technologies

MDPI

Article

Semantic–Sequential Educational Recommendation with Collaborative Enhancement and Parameter-Efficient Language Model Adaptation

Hajar Majjate [1,*], Youssra Bellarhmouch [1], Adil Jeghal [2], Ali Yahyaouy [1], Loubna Laaouina [2], Hamid Tairi [1] and Khalid Alaoui Zidani [1]

[1] L3IA Laboratory, Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez 30003, Morocco
[2] LSATE Laboratory, National School of Applied Science, Sidi Mohamed Ben Abdellah University, Fez 30000, Morocco
* Correspondence: hajar.majjate@usmba.ac.ma

Abstract

The rapid evolution of online learning environments has generated diverse and complex data ecosystems. Recommender systems play a central role in leveraging such heterogeneous data to support personalised learning experiences. However, many deep learning-based recommender systems still rely on identifier-based representations that capture co-occurrence and collaborative patterns while overlooking the semantic information embedded in educational activities and the temporal dynamics of learner behaviour. To address these limitations, this study proposes a collaborative-enhanced semantic–sequential recommendation framework for educational platforms that combines structured semantic representation learning, sequential behavioural modelling, and collaborative preference modelling. The proposed architecture integrates a parameter-efficient MiniLM adaptation strategy to extract semantic representations from structured item-related educational metadata and a bidirectional recurrent encoder to model temporal learning patterns from behavioural logs. A gated fusion mechanism is then used to combine semantic and contextual information into learner representations, which are further integrated with collaborative user–item embeddings for top-K recommendation using a Bayesian personalised ranking objective. Experiments conducted on the EdNet-KT1 dataset under chronological splitting, full-corpus ranking, and fixed candidate-sampling protocols show that the collaborative-enhanced model achieves the highest-ranking performance among popularity-based, matrix factorisation, neural collaborative filtering, recurrent sequential, self-attention sequential, and ablation baselines. The model obtains an NDCG@10 of 0.1344 under full-corpus ranking and 0.5383 under candidate sampling, with statistically significant but practically modest improvements over the strongest baselines. Additional ablation, efficiency, and gate analyses indicate that semantic–contextual modelling is most effective when used as a residual enhancement to collaborative recommendation rather than as a standalone replacement. These results suggest that parameter-efficient semantic–sequential modelling, when combined with collaborative preference signals, offers a promising direction for scalable and evidence-based educational recommender systems.

Keywords: educational recommender systems; semantic representation learning; sequential recommendation; collaborative filtering; parameter-efficient adaptation; learning analytics

Check for updates

Academic Editors: Marta Montenegro-Rueda, Jianxiong Zhu and José Fernández-Cerero

Received: 19 April 2026
Revised: 1 June 2026
Accepted: 4 June 2026
Published: 6 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Every day, millions of learners engage with digital learning environments, including massive open online courses, adaptive tutoring systems, and mobile applications. These interactions generate heterogeneous data, including behavioural logs [1], temporal patterns [2], contextual information [3], and cognitive traces [4]. Leveraging such data through recommender systems enables the delivery of personalised learning resources and adaptive learning pathways, thereby enhancing learning experiences and instructional decision-making [5]. Yet, despite substantial progress in educational data mining and learning analytics, modelling learner needs remains challenging because learning behaviour is dynamic, sequential, and influenced by both contextual and content-related factors [6]. In educational settings, conventional recommendation approaches such as collaborative filtering, matrix factorisation, and neural collaborative filtering, often depend on static representations of learner behaviour and educational content [7]. These methods typically use ID-based representations, where each learner or resource is encoded as a unique identifier associated with a latent embedding [8,9]. In these approaches, recommendations are derived mainly from learner–resource interaction histories. Although this paradigm is efficient and scalable, it does not naturally incorporate user attributes, session context, evolving interests, or semantic information unless these signals are explicitly modelled [6,7]. This limits their ability to capture the rich contextual information that shapes learners' changing needs across diverse real-world situations and fast-evolving learning environments.

To address these limitations, context-aware recommender systems (CARS) have been developed to incorporate additional situational or behavioural factors into the recommendation process [10]. Unlike traditional static systems, CARS integrate relevant contextual factors such as temporal learning patterns, session-level behaviour, prior responses, content structure, and learner progress indicators. These signals enable better alignment of learning modalities with the specific needs of the situation, ultimately enhancing the quality of recommendations and supporting learner motivation [11]. However, the implementation of CARS can be complex due to the need for advanced algorithms and infrastructure [12], which may result in higher computational costs and increased data requirements [12,13]. Additionally, incorporating contextual signals into recommendation systems remains a challenging task, especially with unstructured, noisy and multimodal data such as that found in complex learning environments. Moreover, data quality [14], including the expressiveness of features representing learners, items and their contexts strongly influence the effectiveness of the recommendation process.

In this regard, recent developments in large language models (LLMs), such as GPT and BERT, have shown strong potential for extracting semantic information from text associated with learner interactions [15]. Their main advantage lies in their ability to generate rich contextual representations from unstructured textual inputs [16]. This makes them particularly promising for addressing challenges in context-aware recommender systems, especially in educational environments, such as conducting in-depth analysis of learner interactions, managing ambiguous or sparse data, and adapting to evolving, complex learning environments. Specifically, LLMs can enrich learner–resource representations [17], by capturing semantic relationships among educational concepts, item descriptions, learner responses, interaction histories, and contextual traces that traditional ID-based models may not explicitly represent.

Despite their potential, the use of LLMs as semantic feature extractors in hybrid, context-aware educational recommender systems remains insufficiently explored [18]. Existing approaches rarely exploit LLMs at the feature extraction stage, where shallow encodings and rigid preprocessing pipelines may lead to the loss of important semantic and contextual information [19,20]. Furthermore, fully fine-tuning of large models is

computationally expensive, which restricts their deployment in practical and resource-constrained educational settings. These limitations highlight the need for efficient adaptation strategies that preserve the expressive power of LLMs while enabling scalable personalised recommendation.

Motivated by these considerations, this study proposes a collaborative-enhanced semantic–sequential recommendation framework for educational platforms. The framework builds on a semantic–contextual backbone that combines parameter-efficient language model adaptation with behavioural sequence modelling. A MiniLM-based semantic encoder is adapted using Low-Rank Adaptation to represent structured educational metadata, while a bidirectional recurrent encoder models temporal behavioural patterns from learner interaction logs. A gated fusion mechanism combines semantic and contextual representations into a unified semantic–contextual learner representation. To improve competitiveness against strong collaborative baselines, the architecture further integrates a collaborative enhancement branch initialised from matrix-factorisation embeddings. The final recommendation score combines the semantic–contextual representation with collaborative user–item preference signals.

The central hypothesis of this study is that integrating semantic item-related information and behavioural sequence context with collaborative user–item preference modelling improves educational recommendation quality compared with purely collaborative, purely sequential, and standalone semantic–contextual alternatives. More specifically, the study evaluates the following hypotheses:

H1. *A collaborative-enhanced semantic–sequential model improves top-K ranking performance over strong collaborative and sequential recommendation baselines.*

H2. *Adding collaborative enhancement improves the standalone semantic–contextual backbone.*

H3. *The semantic–contextual branch contributes incremental ranking gains, but its effect is expected to be modest under strict full-corpus evaluation because collaborative preference signals remain highly competitive in large-scale implicit-feedback recommendation.*

Accordingly, the main contributions of this study are as follows:

- A collaborative-enhanced semantic–sequential recommendation framework: We propose an educational recommendation architecture that combines structured semantic representation learning, behavioural sequence modelling, gated semantic–contextual fusion, and collaborative user–item preference modelling.
- Parameter-efficient semantic adaptation for structured educational metadata: We employ a Low-Rank Adaptation strategy to adapt a MiniLM-based semantic encoder to structured item-related educational metadata without full model fine-tuning.
- A gated semantic–contextual learner representation: We integrate semantic and behavioural representations through a learnable gated fusion mechanism that provides a controlled view of the relative contribution of the two internal information sources.
- Collaborative enhancement for competitive top-K ranking: We extend the semantic–contextual backbone with a collaborative branch initialised from matrix-factorisation embeddings, allowing the model to preserve strong collaborative preference structure while adding semantic–contextual information.
- We evaluate the revised framework on EdNet-KT1 using chronological splitting, full-corpus ranking, fixed candidate-sampling evaluation, strong collaborative and sequential baselines, ablation variants, paired statistical tests, effect-size analysis, and computational efficiency measurements.
- A balanced empirical interpretation: The results show that the collaborative-enhanced model achieves the best ranking performance under both evaluation protocols.

However, the improvements over the strongest collaborative baselines are statistically significant but practically modest, indicating that semantic–contextual modelling is most useful as a residual enhancement to collaborative recommendation rather than as a standalone replacement.

The remainder of this paper is organised as follows. Section 2 reviews related work on educational recommender systems, sequential recommendation, language-model-based representation learning, and parameter-efficient adaptation. Section 3 presents the proposed collaborative-enhanced semantic–sequential recommendation framework. Section 4 describes the dataset, preprocessing strategy, experimental protocol, baselines, evaluation metrics, and implementation details. Section 5 reports the experimental results. Section 6 discusses the findings, limitations, and future research directions. Section 7 concludes the paper.

2. Background and Related Work

2.1. Context-Aware Recommender Systems

Conventional recommender systems have predominantly concentrated on modelling general user preferences. In these systems, user–item interactions are typically organised in a two-dimensional structure known as the user–item interaction matrix or utility matrix [21–23]. The appeal of this representation lies in its mathematical simplicity, its compatibility with fundamental algorithms such as collaborative filtering and matrix factorization, and its effectiveness in handling sparse data, where most entries remain unobserved [23]. Despite the prevalence of such approaches, recommendations based on two-dimensional structures can become ineffective when contextual circumstances change. For instance, in educational environments, suggesting group discussions or providing extensive materials during brief mobile commutes or when students move from regular study routines to travelling with limited internet access can be counterproductive and lead to less favourable outcomes. This observation highlights the fact that user preferences are not static; they fluctuate in response to various situational factors. As a result, the concept of context-aware recommender systems (CARS) has emerged to incorporate elements such as time, weather, device type, social and environmental factors [10]. Early research introducing CARS emerged in the early 2000s, recognising that user preferences are not static and are significantly influenced by situational factors and prompting the need to move beyond implicit or explicit user–item interactions alone [24]. Pioneering works, notably by Adomavicius and Tuzhilin [10,24,25] formalised the integration of context into recommendation processes by extending the traditional 2D ($U \times I$) matrix to a multi-dimensional ($U \times I \times C$) model and proposing initial architectural paradigms such as pre-filtering, post-filtering, and contextual modelling.

