Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 20 No. 14 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijim.v20i14.62557

PAPER

Human–Machine Collaborative Knowledge Tracing for Adaptive and Interactive Physics Laboratory Instruction

Lei Su (✉), Changyong Yu, Ping Wang

Faculty of Engineering, Anhui Sanlian University, Hefei, China

suleileisu2026@163.com

ABSTRACT

Accurately characterizing the dynamic evolution of student knowledge states is a foundational prerequisite for designing effective intelligent tutoring systems (ITS) in mobile-supported physics laboratory education. Existing deep learning approaches to knowledge tracing (KT) are predominantly conditioned on student response sequences alone, thereby neglecting two important pedagogical signals in collaborative laboratory environments: human instructor guidance, manifested as teacher demonstrations and corrective interventions, and machine-generated adaptive feedback, manifested as AI-produced hints of varying information quality. To address this limitation, we propose human–machine collaborative knowledge tracing (HMC-KT), a Transformer-based framework that explicitly models the tripartite interaction among student response history, instructor guidance events, and AI feedback quality. Specifically, the human guidance attention (HGA) module encodes instructor intervention signals via gated cross-attention, while the machine feedback attention (MFA) module projects continuous AI hint quality scores into the latent representation space. These two modules are embedded within a causal Transformer backbone and adaptively fused to support context-dependent knowledge state estimation. Extensive experiments on two large-scale public benchmarks—ASSISTments 2009 and ASSISTments 2015—show that HMC-KT achieves consistently superior performance, improving AUC by 3.55 and 4.12 percentage points over the strongest baseline, AKT, respectively, with $p < 0.05$ under paired t-tests. Systematic ablation studies confirm the independent and complementary contributions of HGA and MFA. Attention visualization further reveals that HGA assigns elevated weights to interaction timesteps following teacher demonstration events, forming a "guidance memory" pattern consistent with the Zone of Proximal Development and Cognitive Load Theory.

KEYWORDS

knowledge tracing (KT), intelligent tutoring systems (ITS), physics laboratory education, human–machine collaboration, transformer, gated cross-attention, educational data mining

Su, L., Yu, C., Wang, P. (2026). Human–Machine Collaborative Knowledge Tracing for Adaptive and Interactive Physics Laboratory Instruction. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(14), pp. 132–145. https://doi.org/10.3991/ijim.v20i14.62557

Article submitted 2026-03-25. Revision uploaded 2026-05-27. Final acceptance 2026-06-03.

1 INTRODUCTION

Physics laboratory courses occupy a structurally distinctive position in undergraduate science curricula [1]. Unlike lecture-based instruction, laboratory settings demand the simultaneous integration of conceptual understanding, procedural competence, and experimental reasoning—a convergence that imposes substantial demands on working memory and generates correspondingly elevated levels of cognitive load. The traditional model of laboratory instruction, in which a single instructor iterates among stations to provide demonstrations and corrective feedback, is inherently resource-constrained: it scales poorly with increasing enrollment and yields inconsistent guidance quality across students and sessions.

Over the past decade, computer-based, web-based, and mobile-supported laboratory platforms, together with intelligent tutoring systems (ITS), have emerged as scalable complements to human instruction, enabling students to access experimental guidance, submit responses, request AI-generated hints, and receive adaptive feedback through interactive digital interfaces [2, 3]. However, many existing platforms remain primarily interaction-delivery systems rather than cognition-aware adaptive systems; although they record student actions and feedback events, they rarely model how human instructor guidance and machine-generated support jointly influence the evolution of student knowledge states. The resulting pedagogical environment is fundamentally collaborative in nature: student learning is shaped not by the student's efforts in isolation, but by the joint operation of instructor interventions (demonstrations, corrective feedback, motivational cues) and machine-generated support (structured hints, adaptive difficulty scheduling, error diagnostics). We term this the human–machine collaborative learning environment and argue that it necessitates a commensurate evolution in computational models of student knowledge.

The central computational task is knowledge tracing (KT): given a student's observable history of problem-solving interactions, estimate the latent probability that the student has mastered each relevant knowledge component, and use this estimate to predict future performance [4]. The field was substantially advanced by deep knowledge tracing (DKT) [5], which replaced the Hidden Markov Model formulation of Bayesian Knowledge Tracing (BKT) with Long Short-Term Memory networks, achieving materially higher predictive accuracy. Subsequent architectures introduced external memory structures [6], multi-head self-attention [7], and Rasch-model item difficulty embeddings [8], each yielding incremental improvements on standard benchmarks.

