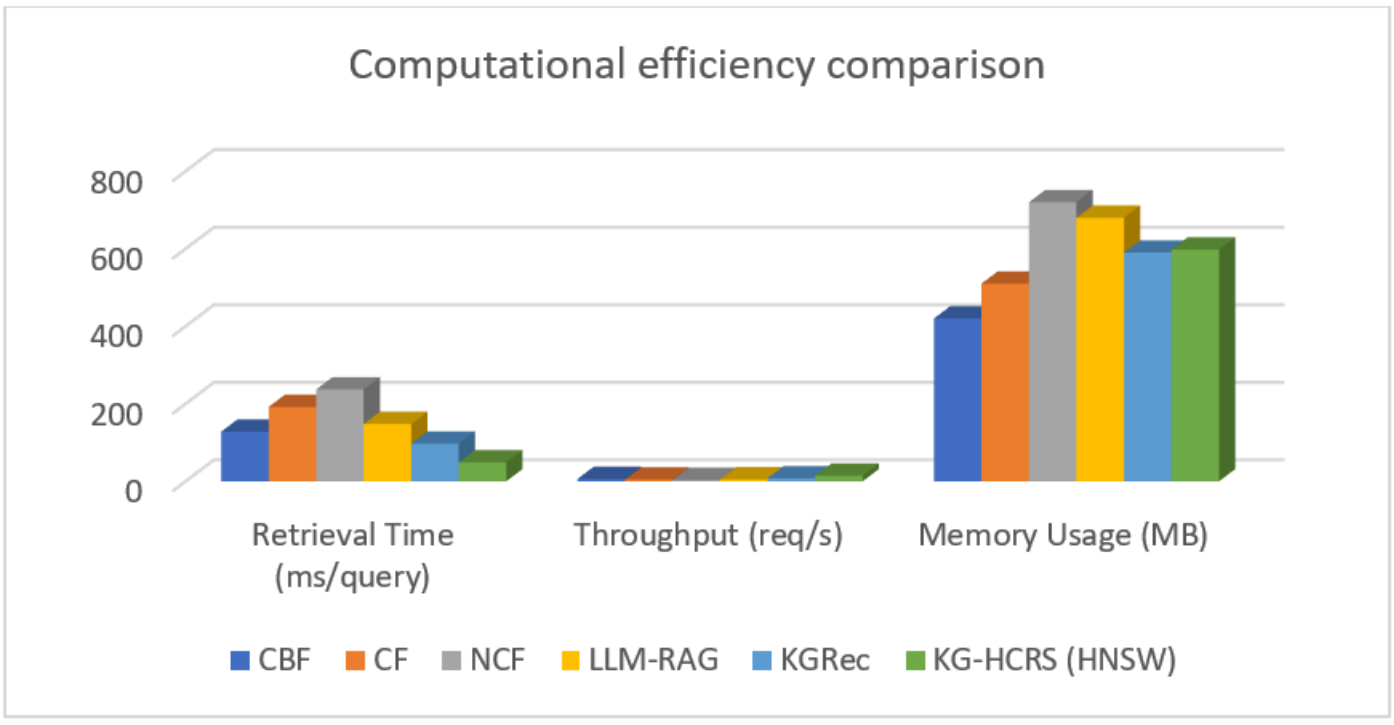


Figure 4

Computational efficiency comparison