Initially, context-aware recommendation systems (CARS) incorporated explicit and limited contextual attributes such as the time of day and social context. However, advances in sensing technologies have significantly influenced their evolution [12]. The widespread adoption of smart mobile devices, Internet of Things (IoT) sensors, and location-based services has transformed how CARS operate, especially in their ability to gather and analyse user data [12,26]. This enabled a more integrated and nuanced understanding of the context in which recommendations are made, allowing for the gathering of increasingly diverse and even implicit contextual data.

The wealth and diversity of available data have led to the development of more sophisticated modelling techniques, including tensor factorization for handling multi-dimensional data [27], fuzzy sets for managing imprecise contextual information [28], and evolutionary computing for optimising contextual feature selection and model parameters [26].

Recently, the trajectory of CARS development accelerated further with the rise in deep learning in the 2010s. Numerous works have employed deep learning architectures, such as Recurrent Neural Networks (RNNs) for sequential context [29] and Convolutional Neural Networks (CNNs) for structured context [30], providing unprecedented capabilities for implicitly learning rich and complex contextual representations directly from raw, high-dimensional data. This marked a significant departure from reliance on manually engineered features, allowing models to discover subtle contextual cues that might otherwise be missed.

In education, context-aware recommender systems are addressing challenges such as data sparsity [31] in high-dimensional learner–context interactions. Growing attention is also being paid to ethical and privacy concerns arising from the collection and use of sensitive learner data, including activity logs, performance records, and behavioural traces [32]. Furthermore, recent works highlight a rising interest in leveraging the representational power of Large Language Models (LLMs) and other advanced AI paradigms to enhance the adaptability [17], transparency, and natural language understanding of next-generation recommender systems [18].

2.2. Collaborative and Sequential Recommendation Models

Collaborative filtering and matrix factorisation constitute foundational paradigms in recommender systems, principally for their capacity to infer latent user–item preference structures from sparse interaction data [21,23]. Matrix factorisation projects users and items into a shared latent space and quantifies relevance through their interaction patterns, offering both computational efficiency and scalability in implicit-feedback scenarios [23]. Neural collaborative filtering advances this paradigm by replacing or augmenting linear interaction functions with non-linear neural architectures, thereby enabling richer characterisation of user–item relationships [33].

In parallel, sequential recommendation models have emerged to explicitly model the temporal nature of user interactions. Recurrent architectures such as GRU4Rec represent interaction histories as ordered sequences and learn evolving preference states [29], while self-attention-based models such as SASRec capture longer-range dependencies through attention mechanisms [34]. These approaches are particularly relevant in educational contexts, where learner behaviour develops progressively and where the sequencing of learning activities encodes shifts in learner needs, conceptual mastery, and intrinsic motivation.

Despite their demonstrated effectiveness, collaborative and sequential models mainly rely on item identifiers and observable interaction signals, thereby underutilising semantic and structural information embedded within educational resources [23,29,33,34]. Specifically, these approaches may overlook concept taxonomies, instructional metadata, answer-response data, and hierarchical content structures, which are important for modelling pedagogically meaningful relationships among learning resources [17,19,20]. This limitation underscores the need for hybrid frameworks that integrate collaborative preference signals with rich contextual and semantic information.

2.3. Large Language Models for Semantic Representation Learning

Large language models (LLMs) have significantly advanced natural language processing through self-attention mechanisms and large-scale pre-training [35]. Their ability to generate rich semantic representations has made them effective for tasks such as question answering [36], information retrieval [37], and summarisation [38]. Recent studies have also highlighted their value for representation learning [9,39], particularly in applications that require expressive semantic embeddings for improved matching and personalisation.

In educational settings, these capabilities are relevant because learner interactions often contain textual information that can support richer modelling than traditional ID-based approaches alone. By extracting semantic information from text associated with learner interactions, LLMs can enhance the expressiveness of learner and item representations and support more adaptive recommendation [11,17]. Recent work has begun to explore the integration of LLMs into educational recommender systems. For example, Retrieval-Augmented Generation (RAG)-based approaches have improved course recommendation by combining external knowledge retrieval with generative reasoning [37,38], while TutorLLM combines LLMs with knowledge-tracing methods to personalise learning trajectories [40]. These studies highlight the promise of LLMs for educational recommendation, but important challenges remain. Many pretrained LLMs are not directly adapted to educational data and may struggle to capture pedagogically relevant semantic patterns [17,41].

In addition, traditional knowledge-tracing models such as Bayesian Knowledge Tracing (BKT) and Deep Knowledge Tracing (DKT) effectively model temporal learning dynamics but do not capture rich semantic information. Conversely, LLMs provide strong semantic representations but do not explicitly model sequential behavioural patterns in learning processes [42,43]. Although recent studies have explored hybrid architectures that combine language models with sequential recommendation frameworks, these approaches are mainly developed in non-educational domains and focus on general item recommendation rather than pedagogically grounded learner modelling. From a practical perspective, the high computational cost of full fine-tuning also limits the deployment of LLMs in resource-constrained educational settings [44]. As a result, there remains a need for efficient approaches that integrate semantic representation learning, behavioural sequence modelling, and collaborative preference signals for scalable educational recommendation.

2.4. Parameter-Efficient Language Model Adaptation

Fine-tuning large language models (LLMs) for specialised applications traditionally involves updating all model parameters [45]. While this full fine-tuning approach ensures high adaptability, it demands substantial computational resources, extended training time, and large storage capacity. Such requirements are impractical in most educational settings, where institutions often operate under hardware and budget constraints. Moreover, maintaining multiple fully fine-tuned models for different subjects or learner populations further increases operational costs and limits scalability in real-world learning management systems.

Parameter-efficient fine-tuning (PEFT) techniques have been proposed to mitigate these challenges. Among them, Low-Rank Adaptation (LoRA) [45,46] has gained particular attention for its simplicity and effectiveness. LoRA injects small, trainable low-rank matrices into the transformer's weight layers while keeping the pre-trained weights frozen. This drastically reduces the number of trainable parameters and GPU memory usage, enabling efficient domain adaptation without sacrificing model performance [47]. Studies have shown that LoRA achieves performance comparable to full fine-tuning across diverse domains, including natural language understanding [48], medical text processing [49], and software engineering tasks [50].

Despite its success across multiple domains, the application of LoRA in educational settings remains limited. Existing educational studies have largely relied on prompting or few-shot learning, with less attention to how parameter-efficient fine-tuning can adapt large language models to pedagogically relevant data. In particular, the use of LoRA in context-aware educational recommender systems remains underexplored, especially for settings that require the joint modelling of semantic information and evolving learner

behaviour. This gap motivates the investigation of parameter-efficient language model adaptation as one component of a broader semantic–sequential and collaborative-enhanced educational recommendation framework.

2.5. Research Gap

Despite substantial progress in recommender systems and the growing use of language models in education, several research gaps remain. First, many language models are pretrained on general-purpose corpora and are not directly adapted to the semantic characteristics of educational data. Second, collaborative and sequential recommender models are effective in capturing user–item preferences and temporal interaction patterns, but they often underuse semantic and structural information embedded in educational resources. Conversely, language-model-based approaches can enrich semantic representations but do not, by themselves, explicitly model behavioural sequences or collaborative preference structures. Third, full fine-tuning of large language models remains computationally expensive for resource-constrained educational settings, which limits practical deployment.

Although parameter-efficient adaptation techniques such as Low-Rank Adaptation provide a practical way to adapt language models with reduced computational cost, their integration into educational recommender systems remains insufficiently explored, particularly in ranking-oriented settings that require the joint modelling of semantic information, behavioural sequence context, and collaborative preference signals. Motivated by this gap, the present study proposes a collaborative-enhanced semantic–sequential framework that combines parameter-efficient semantic representation learning, behavioural sequence modelling, gated semantic–contextual fusion, and collaborative preference modelling for top-K educational recommendation.

3. Materials and Methods

3.1. Framework Overview

The proposed framework is designed to improve educational recommendation by jointly modelling three complementary sources of information: the semantic structure of learning resources, the temporal dynamics of learner behaviour, and the collaborative preference patterns emerging from learner–item interactions. As illustrated in Figure 1, the architecture is organised around a semantic–contextual backbone complemented by a collaborative enhancement branch. The semantic stream encodes structured educational text sequences derived from item metadata, concept tags, learner responses, correct-answer information, and lecture-related descriptors. In parallel, the behavioural stream models temporal interaction logs, including response correctness, elapsed time, interaction order, and time intervals between learning activities. These two streams are integrated through a gated fusion mechanism to produce a semantic–contextual learner representation, which is then combined with collaborative user–item preference signals for top-K recommendation.

The gating mechanism plays a central role in the semantic–contextual backbone by regulating the internal contribution of semantic and behavioural representations before ranking. Rather than treating both information sources as fixed or equally weighted, the gate learns a modality-level fusion pattern that reflects their relative contribution within the semantic–contextual branch. To preserve the strong preference structure captured by collaborative recommendation models, the architecture further includes a collaborative enhancement branch initialised from matrix-factorisation embeddings. The final recommendation layer combines the semantic–contextual score and the collaborative score into a hybrid ranking function optimised with Bayesian personalised ranking. In the remainder of the paper, the semantic–contextual backbone is referred to as CGSR-L, whereas the complete collaborative-enhanced architecture is denoted as CGSR-L+CF.