Despite this sustained progress, a critical representational gap persists: the learner interaction sequence is uniformly modeled as a dyadic exchange between student and content, with no provision for the instructor guidance or AI feedback signals that are, in practice, among the most consequential determinants of knowledge acquisition dynamics. Instructor demonstrations reduce the effective complexity of novel tasks by providing procedural templates, thereby lowering the cognitive load imposed on students [9]. AI hints resolve specific conceptual impasses, with efficacy modulated by informational richness relative to the student's current knowledge state [10, 11]. The omission of both signal categories from extant KT models leads to systematic misestimation of knowledge states, particularly at learning inflection points—precisely the moments at which accurate estimation is most consequential for adaptive intervention decisions in mobile-supported and interactive laboratory instruction.

The principal contributions of this work are as follows:

1. **HMC-KT architecture.** We propose the first KT model explicitly architected for human–machine collaborative learning environments. Human–machine collaborative knowledge tracing HMC-KT extends the Transformer backbone with two novel attention modules—human guidance attention (HGA) and machine feedback attention (MFA)—that encode instructor and AI signals as time-varying modulators of the latent knowledge state representation.
2. **Gated cross-attention fusion.** We formulate a dual-stream gated cross-attention fusion mechanism that learns, from data alone, the degree to which the knowledge state estimate at each timestep should be conditioned on collaborative signals. The adaptive gating enables the model to suppress guidance influence when signals are uninformative and amplify it when guidance follows a pattern of sustained incorrect responses—behavior theoretically motivated by scaffolding theory.
3. **Empirical validation at scale.** We conduct rigorous experiments on two large-scale public benchmarks, comparing HMC-KT against four representative deep learning baselines spanning recurrent, memory-augmented, and attention-based KT paradigms. HMC-KT achieves statistically significant, consistent improvements on all three evaluation metrics (AUC, accuracy, RMSE) across both datasets and all five cross-validation folds.
4. **Systematic ablation and interpretability analysis.** We delineate the independent and joint contributions of each proposed component, and provide attention weight visualizations that corroborate the model's pedagogically coherent behavior. A knowledge state trajectory analysis further demonstrates that HMC-KT captures guidance-induced learning accelerations that elude existing baselines.

2 RELATED WORK

2.1 Knowledge tracing: From Bayesian models to deep architectures

Knowledge tracing aims to maintain a probabilistic estimate of student proficiency over a latent knowledge space, updated sequentially as new interaction evidence accumulates [12]. BKT represents each knowledge component as a two-state Hidden Markov Model parameterized by initial knowledge, learning, slip, and guess probabilities [4]. Although BKT provides an interpretable probabilistic formulation, its Markov assumption and shared skill parameters limit its ability to capture long-range dependencies and individual differences; individualized BKT partially addresses the latter by introducing student-specific learning parameters [13].

Deep knowledge tracing advanced KT by using LSTM networks to encode full interaction histories [5], while DKVMN introduced an external key-value memory structure to distinguish concept representations from mastery states [6]. More recent attention-based models, including SAKT and AKT, employ self-attention and item difficulty embeddings to retrieve relevant historical interactions and improve prediction performance [7, 8]. However, these models mainly rely on the student response sequence (q_t, r_t), leaving pedagogical intervention signals outside the model input space.

Prior studies have examined hint effects on learning outcomes as post-hoc analyses [14], and graph-based KT incorporates prerequisite knowledge structures as

inductive bias [15]. Nevertheless, real-time collaborative intervention signals, such as instructor guidance and AI feedback quality, remain insufficiently integrated into knowledge state estimation—the gap addressed in this work.

2.2 Intelligent tutoring systems for physics laboratory education

Intelligent tutoring systems and virtual laboratories have been widely used to provide adaptive feedback and individualized learning support in science education [16, 17]. Prior studies suggest that the effectiveness of feedback depends not only on its occurrence but also on its specificity and alignment with students' current knowledge gaps [10]. Human-in-the-loop ITS designs have also considered instructor feedback, but such signals are often used at the session level rather than modeled as timestep-resolved events [18–20]. Therefore, a unified student model that jointly represents instructor guidance and AI feedback remains underexplored.

2.3 Transformer architectures in educational artificial intelligence

The Transformer architecture [21], through its parallelizable multi-head attention and superior long-range dependency modeling, has become the predominant backbone for educational sequence modeling tasks. Its adoption in KT [7, 8] yields consistent improvements over LSTM-based predecessors. Applications to automatic essay scoring [22] and learning analytics [23] further establish the paradigm's versatility. The interpretability afforded by attention weight visualization is a practical advantage: attention maps identify which past interactions the model deems most informative, providing actionable insights for instructional design.

Our work extends the Transformer-based KT paradigm by introducing gated cross-attention between the student representation stream and two auxiliary guidance streams—a formulation architecturally distinct from the purely self-attentive approaches of SAKT and AKT, and, to the best of our knowledge, the first application of this mechanism for integrating pedagogical intervention signals into a KT model.

3 PROBLEM FORMULATION

Definition 1 (Student Interaction Sequence). Let a student's learning history over T time steps be represented as a sequence of quadruples:

$$\mathcal{H}_T = \{(q_t, r_t, g_t, a_t)\}_{t=1}^{T} \tag{1}$$

where each component is defined as:

- $q_t \in \{1, \ldots, Q\}$ is the knowledge concept (exercise) identifier at step t, with Q the total number of distinct concepts;
- $r_t \in \{0, 1\}$ is the student's binary response (1 = correct, 0 = incorrect);
- $g_t \in \{0, 1, 2, 3\}$ is a categorical guidance indicator (0 = no guidance, 1 = AI hint only, 2 = instructor demonstration only, 3 = both);
- $a_t \in [0, 1]$ is a continuous AI feedback quality score, with $a_t = 0$ when no AI hint is provided.

Definition 2 (Knowledge Tracing in a Collaborative Setting). Given the history $\mathcal{H}_t$ up to and including step t, and the concept identifier q_{t+1} of the next exercise, the KT task is to estimate the probability that the student responds correctly:

$$\hat{r}_{t+1} = P(r_{t+1} = 1 \mid q_{t+1}, \mathcal{H}_t; \theta) \tag{2}$$

where θ denotes all trainable parameters. Since ASSISTments does not natively provide instructor guidance labels, guidance events are generated following Pardos and Heffernan [24], with higher probabilities assigned after repeated incorrect responses on the same skill. The AI feedback quality score a_t is derived from the platform's hint request metadata.

4 METHODOLOGY

4.1 Architectural overview

HMC-KT processes three synchronized input streams: the exercise-response sequence (q_t, r_t), guidance type sequence (g_t), and AI feedback quality sequence (a_t). These streams are mapped into a shared d-dimensional latent space and passed through N stacked Transformer blocks. Each block contains causal self-attention (CSA), HGA, MFA, and a feed-forward network. HGA and MFA introduce gated cross-attention to adaptively fuse instructor guidance and AI feedback into the student knowledge representation.

4.2 Input encoding

Following established practice in deep KT [5, 7], the exercise-response token at step t is constructed by indexing into an embedding matrix of size $2Q \times d$:

$$\mathbf{z}_t = \mathrm{Emb}\,(q_t \cdot 2 + r_t) \in \mathbb{R}^d \tag{3}$$

This encoding distinguishes correct and incorrect interactions with the same concept. To incorporate item difficulty, Rasch-model scaling [8] is applied:

$$\mathbf{z}_t \leftarrow \mathbf{z}_t \odot \sigma(\mathrm{Diff}(q_t)) \tag{4}$$

where, $\mathrm{Diff}(\cdot)$: $\{1, \ldots, Q\} \rightarrow \mathbb{R}^d$ is a learnable difficulty embedding, $\sigma(\cdot)$ is the element-wise sigmoid function, and $\odot$ denotes the Hadamard product. Equation (4) adjusts each interaction representation according to the learned difficulty profile of the corresponding concept. Sinusoidal positional encodings [21] are added to inject temporal ordering information:

$$\mathbf{z}_t \leftarrow \mathbf{z}_t + \mathbf{PE}(t), \mathbf{PE}(t)_{2i} = \sin\left(\frac{t}{10000^{2i/d}}\right) \tag{5}$$

The sequence of encoded tokens is collected into a matrix $\boldsymbol{X} = [\mathbf{z}_1, \ldots, \mathbf{z}_T] \in \mathbb{R}^{T \times d}$, serving as input to the first HMC-KT block.

4.3 Causal self-attention

The CSA sub-module applies multi-head scaled dot-product attention over the student sequence, enforcing causality through a lower-triangular mask. For a single attention head with weight matrices $\boldsymbol{W}^{Q}, \boldsymbol{W}^{K}, \boldsymbol{W}^{V} \in \mathbb{R}^{d \times d_h}$ ($d_h = d/H$):

$$\mathrm{Att}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \mathrm{softmax}\left(\frac{QK^{T}}{\sqrt{d_h}} + \mathbf{M}_{\mathrm{causal}}\right)\mathbf{V} \tag{6}$$

where, $\boldsymbol{M}_{causal} \in \{0, -\infty\}^{T \times T}$ with $M_{ij} = 0$ if $j \le i$ and $M_{ij} = -\infty$ otherwise. The multi-head variant concatenates H such heads and projects through $\boldsymbol{W}^{O} \in \mathbb{R}^{d \times d}$:

$$\mathrm{CSA}(\mathbf{X}) = \mathrm{MultiHead}\,(\mathbf{X}, \mathbf{X}, \mathbf{X}; \mathbf{M}_{\mathrm{causal}}) \tag{7}$$

$$\mathbf{X}' = \mathrm{LN}(\mathbf{X} + \mathrm{Dropout}(\mathrm{CSA}(\mathbf{X}))) \tag{8}$$