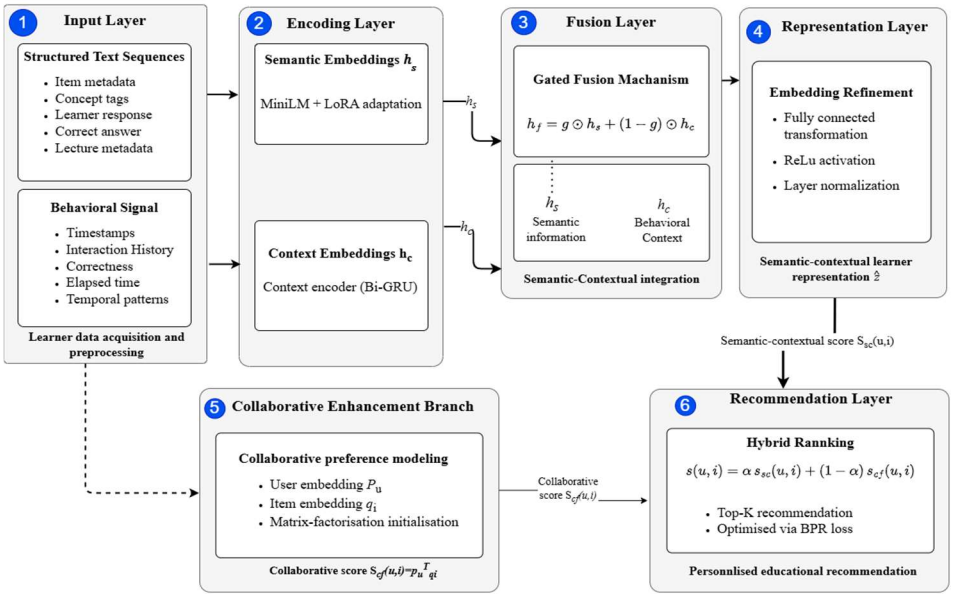


Figure 1. Overview of the proposed CGSR-L+CF architecture.

The overall architecture consists of six main components:

1. **Input Layer**: This layer ingests structured educational text sequences and behavioural logs. The structured text sequences are constructed from item-related metadata, concept tags, learner responses, correct-answer information, and lecture-related metadata. Behavioural logs include timestamps, correctness, elapsed time, and temporal interaction features. The data undergo text normalisation, tokenisation, noise filtering, feature scaling, and temporal alignment.
2. **Encoding Layer**: Two complementary encoders are used. A LoRA-adapted MiniLM encoder transforms structured educational text sequences into dense semantic embeddings. In parallel, a bidirectional GRU encoder processes sequential behavioural data to model temporal dependencies and evolving learning patterns.
3. **Fusion Layer**: The semantic and contextual embeddings are combined using a gated fusion mechanism. This mechanism adaptively balances the contribution of each modality based on their relative relevance, allowing the model to rely more on behavioural context when textual input is limited and more on semantic cues when behavioural history is sparse.
4. **Representation Layer**: The fused embedding is refined through a fully connected projection, non-linear activation, and layer normalisation, producing a compact semantic–contextual learner representation optimised for ranking tasks.
5. **Collaborative Enhancement Branch**: Collaborative learner and item embeddings preserve user–item preference structure and are initialised from the matrix-factorisation baseline. This branch complements the semantic–contextual representation with collaborative signals.
6. **Recommendation Layer**: The final ranking score combines the semantic–contextual score and the collaborative score and is optimised using a Bayesian Personalised Ranking objective.

Overall, this layered design enables an end-to-end transformation from educational interaction data to ranked learning-resource recommendations. The semantic adaptation remains parameter-efficient through LoRA, while the collaborative enhancement branch improves ranking competitiveness by preserving collaborative preference structure.

3.2. Input Layer and Preprocessing

The input layer processes raw learner interaction data collected from educational platforms. These data consist primarily of two components:

- **Structured educational text sequences (T)**: Each text sequence is constructed from educationally meaningful interaction information, including item identifiers, concept tags, module information, learner responses, correct-answer information, and lecture-related metadata. Each text sequence $t_i \in T$, undergoes standard preprocessing, including lowercasing, tokenisation, and truncation to a fixed maximum length $\boldsymbol{L_{max}}$, and is converted into subword embeddings compatible with the MiniLM tokenizer.
- **Contextual behavioural inputs (C):** These inputs consist of behavioural and temporal information, such as task order, response correctness, elapsed time, and time intervals between interactions. Each learner interaction history is represented as:

$$C = \{c_1, c_2, ..., c_T\}, c_t \in \mathbb{R}^{d_c} \quad (1)$$

where d_c is the number of contextual features and is $\boldsymbol{T}$ the sequence length. Continuous features are normalised, and categorical variables are embedded into dense vectors.

After preprocessing, the textual and contextual streams are temporally aligned using activity timestamps, ensuring that text-associated records and behavioural interactions correspond to the same learning events.

3.3. Encoding Layer

The encoding layer extracts semantic and contextual features from preprocessed learner data. It consists of two complementary modules:

- A semantic encoder based on a LoRA-adapted MiniLM model for semantic representation learning.
- A context encoder based on a Bidirectional GRU (Bi-GRU) for modelling behavioural and temporal dependencies.

Together, these encoders capture both structured semantic information associated with educational interactions and the temporal dynamics of learner behaviour.

3.3.1. Semantic Encoder (MiniLM-LoRA)

Structured educational text sequences are encoded using a LoRA-adapted MiniLM model, enabling parameter-efficient adaptation to educational data. LoRA adapts a pretrained language model by learning a small set of low-rank matrices that modify selected attention projections while keeping the original model parameters frozen. This allows efficient domain adaptation with limited computational overhead.

Given a tokenised text sequence:

$$x = \{x_1, x_2, \ldots, x_n\} \quad (2)$$

The semantic embedding is computed as:

$$h_s = f_{\text{MiniLM}-\text{LoRA}}(x) \in \mathbb{R}^{d_s} \quad (3)$$

where $d_s = 384$ corresponds to the hidden size of the MiniLM-L6-v2 model. MiniLM generates contextualised token representations using multi-head self-attention, and the hidden states from the final transformer layer are aggregated via mean pooling across the token dimension to obtain a compact interaction-level representation.

LoRA-based adaptation modifies selected weight matrices as:

$$W' = W + BA \tag{4}$$

with low-rank matrices $A \in \mathbb{R}^{r \times d_{in}}$ and $B \in \mathbb{R}^{d_{out} \times r}$, where $r \ll d_{in}, d_{out}$. This enables efficient domain adaptation without updating the whole model.

3.3.2. Context Encoder

The context encoder models the temporal and behavioural dynamics underlying learner interactions. While the semantic encoder captures structured educational semantic information, the context encoder captures how the learner behaves over time through actions, responses, and performance-related signals.

Let each learner's interaction sequence be represented as:

$$C = \{c_1, c_2, ..., c_T\}, c_t \in \mathbb{R}^{d_c} \tag{5}$$

where each vector c_t encodes contextual features such as task correctness, response time, and time interval between tasks, with interaction order implicitly encoded by the sequential structure. These features are normalised and temporally aligned with the corresponding text-associated events processed by the semantic encoder.

To capture sequential dependencies, a Bidirectional Gated Recurrent Unit network is employed. The Bi-GRU processes the sequence in both forward and backward directions to capture temporal dependencies within the learner activity stream:

$$\overrightarrow{h_t} = \mathrm{GRU}_f(c_t, \overrightarrow{h_{t-1}}), \overleftarrow{h_t} = \mathrm{GRU}_b(c_t, \overleftarrow{h_{t+1}}) \tag{6}$$

The hidden representations are concatenated at each time step to form:

$$h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}] \tag{7}$$

To obtain a single behavioural embedding for the learner, temporal pooling is applied:

$$h_c = \frac{1}{T}\sum_{t=1}^{T} h_t \in \mathbb{R}^{d'_c} \tag{8}$$

We chose Bi-GRU due to its efficiency and strong performance in modelling short-to-medium sequences, which aligns with the structure of the EdNet dataset, where most learner sessions consist of tens to hundreds of actions rather than very long sequences. Compared with traditional recurrent networks, Bi-GRUs mitigate vanishing gradients while maintaining a lower computational cost than LSTMs or Transformers. Bidirectionality enables the encoder to integrate both prior and subsequent behavioural cues in the offline representation-learning setting used in this study.

3.4. Fusion Layer

The fusion layer integrates semantic and contextual representations into a unified learner embedding, enabling the combination of semantic information from text associated with learner interactions and contextual patterns derived from sequential logs prior to recommendation. The semantic embedding produced by the LoRA-adapted language model is denoted as $h_s \in \mathbb{R}^{d_s}$, and the contextual embedding derived from the Bi-GRU encoder is denoted as $h_c \in \mathbb{R}^{d'_c}$.

To align the two feature spaces, both embeddings are projected into a shared latent dimension d_f using linear transformations:

$$\tilde{h}_s = W_s h_s + b_s \,, \quad \tilde{h}_c = W_c h_c + b_c \tag{9}$$

where $W_s \in \mathbb{R}^{d_f \times d_s}$, $W_c \in \mathbb{R}^{d_f \times d'_c}$ and $b_s, b_c \in \mathbb{R}^{d_f}$ are learnable parameters.

Gated Fusion Mechanism

To regulate the relative contribution of semantic and contextual information within the semantic–contextual branch, a gated fusion mechanism is applied. A gating vector $g \in \left[0,1\right]^{d_f}$ is computed as:

$$g = \sigma\left(W_g \left[\tilde{h}_s; \tilde{h}_c\right] + b_g\right) \tag{10}$$

where $[\cdot;\cdot]$ denotes vector concatenation, $W_g \in \mathbb{R}^{d_f \times 2d_f}$, $b_g \in \mathbb{R}^{d_f}$ and $\sigma(\cdot)$ is the sigmoid activation function. The final fused embedding is obtained as:

$$h_f = g \odot \tilde{h}_s + (1 - g) \odot \tilde{h}_c \tag{11}$$

where $\odot$ denotes element-wise multiplication.