4.4 Human guidance attention

The HGA module incorporates instructor guidance signals that are not represented in the standard exercise-response sequence. We first embed the guidance type sequence through a learnable embedding matrix $\boldsymbol{E}_{G} \in \mathbb{R}^{4 \times d}$:

$$\mathbf{G} = [\mathbf{E}_{G}[g_1], \ldots, \mathbf{E}_{G}[g_T]] \in \mathbb{R}^{T \times d} \tag{9}$$

Cross-attention is then computed with the student representation as queries and the guidance embedding as keys and values:

$$\mathbf{X}_{\mathrm{HGA}} = \mathrm{MultiHead}(\mathbf{X}', \mathbf{G}, \mathbf{G}; \mathbf{M}_{\mathrm{causal}}) \tag{10}$$

A learned gating vector $\boldsymbol{\lambda}^{H} \in [0, 1]^{T \times d}$ controls the contribution of the guidance-attended representation:

$$\lambda^{H} = \sigma(\mathbf{W}_{g}[\mathbf{X}'; \mathbf{X}_{\mathrm{HGA}}] + \mathbf{b}_{g}) \tag{11}$$

$$\mathbf{X}'' = \mathrm{LN}(\mathbf{X}' + \lambda^{H} \odot \mathbf{X}_{\mathrm{HGA}}) \tag{12}$$

where, $\boldsymbol{W}_{g} \in \mathbb{R}^{d \times 2d}$ and $\boldsymbol{b}_{g} \in \mathbb{R}^{d}$ are learnable parameters, and $[\cdot;\cdot]$ denotes column-wise concatenation. The gating mechanism is theoretically motivated by scaffolding theory [11]: guidance interventions are most beneficial—and hence should most strongly modulate the knowledge state representation—when the student is operating within the Zone of Proximal Development. The gate learns to approach zero when recent responses indicate mastery (suppressing guidance influence) and to approach one when the student history exhibits sustained difficulty (amplifying guidance influence). This adaptive behavior emerges entirely from data, without any hard-coded mastery threshold.

4.5 Machine feedback attention

The MFA module encodes the continuous AI feedback quality signal. Unlike categorical instructor guidance, AI feedback quality is represented as a continuous signal.

Raw quality scores $a_t \in [0, 1]$ are projected to the d-dimensional latent space via a two-layer MLP with ReLU activation:

$$\mathbf{A} = \mathrm{MLP}(a_{1:T}) \equiv \left[\phi(a_1), \ldots, \phi(a_T)\right] \in \mathbb{R}^{T \times d} \tag{13}$$

where, $\phi(a_t) = \boldsymbol{W}_2 \mathrm{ReLU}(\mathbf{W}_1 a_t + \boldsymbol{b}_1) + \boldsymbol{b}_2$, with $\boldsymbol{W}_1 \in \mathbb{R}^{d \times 1}$, $\boldsymbol{W}_2 \in \mathbb{R}^{d \times d}$. The MLP is preferable to a direct linear projection because the relationship between hint quality and its effect on knowledge state is non-linear: high-quality hints on mastered concepts may have negligible impact, while identical hints at the boundary of comprehension may substantially alter the knowledge state. Cross-attention and gated fusion are applied analogously to HGA:

$$\mathbf{X}_{\mathrm{MFA}} = \mathrm{MultiHead}(\mathbf{X}'', \mathbf{A}, \mathbf{A}; \mathbf{M}_{\mathrm{causal}}) \tag{14}$$

$$\boldsymbol{\lambda}^M = \sigma(\mathbf{W}_m [\mathbf{X}''; \mathbf{X}_{\mathrm{MFA}}] + \mathbf{b}_m) \tag{15}$$

$$\mathbf{X}''' = \mathrm{LN}(\mathbf{X}'' + \lambda^M \odot \mathbf{X}_{\mathrm{MFA}}) \tag{16}$$

MFA is applied after HGA so that AI feedback can be modeled on top of a guidance-aware student representation.

4.6 Feed-forward network and output prediction

Each HMC-KT block concludes with a position-wise feed-forward network with GELU activation:

$$\mathrm{FFN}(\mathbf{x}) = \mathbf{W}_2\ \mathrm{GELU}\ (\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2 \tag{17}$$

$$\mathbf{X}_{\mathrm{out}} = \mathrm{LN}(\mathbf{X}''' + \mathrm{Dropout}(\mathrm{FFN}(\mathbf{X}'''))) \tag{18}$$

where, $\boldsymbol{W}_1 \in \mathbb{R}^{d_{ff} \times d}$, $\boldsymbol{W}_2 \in \mathbb{R}^{d \times d_{ff}}$, and $d_{ff} = 4d$. After N stacked blocks, the prediction for timestep $t + 1$ is produced by a linear layer followed by sigmoid:

$$\hat{r}_{t+1} = \sigma(\mathbf{w}_o^{\top} \mathbf{X}_{\mathrm{out},t}) \tag{19}$$

where, $\boldsymbol{w}_o \in \mathbb{R}^d$ is a learnable weight vector.

4.7 Training objective

The model minimizes the mean binary cross-entropy loss over all valid (non-padding) positions across the mini-batch:

$$\mathcal{L}(\theta) = -\frac{1}{|\mathcal{M}|} \sum_{(b,t) \in \mathcal{M}} \left[r_{t+1}^{(b)} \widehat{\log r_{t+1}^{(b)}} + (1 - r_{t+1}^{(b)}) \log(1 - \hat{r}_{t+1}^{(b)}) \right] \tag{20}$$

where, $\mathcal{M} = \{(b, t)$: position t in batch item b is unmasked$\}$. L2 weight regularization with coefficient $\lambda_{\mathrm{reg}} = 10^{-5}$ is applied to all non-embedding parameters.

4.8 Computational complexity

For a sequence of length T with model dimension d and H attention heads, the per-layer time complexity of each attention sub-module is $\mathcal{O}(T^2 d + T d^2)$, consistent

with standard Transformer scaling. The HGA and MFA modules each add one cross-attention and one gating computation per layer, contributing a factor of approximately 3× over a single self-attention layer. With $N = 3$ layers, the overall complexity is $\mathcal{O}(NT^2d)$, rendering HMC-KT tractable for sequences up to $T = 200$ steps on modern GPU hardware.

5 EXPERIMENTS

5.1 Datasets

We evaluate on two large-scale benchmark datasets drawn from the ASSISTments intelligent tutoring platform [18], one of the most extensively used sources for KT research. Table 1 summarizes the dataset statistics.

The two datasets differ in sequence length, student population size, and correct rate, providing complementary settings for evaluating HMC-KT. ASSISTments 2009 contains longer sequences and a lower correct rate, while ASSISTments 2015 includes more students and shorter sequences. These differences allow us to examine the robustness of the proposed model across distinct KT scenarios.

Table 1. Correct Rate denotes

Dataset	Students	Skills	Interactions	Avg. Seq. Len.	Correct Rate
ASSISTments 2009	4,151	110	325,637	78.4	65.7%
ASSISTments 2015	19,840	100	683,801	34.5	74.1%

Sequences exceeding 200 interaction steps are partitioned into overlapping windows of length 200 with stride 100 [8]. Students with fewer than two interactions are excluded. Skill identifiers are re-indexed from zero.

5.2 Baseline models

We compare HMC-KT with representative KT baselines, including BKT [4], DKT [5], DKVMN [6], SAKT [7], and AKT [8]. These baselines cover classical Bayesian modeling, recurrent neural networks, memory-augmented networks, and attention-based architectures. For deep learning baselines, we follow commonly used configurations reported in prior KT studies. None of the baselines receives instructor guidance or AI feedback quality as input, ensuring that the comparison evaluates the contribution of the proposed collaborative modeling modules.

5.3 Implementation details

HMC-KT uses 3 Transformer blocks, model dimension 128, 8 attention heads, feed-forward dimension 512, and dropout rate 0.2. All models are trained with Adam using cosine annealing for 100 epochs and batch size 64. We conduct 5-fold cross-validation stratified by student identifier on an NVIDIA RTX 3090 GPU. Statistical significance is tested using paired t-tests against AKT across five folds, with significance declared at $p < 0.05$.

5.4 Main results

Table 2 reports on performance on both benchmarks. HMC-KT achieves the highest AUC, accuracy, and lowest RMSE on all six metric–dataset combinations.

On ASSISTments 2009, HMC-KT outperforms AKT by +3.55% AUC, +2.52% accuracy, and –0.0211 RMSE. On ASSISTments 2015, the corresponding improvements are +4.12% AUC, +2.59% accuracy, and –0.0223 RMSE. The larger gain on ASSISTments 2015 may be attributed to its richer hint-related interactions and larger student population, which provide more informative AI feedback quality signals and more diverse guidance patterns for model training. Overall, the consistent superiority of HMC-KT over DKT, DKVMN, SAKT, and AKT indicates that collaborative guidance signals offer complementary information beyond student response sequences and item difficulty. The low standard deviations across folds further suggest that the observed improvements are stable rather than fold specific.