This mechanism provides an internal weighting between semantic and behavioural representations. The fused representation $h_f \in \mathbb{R}^{d_f}$ therefore serves as a context-aware learner embedding that captures both semantic and behavioural information. This gate is interpreted as a fusion mechanism inside the semantic–contextual branch, not as the final contribution of this branch to the recommendation score.

3.5. Representation Layer

The representation layer refines the fused embedding h_f obtained from the gated fusion mechanism into a task-specific semantic–contextual learner representation suitable for personalised recommendation. This is achieved through a fully connected transformation followed by a non-linear activation and layer normalisation, which improves representational stability and task relevance.

The fused learner embedding $h_f \in \mathbb{R}^{d_f}$ is first transformed into a higher-level latent space:

$$z = \mathrm{ReLU}\left(W_r h_f + b_r\right) \tag{12}$$

where $W_r \in \mathbb{R}^{d_z \times d_f}$ and $b_r \in \mathbb{R}^{d_z}$ are trainable parameters, and d_z denotes the dimensionality of the final learner representation.

To ensure embedding stability and mitigate internal covariate shifts, layer normalisation is applied:

$$\hat{z} = \mathrm{LayerNorm}(z) \tag{13}$$

The resulting normalised vector z_hat serves as a compact semantic–contextual learner representation for downstream recommendation.

3.6. Collaborative Enhancement Branch

To complement the semantic–contextual learner representation with collaborative preference information, the proposed architecture incorporates a collaborative enhancement branch. This branch is designed to preserve the user–item preference structure captured from historical interactions while allowing the model to benefit from the semantic and behavioural information encoded by the semantic–contextual backbone.

Let $p_u \in \mathbb{R}^{d_z}$ denote the collaborative embedding of learner u, and let $q_i \in \mathbb{R}^{d_z}$ denote the collaborative embedding of item i. These embeddings are initialised from the matrix-factorisation model and further optimised during ranking training. The semantic–contextual branch computes the relevance between learner u and item i as follows:

$$s_{sc}(u,i) = \hat{z}_u^\top v_i \tag{14}$$

where $\hat{z}_u$ denotes the semantic–contextual learner representation obtained after gated fusion and representation refinement, and $v_i \in \mathbb{R}^{d_z}$ denotes the semantic–contextual item embedding.

In parallel, the collaborative branch estimates the relevance score using the collaborative learner and item embeddings:

$$s_{cf}(u,i) = p_u^\top q_i \tag{15}$$

The final relevance score is obtained by combining the semantic–contextual score and the collaborative score:

$$s(u,i) = \alpha s_{sc}(u,i) + (1-\alpha) s_{cf}(u,i) \tag{16}$$

where α is a learnable coefficient that controls the branch-level contribution of the semantic–contextual component relative to the collaborative component. This formulation enables the model to jointly exploit semantic–behavioural representations and collaborative preference signals within a unified ranking function.

The architecture separates modality-level fusion from branch-level scoring. The gated fusion mechanism operates within the semantic–contextual branch by combining semantic and behavioural representations, whereas α weights the semantic–contextual score against the collaborative score in the final relevance function. As a result, the model first constructs a fused semantic–behavioural learner representation and then integrates it with collaborative preference information at the scoring stage.

3.7. Recommendation Layer

The recommendation layer transforms the semantic–contextual and collaborative representations into ranked educational item recommendations. For each learner–item pair (ui), the model computes a hybrid relevance score $s(u,i)$ by combining the semantic–contextual score produced by the backbone and the collaborative score produced by the collaborative enhancement branch. Candidate items are then ranked according to this final score, where higher values indicate stronger predicted relevance for the learner.

During training, the ranking order is optimised using the Bayesian Personalised Ranking objective:

$$\mathcal{L}_{BPR} = -\sum_{(u,i,j)\in\mathcal{D}} \ln\sigma(s(u,i) - s(u,j)) + \lambda \parallel \Theta \parallel_2^2 \tag{17}$$

where (uij) denotes a training triplet in which learner u interacted with item i, while item j is treated as a non-observed or negative item. The function $\sigma(\cdot)$ denotes the sigmoid activation, Θ represents the trainable parameters of the model, and $\lambda \parallel \Theta \parallel_2^2$ is the regularisation term used to reduce overfitting. This pairwise ranking objective encourages observed or relevant items to receive higher scores than non-observed items, making it suitable for implicit-feedback educational recommendation.

By combining semantic–contextual relevance with collaborative preference structure, the recommendation layer produces top-K learning-resource recommendations that reflect

both the learner's behavioural-semantic profile and the collective interaction patterns captured from historical data.

The following section presents the experimental setup, including the dataset configuration, preprocessing strategy, model training procedures, evaluation metrics, and comparisons with popularity-based, matrix-factorisation, neural collaborative filtering, sequential recommendation, and ablation baselines.

4. Experimental Setup and Evaluation

The experimental setup evaluates the proposed framework under rigorous and reproducible recommendation conditions. Specifically, the experiments assess: (i) the ability of the model to exploit semantic and behavioural information for educational recommendation; (ii) the contribution of collaborative enhancement to ranking performance; (iii) the effect of the main architectural components through ablation analysis; (iv) robustness under both full-corpus and fixed candidate-sampling evaluation; and (v) computational efficiency in terms of trainable parameters, runtime, and memory consumption. All models were trained and evaluated under consistent chronological splitting and top-K ranking protocols to ensure fair comparison.

This section describes the dataset, preprocessing strategy, baseline models, ablation variants, implementation settings, evaluation metrics, statistical tests, and efficiency measurements used in the experiments.

4.1. Dataset and Preprocessing

The experiments were conducted on EdNet-KT1, a large-scale educational interaction dataset [51] collected from an intelligent tutoring platform. The dataset contains timestamped learner–item interaction records, including question identifiers, learner responses, correctness information, elapsed time, and interaction timestamps. In addition, item-related content metadata were used to construct structured educational text sequences for semantic representation learning.

Each learner's interaction history was ordered chronologically to preserve the temporal structure of learning behaviour. Learners with fewer than five valid interactions were removed to ensure a minimum behavioural history. For each remaining learner, the final interaction was reserved for testing, the preceding interaction was used for validation, and all earlier interactions were used for training. This chronological split prevents future interactions from being used when generating recommendations and therefore reflects a realistic next-item recommendation setting. To operationalise semantic representation learning, each learner–item interaction was transformed into a structured educational text sequence. These sequences were constructed from item-related metadata, including question identifiers, concept tags, module information, learner responses, correct-answer information, and lecture-related metadata. Textual sequences were normalised, tokenised using the MiniLM tokenizer, and temporally aligned with the corresponding behavioural records.

The behavioural stream included temporal and performance-related features such as correctness, elapsed time, time interval between consecutive interactions, item order, and module information. Continuous variables were normalised, and categorical variables were encoded before being passed to the sequential encoder. To control computational cost, interaction sequences were truncated or padded to a maximum length of 50.

Table 1 summarises the main statistics of the EdNet-KT1 subset used in our experiments after preprocessing. The final evaluation set contained 87,389 valid learners and 12,275 items, which were also used in the baseline- and candidate-sampling evaluation pipelines.

Table 1. Dataset statistics after preprocessing.

Statistic	Value
Learners before filtering	100,000
Valid learners after filtering	87,389
Total valid interactions	12,217,872
Training interactions	12,043,094
Validation interactions	87,389
Test interactions	87,389
Unique items	12,275
Average full sequence length	139.81
Maximum sequence length used by the model	50

4.2. Evaluation Protocol

In this study a strict chronological interaction-level evaluation protocol was adopted. For each learner, all interactions were sorted by timestamp. The final interaction was used as the test item, the preceding interaction was used for validation, and all earlier interactions were used for training. During evaluation, candidate items previously observed in the learner's training history were excluded from the ranked list. This prevents the model from recommending items already encountered during training.

Two complementary evaluation settings were used. First, full-corpus ranking evaluates each test interaction against the complete item vocabulary, after removing items previously observed by the learner during training. This setting is more demanding because the model must rank the relevant item among thousands of candidate resources.

Second, fixed candidate-sampling evaluation ranks the held-out positive item against 100 randomly sampled negative items. The candidate sets were generated once and kept identical across all models to ensure fair comparison. This protocol provides a controlled evaluation setting and allows direct comparison between collaborative, sequential, semantic–contextual, and hybrid models.

For both protocols, candidate items were ranked according to the predicted relevance score, and performance was computed at multiple top-K cut-offs.

4.3. Compared Models and Ablation Variants

The proposed model was compared with representative baselines from collaborative filtering, neural collaborative filtering, sequential recommendation, and internal ablation variants. The evaluated models are presented in Table 2.

Table 2. Evaluated models and baselines.

Popularity	A non-parametric baseline that ranks items according to their frequency in the training set.
BPR-MF	A matrix-factorisation model optimised with Bayesian Personalised Ranking.
NCF-GMF	A neural collaborative filtering baseline using a generalised matrix factorisation formulation. It provides a neural collaborative comparison against the proposed architecture.
GRU4Rec	A recurrent sequential recommendation model that captures ordered interaction histories.
SASRec	A self-attention-based sequential recommendation model that captures longer-range dependencies.
CGSR-L	The semantic–contextual backbone combining MiniLM-LoRA semantic encoding, Bi-GRU behavioural modelling, gated fusion, and ranking optimisation.
No-LoRA	An ablation variant without parameter-efficient semantic adaptation.

Table 2. *Cont.*

No-Gate	An ablation variant without the learnable gated fusion mechanism.
Static-Fusion	An ablation variant using fixed equal-weight fusion between semantic and behavioural representations.
CGSR-L+CF	The complete collaborative-enhanced architecture combining the semantic–contextual backbone with collaborative preference modelling.

All models were evaluated under both full-corpus and fixed candidate-sampling protocols, including popularity, BPR-MF, NCF-GMF, GRU4Rec, SASRec, CGSR-L, and all ablation variants. This comparison allows the evaluation to distinguish between the contributions of semantic representation learning, behavioural sequence modelling, gated fusion, and collaborative enhancement.