Table 2. Performance comparison (5-fold CV, mean ± std)

Model	AUC (09)	ACC (09)	RMSE (09)	AUC (15)	ACC (15)	RMSE (15)
DKT	0.7423 ± 0.0031	0.7512 ± 0.0028	0.4312 ± 0.0022	0.7198 ± 0.0035	0.7401 ± 0.0031	0.4498 ± 0.0025
DKVMN	0.7561 ± 0.0027	0.7598 ± 0.0024	0.4218 ± 0.0019	0.7312 ± 0.0030	0.7489 ± 0.0027	0.4389 ± 0.0022
SAKT	0.7714 ± 0.0024	0.7723 ± 0.0021	0.4101 ± 0.0018	0.7445 ± 0.0027	0.7563 ± 0.0024	0.4278 ± 0.0019
AKT	0.7832 ± 0.0022	0.7791 ± 0.0019	0.4023 ± 0.0016	0.7531 ± 0.0025	0.7612 ± 0.0022	0.4201 ± 0.0018
HMC-KT (Ours)†	0.8187 ± 0.0018	0.8043 ± 0.0016	0.3812 ± 0.0014	0.7943 ± 0.0021	0.7871 ± 0.0018	0.3978 ± 0.0015

Notes: †Significant improvement over AKT ($p < 0.05$, paired t-test). Best results in bold.

5.5 Ablation study

As shown in Table 3, the ablation results confirm the effectiveness of the proposed collaborative attention modules. Removing HGA leads to the largest individual AUC drop (–0.0223), indicating the strong contribution of instructor guidance, while removing MFA also causes clear performance degradation (–0.0175), showing that AI feedback quality provides complementary information. The joint removal of HGA and MFA results in the largest overall decline (–0.0355), further demonstrating the complementary effect of human guidance and machine feedback. In contrast, removing positional encoding or Rasch embedding produces smaller but consistent decreases.

Table 3. Ablation study on ASSISTments 2009 (5-fold CV)

Variant	AUC	ACC	RMSE	ΔAUC
HMC-KT (Full)	0.8187	0.8043	0.3812	—
w/o HGA	0.7964	0.7882	0.3994	–0.0223
w/o MFA	0.8012	0.7913	0.3967	–0.0175
w/o HGA + MFA (≡ AKT)	0.7832	0.7791	0.4023	–0.0355
w/o Positional Encoding	0.8051	0.7961	0.3941	–0.0136
w/o Rasch Embedding	0.8103	0.7997	0.3891	–0.0084

5.6 Effect of guidance type on prediction performance

Table 4 stratify performance by the predominant guidance condition in each student's validation-set sequence.

As shown in Table 4, the guidance-type analysis further shows that prediction performance improves as collaborative support becomes richer. Sequences involving both teacher demonstrations and AI hints achieve the highest AUC and accuracy, followed by teacher-only and AI-only conditions. This pattern supports the complementarity between human guidance and machine-generated feedback in collaborative learning environments.

Table 4. Performance by guidance condition (ASSISTments 2009, HMC-KT)

Guidance Condition	Sequences	AUC	ACC
No Guidance	1,243	0.7931	0.7841
AI Hint Only	892	0.8074	0.7968
Teacher Demo Only	548	0.8213	0.8089
Both (HMC Condition)	321	0.8412	0.8247

5.7 Attention visualization

The CSA map exhibits the expected monotonic decay pattern: attention weights diminish approximately geometrically with temporal distance, reflecting the spacing effect whereby the predictive value of past interactions decays over time. The HGA map exhibits a qualitatively different structure: attention weights spike sharply at positions corresponding to teacher demonstration events and remain elevated for several subsequent timesteps before gradually returning to baseline. This "guidance memory" pattern is coherent with cognitive load theory [9]: a teacher demonstration reduces the schema construction burden on the student, and its representational benefit persists until the student consolidates the demonstrated procedure into long-term memory or encounters a sufficiently different task context. The alignment between the model's learned attention behavior and these theoretical predictions provides qualitative evidence that HGA encodes pedagogically meaningful structure rather than statistical artifacts.

5.8 Knowledge state evolution

As shown in Figure 1, HGA assigns elevated attention weights to timesteps following teacher demonstration events, indicating that the model captures a "guidance memory" effect in the representation space. This pattern is further reflected in the estimated knowledge state trajectories shown in Figure 2. Compared with AKT, which produces relatively smooth and synchronized trajectories across concepts, HMC-KT generates more differentiated trajectories and clear stepwise increases around guidance events. These results suggest that HMC-KT can translate instructor interventions into upward revisions of the estimated knowledge state and capture guidance-induced learning acceleration, thereby supporting more timely adaptive instructional decisions.