4.4. Implementation Details and Hyperparameters

All models were implemented in PyTorch 2.11.0+cu128 and trained using CUDA 12.8 acceleration on an NVIDIA A100 (NVIDIA Corporation, Santa Clara, CA, USA) with 80 GB of memory. Random seeds were fixed to improve reproducibility. The language-model component used the MiniLM-L6-v2 encoder, which produces 384-dimensional semantic representations. Low-Rank Adaptation was applied specifically to the query, key, and value projection modules of the MiniLM attention layers. The LoRA configuration used a rank of 8, an adaptation scaling factor of 32, and a dropout rate of 0.05. Table 3 summarises the main implementation settings.

Table 3. Main implementation settings.

Component	Setting
Semantic encoder	MiniLM-L6-v2
Semantic hidden size	384
Parameter-efficient adaptation	Low-Rank Adaptation
LoRA rank	8
LoRA scaling factor	32
LoRA dropout	0.05
LoRA target modules	query, key, value
Context encoder	Bidirectional GRU
Maximum sequence length	50
Ranking objective	Bayesian Personalised Ranking
Negative sampling	One sampled unobserved item per positive interaction
Full-corpus K values	5, 10, 20
Candidate-sampling K values	5, 10
Candidate-sampling negatives	100
Hardware	NVIDIA A100 GPU, 80 GB
Framework	PyTorch, Transformers, PEFT

The semantic–contextual backbone used a bidirectional GRU encoder for behavioural sequence modelling, followed by gated fusion and a representation projection layer. The final semantic–contextual learner representation was aligned with item embeddings for ranking. For CGSR-L+CF, the collaborative branch was initialised using embeddings learned by the BPR-MF baseline and then further optimised during ranking training.

Training used the Bayesian Personalised Ranking objective with one positive item and one sampled negative item per learner–item training instance. Negative items were sampled from items not observed in the learner's training history. Gradient clipping was applied to improve stability. For the semantic–contextual variants, training followed a staged procedure in which the language-model backbone was initially frozen and the LoRA parameters were subsequently unfrozen for semantic adaptation. The final collaborative-enhanced model was trained after matrix-factorisation initialization.

4.5. Evaluation Metrics and Statistical Analysis

Recommendation performance was measured using standard top-K ranking metrics: Precision@K, Recall@K, and Normalised Discounted Cumulative Gain at K. Precision@K measures the proportion of recommended items in the top-K list that are relevant. Recall@K measures the proportion of relevant items retrieved in the top-K list. Normalised Discounted Cumulative Gain accounts for the rank position of the relevant item and assigns higher value when relevant items appear closer to the top of the recommendation list.

NDCG@10 was used as the primary metric because it reflects both ranking quality and the practical importance of presenting relevant educational resources near the top of the recommendation list. For full-corpus ranking, metrics were reported at K = 5, 10, and 20. For candidate-sampling evaluation, metrics were reported at K = 5 and 10.

To assess whether observed differences were statistically reliable, paired statistical tests were conducted on per-user NDCG@10 scores. For each comparison, we report paired *t*-test *p*-values, Wilcoxon signed-rank *p*-values, Cohen's d, and effect-size interpretation. This statistical protocol was applied to both full-corpus and candidate-sampling settings. The statistical analysis includes comparisons between CGSR-L+CF and the main collaborative, neural, sequential, and ablation baselines.

Confidence intervals were not reported separately in order to keep the empirical presentation concise, as the interpretation already combines paired significance tests, effect-size estimates, and relative performance gains.

4.6. Computational Efficiency Measurement

Computational efficiency was evaluated to characterise the cost–performance trade-off of the proposed framework. For each model, the following quantities were recorded: total number of parameters, number of trainable parameters, trainable parameter percentage, number of training epochs, average training time per epoch, peak GPU memory consumption, and final training loss.

This analysis distinguishes between the parameter-efficient semantic adaptation mechanism and the overall cost of the complete collaborative-enhanced architecture. The MiniLM semantic adaptation remains parameter-efficient through LoRA, whereas CGSR-L+CF introduces additional trainable collaborative embeddings to improve ranking competitiveness.

The efficiency analysis includes popularity, BPR-MF, NCF-GMF, GRU4Rec, SASRec, No-LoRA, No-Gate, Static-Fusion, CGSR-L, and CGSR-L+CF. These measurements provide a transparent view of the computational trade-off associated with combining parameter-efficient semantic adaptation and collaborative enhancement.

5. Results

This section reports the empirical performance of the proposed collaborative-enhanced semantic–sequential recommendation framework. The evaluation focuses on six aspects: full-corpus ranking performance, fixed candidate-sampling performance, ablation analysis, statistical significance, computational efficiency, and gate-level behaviour. The complete architecture, denoted as CGSR-L+CF, is compared with popularity-based,

matrix-factorisation, neural collaborative filtering, recurrent sequential, self-attention sequential, semantic–contextual, and ablation baselines.

5.1. Full-Corpus Ranking Performance

The full-corpus ranking results are reported in Table 4 and illustrated in Figure 2. This evaluation setting is the most demanding because each held-out test item is ranked against the complete item vocabulary, after excluding items already observed in the learner's training history.

Table 4. Full-corpus ranking performance across all evaluated models.

Model	Precision@5	Recall@5	NDCG@5	Precision@10	Recall@10	NDCG@10	Precision@20	Recall@20	NDCG@20
Popularity	0.0102	0.0511	0.0364	0.0125	0.1246	0.0609	0.0064	0.1288	0.0619
BPR-MF	0.0270	0.1349	0.1232	0.0167	0.1671	0.1335	0.0099	0.1974	0.1411
NCF-GMF	0.0273	0.1365	0.1068	0.0163	0.1634	0.1155	0.0100	0.2001	0.1249
GRU4Rec	0.0259	0.1297	0.1030	0.0152	0.1518	0.1101	0.0091	0.1828	0.1179
SASRec	0.0263	0.1315	0.1075	0.0151	0.1505	0.1136	0.0090	0.1790	0.1208
No-LoRA	0.0268	0.1341	0.0775	0.0158	0.1576	0.0850	0.0092	0.1837	0.0916
No-Gate	0.0261	0.1306	0.1095	0.0155	0.1548	0.1174	0.0090	0.1800	0.1237
Static-Fusion	0.0259	0.1293	0.0737	0.0152	0.1517	0.0808	0.0095	0.1907	0.0907
CGSR-L	0.0268	0.1339	0.0841	0.0157	0.1575	0.0918	0.0095	0.1893	0.0998
CGSR-L+CF	**0.0284**	**0.1418**	**0.1257**	**0.0169**	**0.1688**	**0.1344**	**0.0102**	**0.2035**	**0.1431**

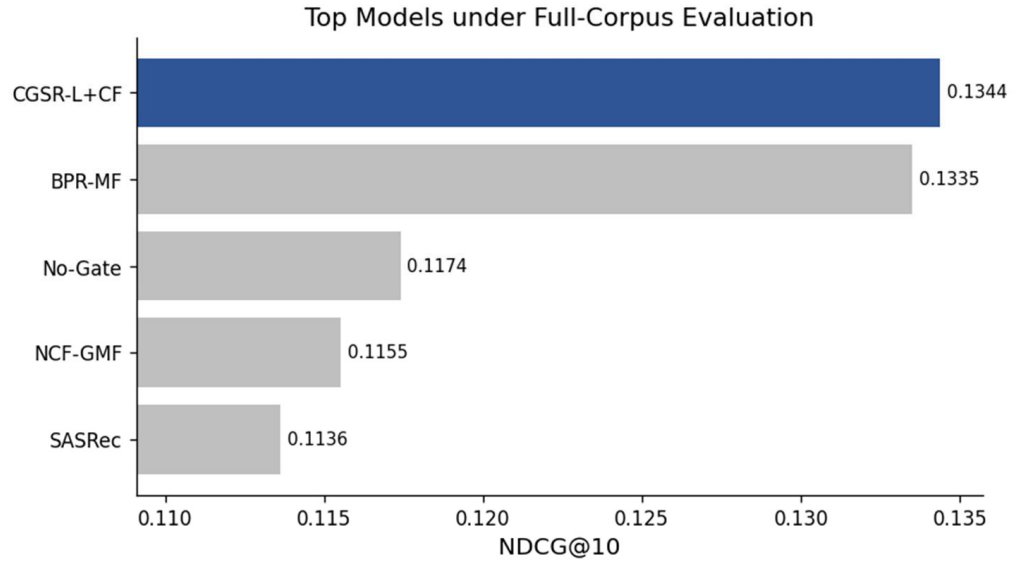


Figure 2. Full-corpus NDCG@10 comparison across all evaluated models.

As shown in Table 4 and Figure 2, CGSR-L+CF achieves the highest performance across all full-corpus metrics. At NDCG@10, it obtains 0.1344, slightly exceeding the strongest collaborative baseline, BPR-MF, which obtains 0.1335.

The margin over BPR-MF is narrow, but CGSR-L+CF also improves over neural collaborative filtering, recurrent sequential recommendation, self-attention sequential recommendation, and the semantic–contextual backbone. These results show that the collaborative-enhanced architecture preserves the strength of collaborative preference modelling while adding useful semantic–contextual information.

5.2. Candidate-Sampling Ranking Performance

Table 5 and Figure 3 report the candidate-sampling results using one held-out positive item and 100 fixed negative items per learner. This protocol provides a controlled setting for comparing all models under the same constrained candidate pool.

Table 5. Candidate-sampling ranking performance using one positive item and 100 fixed negatives.

Model	Precision@5	Recall@5	NDCG@5	Precision@10	Recall@10	NDCG@10
Popularity	0.0818	0.4090	0.3082	0.0534	0.5345	0.3486
BPR-MF	0.1228	0.6141	0.4859	0.0726	0.7262	0.5223
NCF-GMF	0.1232	0.6159	0.4923	0.0728	0.7278	0.5286
GRU4Rec	0.1152	0.5758	0.4509	0.0701	0.7006	0.4913
SASRec	0.1178	0.5890	0.4574	0.0710	0.7102	0.4967
No-LoRA	0.1218	0.6088	0.4759	0.0722	0.7219	0.5125
No-Gate	0.1217	0.6086	0.4749	0.0719	0.7193	0.5108
Static-Fusion	0.1223	0.6115	0.4779	0.0724	0.7239	0.5143
CGSR-L	0.1221	0.6107	0.4763	0.0724	0.7243	0.5132
CGSR-L+CF	**0.1270**	**0.6350**	**0.5041**	**0.0740**	**0.7405**	**0.5383**

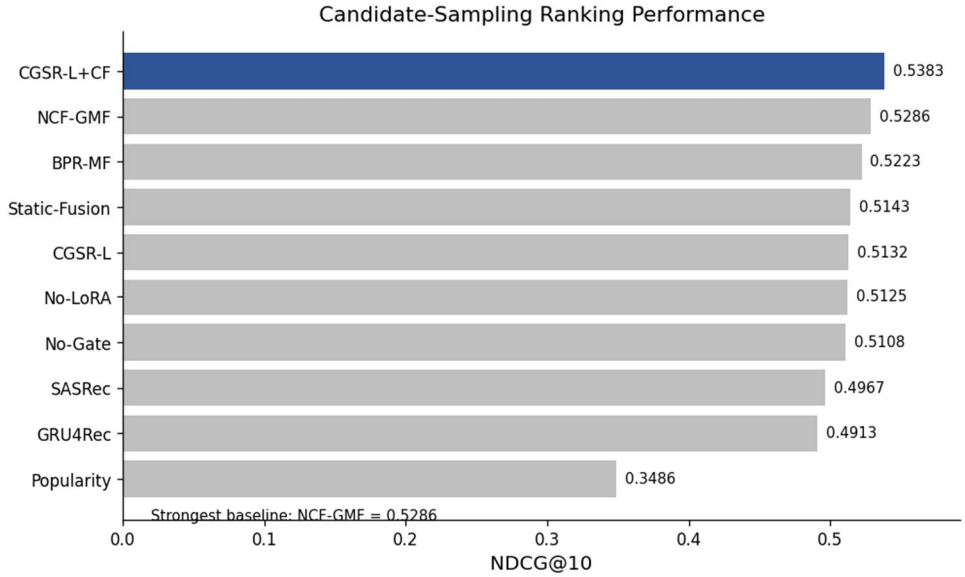


Figure 3. Candidate-sampling NDCG@10 comparison across all evaluated models.

Under candidate sampling, CGSR-L+CF again achieves the strongest overall ranking performance. It obtains an NDCG@10 of 0.5383, compared with 0.5286 for NCF-GMF, and 0.5223 for BPR-MF.

The improvement over the strongest candidate-sampling baseline, NCF-GMF, corresponds to a relative gain of approximately 1.85%. The higher absolute scores compared with full-corpus ranking are expected because the candidate set is smaller; however, the relative ordering remains informative because all models are evaluated using the same fixed candidate sets.

5.3. Ablation Study

The ablation study evaluates the contribution of the main architectural components: parameter-efficient semantic adaptation, gated semantic–contextual fusion, static versus learnable fusion, and collaborative enhancement. The results are reported in Table 6.

Table 6. Ablation study under full-corpus and candidate-sampling evaluation.

Model	NDCG@5 Full	NDCG@10 Full	NDCG@20	NDCG@5 Sampled	NDCG@10 Sampled
No-LoRA	0.0775	0.0850	0.0916	0.4759	0.5125
No-Gate	0.1095	0.1174	0.1237	0.4749	0.5108
Static-Fusion	0.0737	0.0808	0.0907	0.4779	0.5143
CGSR-L	0.0841	0.0918	0.0998	0.4763	0.5132
CGSR-L+CF	**0.1257**	**0.1344**	**0.1431**	**0.5041**	**0.5383**

The ablation results show that the collaborative enhancement branch provides the most substantial improvement over the standalone semantic–contextual backbone. In full-corpus ranking, CGSR-L+CF improves NDCG@10 from 0.0918 for CGSR-L to 0.1344, while under candidate sampling, NDCG@10 increases from 0.5132 to 0.5383. This indicates that semantic–contextual modelling becomes most effective when integrated with collaborative preference modelling, especially under strict ranking conditions. The results also reveal that the contribution of gated fusion is more nuanced: in full-corpus evaluation, the No-Gate variant obtains a higher NDCG@10 than the standalone CGSR-L backbone, suggesting that the gate alone does not uniformly improve performance. However, CGSR-L+CF outperforms all ablation variants, showing that the strongest performance is achieved when semantic adaptation, behavioural modelling, gated fusion, and collaborative preference structure are combined within the same ranking architecture.

5.4. Statistical Significance and Practical Effect Size

To determine whether the observed differences are statistically reliable, paired statistical tests were conducted on per-learner NDCG@10 scores. Table 7 reports paired *t*-test *p*-values, Wilcoxon signed-rank *p*-values, relative gains, Cohen's d, and effect-size interpretation.

Table 7. Statistical significance and effect-size analysis based on per learner NDCG@10.

Evaluation	Comparison	Proposed Mean	Baseline Mean	Relative Gain (%)	*t*-Test *p*	Wilcoxon *p*	Cohen's d	Effect Size
Full-corpus	CGSR-L+CF vs. BPR-MF	0.1344	0.1335	0.65	6.696×10^{-3}	4.870×10^{-2}	0.009	Negligible
Full-corpus	CGSR-L+CF vs. NCF-GMF	0.1344	0.1155	16.31	1.424×10^{-303}	$<1 \times 10^{-300}$	0.127	Negligible
Full-corpus	CGSR-L+CF vs. SASRec	0.1344	0.1136	18.26	4.181×10^{-302}	7.515×10^{-282}	0.126	Negligible
Full-corpus	CGSR-L+CF vs. CGSR-L	0.1344	0.0918	46.45	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$	0.250	Small
Candidate-sampling	CGSR-L+CF vs. NCF-GMF	0.5383	0.5286	1.85	4.503×10^{-36}	1.822×10^{-38}	0.042	Negligible
Candidate-sampling	CGSR-L+CF vs. BPR-MF	0.5383	0.5223	3.07	3.255×10^{-194}	1.073×10^{-238}	0.101	Negligible
Candidate-sampling	CGSR-L+CF vs. SASRec	0.5383	0.4967	8.38	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$	0.178	Negligible
Candidate-sampling	CGSR-L+CF vs. CGSR-L	0.5383	0.5132	4.89	1.338×10^{-294}	$<1 \times 10^{-300}$	0.125	Negligible

The statistical results confirm that CGSR-L+CF significantly outperforms the strongest baselines under both evaluation protocols. However, the effect-size analysis provides a more nuanced interpretation. The improvement over BPR-MF in full-corpus ranking is statistically significant but practically small, with a relative gain of 0.65% and negligible

Cohen's d. Similarly, the improvement over NCF-GMF in candidate sampling is statistically significant but modest, with a relative gain of 1.85% and negligible effect size. These findings show that CGSR-L+CF delivers statistically reliable and consistent gains over strong collaborative baselines, while the effect-size analysis indicates that these gains are incremental in practical magnitude.

5.5. Computational Efficiency

Table 8 reports the computational efficiency measurements, including total parameters, trainable parameters, trainable percentage, training epochs, average time per epoch, peak GPU memory, and final training loss.

Table 8. Computational efficiency and cost–performance characteristics of the evaluated models.

Model	Total Parameters	Trainable Parameters	Trainable %	Epochs	Avg Time/Epoch (s)	Peak GPU Memory (GB)	Final Loss
Popularity	0	0	0.00	0	0.00	0.00	—
BPR-MF	25,513,984	25,513,984	100.00	20	1.04	0.48	0.1843
NCF-GMF	25,514,240	25,514,240	100.00	20	1.04	0.87	0.1450
GRU4Rec	3,603,712	3,603,712	100.00	20	4.76	1.02	0.1294
SASRec	4,735,488	4,735,488	100.00	20	6.08	2.02	0.1803
CGSR-L	26,443,264	3,730,048	14.11	5	971.31	9.76	0.2323
No-LoRA	26,332,672	3,619,456	13.75	5	708.24	2.20	0.2267
No-Gate	26,311,680	3,598,464	13.68	5	969.21	9.99	0.2338
Static-Fusion	26,311,936	3,598,720	13.68	5	971.31	10.10	0.2289
CGSR-L+CF	51,957,249	29,244,033	56.28	3	945.53	19.25	0.1737

The efficiency results clarify the cost–performance trade-off of the proposed approach. The standalone CGSR-L backbone remains parameter-efficient because LoRA updates only a subset of the language-model parameters; it contains 26.44 million total parameters, of which 3.73 million are trainable, corresponding to 14.11% of the model. In contrast, CGSR-L+CF introduces additional trainable collaborative embeddings, increasing the trainable parameter count to 29.24 million and peak GPU memory usage to 19.25 GB. This indicates that the efficiency contribution of the framework lies primarily in parameter-efficient semantic adaptation, while the complete collaborative-enhanced architecture prioritises stronger ranking performance through additional collaborative preference modelling.

5.6. Gate and Hybrid-Fusion Analysis

The gate distribution and hybrid coefficient provide complementary evidence on how CGSR-L+CF combines semantic, behavioural, and collaborative signals. Within the semantic–contextual branch, the average gate value is 0.4888, with a standard deviation of 0.0033. As shown in Figure 4, this indicates a nearly balanced internal fusion between semantic and behavioural representations, with a slight preference toward behavioural context. The narrow distribution also shows stable fusion behaviour across learners rather than highly variable modality switching.

At the branch level, the hybrid coefficient α converges to approximately 0.0959. This indicates that the collaborative branch contributes the dominant part of the final relevance score, while the semantic–contextual branch provides a targeted residual enhancement. This behaviour aligns with the ranking results: the standalone CGSR-L backbone is less competitive than the strongest collaborative baselines, whereas CGSR-L+CF achieves the best performance by integrating semantic–contextual information with collaborative preference modelling.