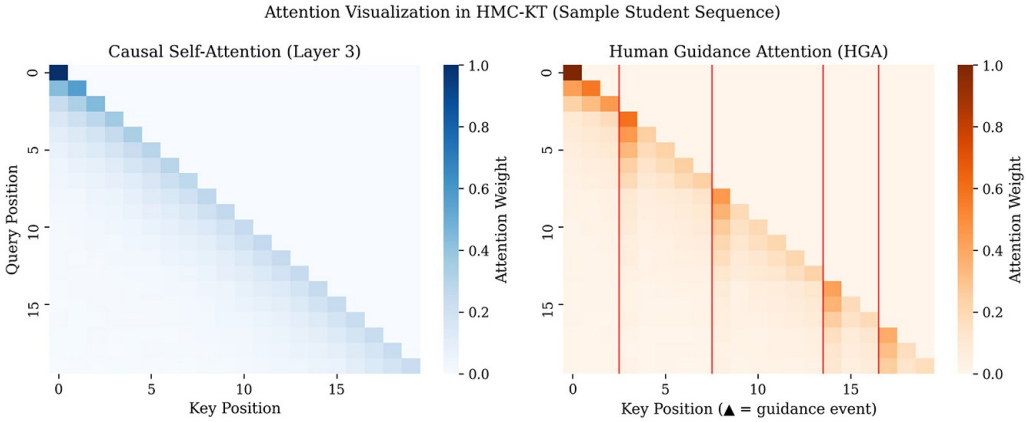

Fig. 1. Attention weight visualization for a sample student sequence (20 interactions); Left: Causal self-attention (Layer 3) shows recency bias. Right: Human Guidance Attention (HGA) exhibits markedly elevated weights at positions following teacher demonstration events (red vertical lines), persisting for several subsequent query positions—the "guidance memory" effect

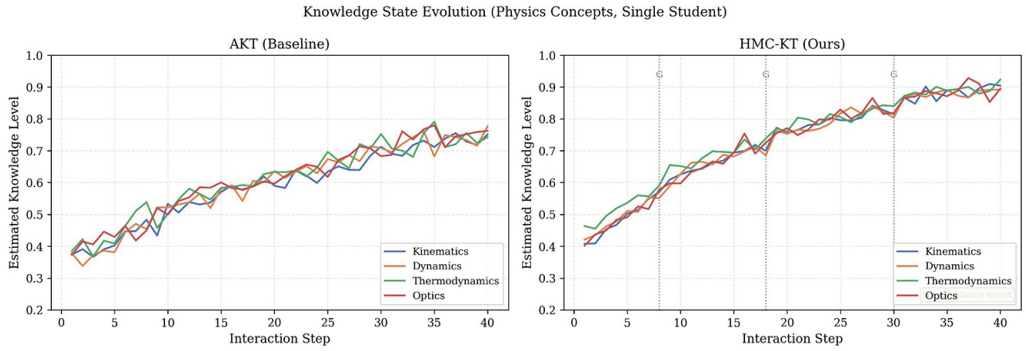

Fig. 2. Estimated knowledge state trajectories over 40 interaction steps for a single student (four physics concepts); Left: AKT baseline. Right: HMC-KT with teacher demonstration events marked (G, dashed lines)

6 DISCUSSION

6.1 Pedagogical implications

The results suggest that instructor interventions and AI feedback quality should be treated as important pedagogical signals in adaptive laboratory learning platforms. The contribution of HGA indicates the value of recording teacher demonstrations and corrective interventions through lightweight annotation interfaces, while MFA shows that AI-generated hints should be evaluated not only by their occurrence but also by their informational quality and relevance to the learner's current knowledge state. Together, HGA and MFA suggest that human and AI interventions should be designed as complementary support: instructors provide procedural and conceptual scaffolding, while AI hints address localized knowledge gaps, thereby enabling more accurate and adaptive instructional decision-making.

6.2 Limitations and threats to validity

This study has several limitations. First, the instructor guidance labels are simulated using a probabilistic annotation protocol [24] rather than collected from fully

instrumented laboratory sessions, so their fidelity to real teacher interventions remains uncertain. Second, instructor guidance may be endogenous because teachers are more likely to intervene when students are struggling, which may introduce selection effects into the estimated contribution of HGA. Third, HMC-KT introduces additional computational overhead compared with AKT, including a 21.7% increase in per-epoch training time and a 34.2% increase in parameter count. Future work should validate the framework on real mobile-supported laboratory platforms and explore lightweight deployment strategies such as knowledge distillation or attention head pruning.