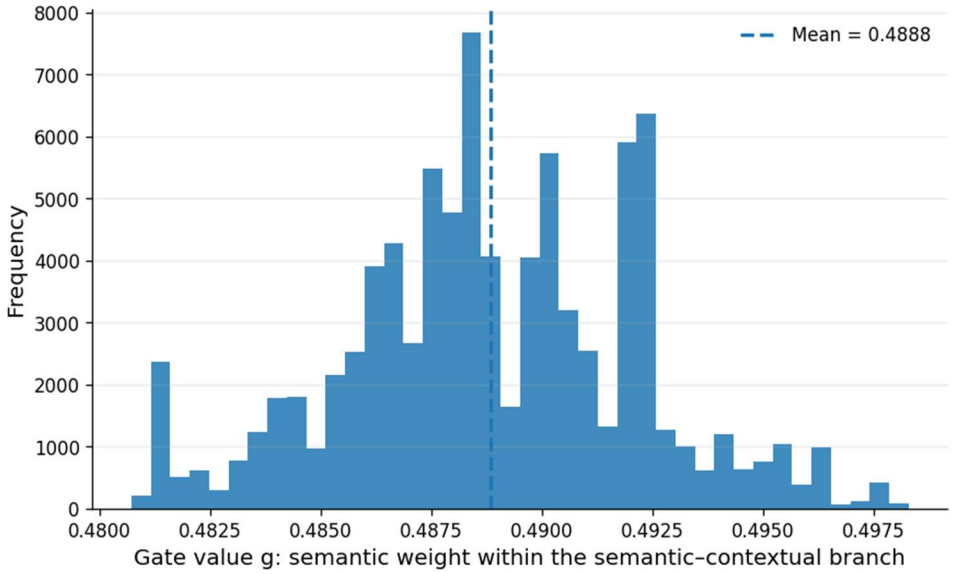


Figure 4. Distribution of gate values in CGSR-L+CF.

Overall, the gate analysis shows that CGSR-L+CF does not rely on semantic or behavioural information in isolation. Instead, it first builds a balanced semantic–behavioural representation within the semantic–contextual branch and then combines this representation with collaborative user–item preference signals at the final scoring stage. This layered fusion strategy explains why the complete architecture outperforms both the standalone semantic–contextual backbone and the strongest collaborative baselines.

6. Discussion

6.1. Main Findings

The results confirm the effectiveness of CGSR-L+CF as a collaborative-enhanced semantic–sequential framework for educational recommendation. Across both full-corpus and candidate-sampling evaluation, the proposed model achieves the best ranking performance among all evaluated baselines and ablation variants. This consistency is important because the two protocols assess different levels of ranking difficulty: full-corpus evaluation tests recommendation under a strict complete-item setting, while candidate sampling provides a controlled comparison using fixed negative sets.

The main finding is that semantic–contextual modelling provides its strongest contribution when integrated with collaborative learner–item preference signals. CGSR-L+CF preserves the strength of collaborative filtering while enriching the final ranking function with semantic and behavioural information. This confirms the value of combining educational metadata, learner behaviour, and collaborative interaction patterns within a unified recommendation architecture.

6.2. Contribution of the Model Components

The ablation results show that the collaborative enhancement branch is the most decisive component for improving ranking performance over the standalone semantic–contextual backbone. This result confirms that learner–item interaction patterns remain essential in educational recommendation, particularly under strict full-corpus ranking.

At the same time, the semantic–contextual branch adds complementary information that is not fully captured by collaborative signals alone. It represents educational descriptors, concept tags, learner responses, instructional metadata, and behavioural sequences. The gate analysis further shows that semantic and behavioural representations are fused

in a nearly balanced manner before being combined with collaborative preference signals. Therefore, the strength of CGSR-L+CF comes from the layered integration of semantic, behavioural, and collaborative evidence rather than from a single isolated component.

6.3. Statistical and Computational Implications

The statistical analysis confirms that the improvements achieved by CGSR-L+CF are reliable across learners. The model significantly outperforms the strongest baselines under both evaluation protocols, including BPR-MF in full-corpus ranking and NCF-GMF in candidate sampling. The effect-size analysis shows that the gains over highly competitive collaborative baselines are incremental in practical magnitude, which is expected in large-scale recommendation settings where matrix-factorisation and neural collaborative models are already strong.

The computational results highlight an accuracy–efficiency trade-off. LoRA enables parameter-efficient adaptation of the semantic encoder by updating only a subset of language-model parameters. The complete CGSR-L+CF model introduces additional collaborative embeddings, increasing trainable parameters and memory usage, but this additional cost supports stronger ranking performance. Thus, the framework combines parameter-efficient semantic adaptation with collaborative enhancement to improve recommendation accuracy under rigorous evaluation conditions.

6.4. Limitations and Future Work

The study has several directions for extension. First, the model was evaluated on EdNet-KT1; future work should validate CGSR-L+CF on additional educational datasets, learning platforms, and subject domains to assess generalisability. Second, the semantic inputs are mainly derived from structured educational metadata and interaction-related fields. Future studies could incorporate richer textual sources, such as learner reflections, teacher feedback, forum discussions, explanations, and open-ended responses.

Third, the current evaluation focuses on offline ranking metrics. Future work should examine educational outcomes in real learning environments, including learner engagement, knowledge gain, persistence, and perceived usefulness. Finally, deployment-oriented improvements such as semantic-vector caching, embedding compression, approximate retrieval, and model distillation could further improve the scalability of CGSR-L+CF in educational platforms with limited computational resources.

7. Conclusions

This study introduced CGSR-L+CF, a collaborative-enhanced semantic–sequential framework for educational recommendation. The model integrates parameter-efficient semantic adaptation, behavioural sequence modelling, gated semantic–contextual fusion, and collaborative learner–item preference modelling within a unified ranking architecture.

The experimental results demonstrate that CGSR-L+CF achieves the strongest overall ranking performance under both full-corpus and fixed candidate-sampling evaluation. Its advantage over collaborative, neural, sequential, semantic–contextual, and ablation baselines confirms the value of combining educational meaning, learner behaviour, and collaborative preference structure rather than treating recommendation as a purely interaction-based task.

The main contribution of this work lies in showing how next-generation educational recommender systems can move beyond conventional ID-based preference modelling toward more context-aware and semantically enriched recommendation. CGSR-L+CF does not replace collaborative filtering; it strengthens it by integrating semantic and behavioural evidence that better reflects the complexity of learning processes. The results also highlight

a realistic accuracy–efficiency trade-off. LoRA enables parameter-efficient adaptation of the semantic encoder, while the collaborative enhancement branch increases model capacity to improve ranking performance. Overall, CGSR-L+CF provides a strong step toward intelligent educational recommendation systems that combine accuracy, scalability, and pedagogical relevance within a unified hybrid framework.

Author Contributions: Conceptualization, H.M., A.J. and L.L.; methodology, H.M. and A.J.; software, H.M. and Y.B.; validation, L.L., A.J., A.Y. and K.A.Z.; formal analysis, L.L. and A.J.; investigation, H.M. and Y.B.; resources, L.L., A.J. and H.M.; data curation, H.M.; writing—original draft preparation, H.M., A.J. and Y.B.; writing—review and editing, H.M. and A.J.; visualisation, H.M. and A.J.; supervision, H.T., K.A.Z. and A.Y.; project administration, H.T., K.A.Z. and A.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data analysed in this study are openly available in the EdNet repository at https://github.com/riiid/ednet (accessed on 4 January 2026). The study used the EdNet-KT1 dataset and associated educational content metadata. The processed data, experimental splits, and results generated during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: During the preparation of this manuscript, the authors used Grammarly for Chrome, version 14.1299.0 for language refinement, editing support, and improvement of clarity and presentation. Grammarly was not used to generate research data or images, conduct the experiments, perform the statistical analyses, or make scientific interpretations. All scientific content, methodological choices, experimental results, and conclusions were reviewed, verified, and approved by the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yürüm, O.R. Technology-Enhanced Multimodal Learning Analytics in Higher Education: A Systematic Literature Review. *IEEE Access* **2025**, *13*, 92057–92073. [CrossRef]
2. Sun, J.C.-Y.; Liu, Y.; Lin, X.; Hu, X. Temporal learning analytics to explore traces of self-regulated learning behaviors and their associations with learning performance, cognitive load, and student engagement in an asynchronous online course. *Front. Psychol.* **2023**, *13*, 1096337. [CrossRef] [PubMed]
3. Wu, J.-Y.; Liao, C.-H.; Tsai, C.-C.; Kwok, O.-M. Using learning analytics with temporal modeling to uncover the interplay of before-class video viewing engagement, motivation, and performance in an active learning context. *Comput. Educ.* **2024**, *212*, 104975. [CrossRef]
4. Ahuja, R.; Jha, A.; Maurya, R.; Srivastava, R. Analysis of Educational Data Mining. In *Harmony Search and Nature Inspired Optimization Algorithms*; Yadav, N., Yadav, A., Bansal, J., Deep, K., Kim, J., Eds.; Springer: Singapore, 2019; Volume 741, pp. 897–907. [CrossRef]
5. Kumar, N.; Sneha, Y.S.; Mungara, J.; Raghavendra Prasad, S.G. A Survey on Data Mining Methods Available for Recommendation System. In Proceedings of the 2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), Bengaluru, India, 21–23 December 2017; pp. 1–6. [CrossRef]
6. Saleem, R.; Aslam, M. A Multi-Faceted Deep Learning Approach for Student Engagement Insights and Adaptive Content Recommendations. *IEEE Access* **2025**, *13*, 69236–69256. [CrossRef]
7. Ahmadian Yazdi, H.; Seyyed Mahdavi, S.J.; Ahmadian Yazdi, H. Dynamic educational recommender system based on improved LSTM neural network. *Sci. Rep.* **2024**, *14*, 4381. [CrossRef]
8. Wang, M.; Zhao, X.; Wang, W.; Zhang, S.; Li, J.; Yu, B.; Wang, B.; Zhou, S.; Yin, D.; Li, Q.; et al. Embedding in Recommender Systems: A Survey. *arXiv* **2023**, arXiv:2310.18608. [CrossRef]
9. Chen, L.; Gao, C.; Du, X.; Luo, H.; Jin, D.; Li, Y.; Wang, M. Enhancing ID-based recommendation with large language models. *ACM Trans. Inf. Syst.* **2025**, *43*, 112. [CrossRef]

10. Adomavicius, G.; Mobasher, B.; Ricci, F.; Tuzhilin, A. Context-aware recommender systems. *AI Mag.* **2011**, *32*, 67–80. [CrossRef]
11. Abu-Rasheed, H.; Weber, C.; Fathi, M. Context based learning: A survey of contextual indicators for personalized and adaptive learning recommendations—A pedagogical and technical perspective. *Front. Educ.* **2023**, *8*, 1210968. [CrossRef]
12. Adomavicius, G.; Bauman, K.; Tuzhilin, A.; Unger, M. Context-aware recommender systems: From foundations to recent developments. In *Recommender Systems Handbook*, 3rd ed.; Springer: New York, NY, USA, 2022; pp. 211–250. [CrossRef]
13. Mastkar, S.; Thakar, U. An exhaustive study on context based recommender systems. *Int. J. Comput. Appl.* **2024**, *186*, 17–21. [CrossRef]
14. Kulkarni, S.; Rodd, S.F. Context aware recommendation systems: A review of the state of the art techniques. *Comput. Sci. Rev.* **2020**, *37*, 100255. [CrossRef]
15. Li, J.; Gao, Y.; Yang, Y.; Bai, Y.; Zhou, X.; Li, Y.; Sun, H.; Liu, Y.; Si, X.; Ye, Y.; et al. Fundamental capabilities and applications of large language models: A survey. *ACM Comput. Surv.* **2025**, *58*, 1–42. [CrossRef]
16. Perez-Toro, P.A.; Dineley, J.; Iniesta, R.; Zhang, Y.; Matcham, F.; Siddi, S.; Lamers, F.; Haro, J.M.; Penninx, B.W.J.H.; Folarin, A.A.; et al. Exploring biases related to the use of large language models in a multilingual depression corpus. *Sci. Rep.* **2025**, *15*, 36197. [CrossRef] [PubMed]
17. Mannekote, A.; Davies, A.; Pinto, J.D.; Zhang, S.; Olds, D.; Schroeder, N.L.; Lehman, B.; Zapata-Rivera, D.; Zhai, C. Large language models for whole-learner support: Opportunities and challenges. *Front. Artif. Intell.* **2024**, *7*, 1460364. [CrossRef]
18. Shehmir, S.; Kashef, R. LLM4Rec: A comprehensive survey on the integration of large language models in recommender systems—Approaches, applications and challenges. *Future Internet* **2025**, *17*, 252. [CrossRef]
19. Lin, J.; Dai, X.; Shan, R.; Chen, B.; Tang, R.; Yu, Y.; Zhang, W. Large language models make sample-efficient recommender systems. *arXiv* **2024**, arXiv:2406.02368. [CrossRef]
20. Nawara, D.; Kashef, R. A comprehensive survey on LLM-powered recommender systems: From discriminative, generative to multi-modal paradigms. *IEEE Access* **2025**, *13*, 145772–145798. [CrossRef]
21. Rajesh, D.B.; Kumar, A. Collaborative filtering models: An experimental and detailed comparative study. *Sci. Rep.* **2025**, *15*, 31667. [CrossRef]
22. Ricci, F.; Rokach, L.; Shapira, B. Recommender systems: Introduction and challenges. In *Recommender Systems Handbook*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2015; pp. 1–34.
23. Koren, Y.; Bell, R.; Volinsky, C. Matrix factorization techniques for recommender systems. *Computer* **2009**, *42*, 30–37. [CrossRef]
24. Adomavicius, G.; Sankaranarayanan, R.; Sen, S.; Tuzhilin, A. Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Trans. Inf. Syst.* **2005**, *23*, 103–145. [CrossRef]
25. Adomavicius, G.; Tuzhilin, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 734–749. [CrossRef]
26. Livne, A.; Shem Tov, E.; Solomon, A.; Elyasaf, A.; Shapira, B.; Rokach, L. Evolving context-aware recommender systems with users in mind. *Expert Syst. Appl.* **2022**, *189*, 116042. [CrossRef]
27. Karatzoglou, A.; Amatriain, X.; Baltrunas, L.; Oliver, N. Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering. In Proceedings of the Fourth ACM Conference on Recommender Systems (RecSys '10), Barcelona, Spain, 26–30 September 2010; pp. 79–86. [CrossRef]
28. Muñoz, E.G.; Parraga-Alava, J.; Meza, J.; Proaño Morales, J.J.; Ventura, S. Housing fuzzy recommender system: A systematic literature review. *Heliyon* **2024**, *10*, e26444. [CrossRef]
29. Hidasi, B.; Karatzoglou, A.; Baltrunas, L.; Tikk, D. Session-based recommendations with recurrent neural networks. *arXiv* **2016**, arXiv:1511.06939. [CrossRef]
30. Kim, D.; Park, C.; Oh, J.; Lee, S.; Yu, H. Convolutional matrix factorization for document context-aware recommendation. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), San Francisco, CA, USA, 13–17 August 2016; pp. 895–904. [CrossRef]
31. Hartatik; Winarko, E.; Heryawan, L. Context-aware recommendation system survey: Recommendation when adding contextual information. In Proceedings of the 2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia, 23–24 November 2022; pp. 7–13. [CrossRef]
32. Majjate, H.; Bellarhmouch, Y.; Jeghal, A.; Yahyaouy, A.; Tairi, H.; Zidani, K.A. Assessing the impact of ethical aspects of recommendation systems on student trust and engagement in e-learning platforms: A multifaceted investigation. *Educ. Inf. Technol.* **2025**, *30*, 3953–3977. [CrossRef]
33. He, X.; Liao, L.; Zhang, H.; Nie, L.; Hu, X.; Chua, T.-S. Neural collaborative filtering. In Proceedings of the 26th International Conference on World Wide Web (WWW '17), Perth, Australia, 3–7 April 2017; pp. 173–182. [CrossRef]
34. Kang, W.-C.; McAuley, J. Self-attentive sequential recommendation. In Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM), Singapore, 17–20 November 2018; pp. 197–206. [CrossRef]
35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: New York, NY, USA, 2017; Volume 30.

36. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
37. Khattab, O.; Zaharia, M. ColBERT: Efficient and effective passage search via contextualized late interaction over BERT. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20), Virtual Event, 25–30 July 2020; pp. 39–48. [CrossRef]
38. Zhang, J.; Zhao, Y.; Saleh, M.; Liu, P.J. PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. In Proceedings of the 37th International Conference on Machine Learning (ICML 2020), Virtual Event, 13–18 July 2020.
39. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. In Proceedings of the 34th International Conference on Neural Information Processing Systems (NeurIPS '20), Vancouver, BC, Canada, 6–12 December 2020.
40. Li, Z.; Wang, Z.; Wang, W.; Hung, K.; Xie, H.; Wang, F.L. Retrieval-augmented generation for educational application: A systematic survey. *Comput. Educ. Artif. Intell.* **2025**, *8*, 100417. [CrossRef]
41. Li, Z.; Wang, J.; Gu, W.; Yazdanpanah, V.; Shi, L.; Cristea, A.I.; Stein, S. TutorLLM: Customizing learning recommendations with knowledge tracing and retrieval-augmented generation. In Proceedings of the IFIP Conference on Human-Computer Interaction, Marseille, France, 1–5 September 2025; pp. 137–146.
42. Delikoura, I.; Fung, Y.R.; Hui, P. From superficial outputs to superficial learning: Risks of large language models in education. *arXiv* **2025**, arXiv:2509.21972. [CrossRef]
43. Li, Y.; Grasman, R. DUALRec: A hybrid sequential and language model framework for context-aware movie recommendation. *arXiv* **2025**, arXiv:2507.13957. [CrossRef]
44. Sakurai, K.; Togo, R.; Ogawa, T.; Haseyama, M. LLM is knowledge graph reasoner: LLM's intuition-aware knowledge graph reasoning for cold-start sequential recommendation. In Proceedings of the Advances in Information Retrieval: 47th European Conference on Information Retrieval (ECIR 2025), Lucca, Italy, 6–10 April 2025; pp. 263–278. [CrossRef]
45. Lv, K.; Yang, Y.; Liu, T.; Guo, Q.; Qiu, X. Full parameter fine-tuning for large language models with limited resources. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024), Bangkok, Thailand, 11–16 August 2024; pp. 8187–8198. [CrossRef]
46. Hu, E.J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W. LoRA: Low-Rank Adaptation of Large Language Models. In Proceedings of the International Conference on Learning Representations (ICLR), Virtual Event, 25–29 April 2022.
47. Dettmers, T.; Pagnoni, A.; Holtzman, A.; Zettlemoyer, L. QLoRA: Efficient finetuning of quantized LLMs. In Proceedings of the 37th International Conference on Neural Information Processing Systems (NeurIPS '23), New Orleans, LA, USA, 10–16 December 2023.
48. Houlsby, N.; Giurgiu, A.; Jastrzebski, S.; Morrone, B.; de Laroussilhe, Q.; Gesmundo, A.; Attariyan, M.; Gelly, S. Parameter-efficient transfer learning for NLP. In Proceedings of the 36th International Conference on Machine Learning (ICML 2019), Long Beach, CA, USA, 9–15 June 2019; pp. 2790–2799.
49. Le, T.D.; Nguyen, T.T.; Nguyen Ha, V. The impact of LoRA adapters for LLMs on clinical NLP classification under data limitations. *arXiv* **2024**, arXiv:2407.19299.
50. Weyssow, M.; Zhou, X.; Kim, K.; Lo, D.; Sahraoui, H. Exploring parameter-efficient fine-tuning techniques for code generation with large language models. *ACM Trans. Softw. Eng. Methodol.* **2025**, *34*, 1–25. [CrossRef]
51. Choi, Y.; Lee, Y.; Shin, D.; Cho, J.; Park, S.; Lee, S.; Heo, J. EdNet: A large-scale hierarchical dataset in education. In Proceedings of the International Conference on Artificial Intelligence in Education, Ifrane, Morocco, 6–10 July 2020; pp. 69–73.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.