6.3 Future directions

Several extensions merit systematic investigation. First, knowledge graph structure [15]—encoding prerequisite relationships between physics concepts—could be incorporated as a third cross-attention stream alongside HGA and MFA, providing an inductive bias that constrains knowledge state estimates to be consistent with the curriculum dependency structure. Second, the natural downstream application of HMC-KT is guidance *policy optimization*: using HMC-KT as the student model within a reinforcement learning framework to learn when and how to intervene with each individual student. Third, extension to multimodal inputs—including video recordings of student experimental procedures, audio of instructor comments, and sensor data from laboratory instruments—would enable a more complete representation of the human–machine collaborative environment, at the cost of substantially increased data collection and modeling complexity.

7 CONCLUSION

This paper introduced HMC-KT, a Human–Machine Collaborative Knowledge Tracing framework that incorporates instructor guidance and AI feedback quality into knowledge state estimation. By integrating HGA and MFA into a causal Transformer backbone with adaptive gating, HMC-KT explicitly models the tripartite interaction among students, instructors, and AI systems. Experiments on ASSISTments 2009 and ASSISTments 2015 demonstrate that HMC-KT achieves statistically significant improvements over the strongest baseline, with AUC gains of +3.55% and +4.12%, respectively.

Ablation studies confirm the independent and complementary contributions of HGA and MFA, while attention visualization and knowledge state trajectory analysis show that the model captures pedagogically meaningful guidance effects and guidance-induced learning acceleration. These results suggest that modeling collaborative instructional signals can improve the accuracy and interpretability of KT, providing a useful student-modeling foundation for adaptive and interactive physics laboratory instruction.

8 ACKNOWLEDGMENT

This work was funded by the Quality Engineering Project of Anhui Sanlian University: General Education Course in Artificial Intelligence (University Physics) (Grant No.: 25zlgc032); and the Construction of Integrated Ideological and Political Demonstration Course in College Physics (University Physics–Middle School Physics) (Grant No.: 25zlgc034).

9 REFERENCES

[1] N. G. Holmes, C. E. Wieman, and D. A. Bonn, "Teaching critical thinking," *Proceedings of the National Academy of Sciences*, vol. 112, no. 36, pp. 11199–11204, 2015. https://doi.org/10.1073/pnas.1505329112

[2] J. Hu, "Integrated UAV swarm and human-machine interaction platform for real-time mobile training and evaluation," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 19, no. 23, pp. 83–97, 2025. https://doi.org/10.3991/ijim.v19i23.59249

[3] J. Alostad, "Integrating human-computer interaction and software engineering for enhanced usability using support vector machines," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 19, no. 16, pp. 41–59, 2025. https://doi.org/10.3991/ijim.v19i16.53159

[4] A. T. Corbett and J. R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," *User Modeling and User-Adapted Interaction*, vol. 4, no. 4, pp. 253–278, 1995. https://doi.org/10.1007/bf01099821

[5] C. Piech, *et al.*, "Deep knowledge tracing," *Advances in Neural Information Processing Systems*, vol. 28, pp. 505–513, 2015.

[6] J. Zhang, X. Shi, I. King, and D. Yeung, "Dynamic key-value memory networks for knowledge tracing," in *WWW '17: 26th International World Wide Web Conference*, Perth Australia, 2017, pp. 765–774. https://doi.org/10.1145/3038912.3052580

[7] H. Z. Gu, J. L. Ma, Y. R. Zhao, and A. Khadka, "Enhanced named entity recognition based on multi-feature fusion using dual graph neural networks," *Acadlore Transactions on AI and Machine Learning*, vol. 3, no. 2, pp. 84–93, 2024. https://doi.org/10.56578/ataiml030202

[8] A. Ghosh, N. Heffernan, and A. S. Lan, "Context-aware attentive knowledge tracing," in *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Virtual Event CA USA, 2020, pp. 2330–2339. https://doi.org/10.1145/3394486.3403282

[9] J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cognitive Science*, vol. 12, no. 2, pp. 257–285, 1988. https://doi.org/10.1016/0364-0213(88)90023-7

[10] K. R. Koedinger and V. Aleven, "Exploring the assistance dilemma in experiments with cognitive tutors," *Educational Psychology Review*, vol. 19, no. 3, pp. 239–264, 2007. https://doi.org/10.1007/s10648-007-9049-0

[11] L. S. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes*. vol. 86. Harvard University Press, 1978.

[12] R. Pelánek, "Bayesian knowledge tracing, logistic models, and beyond: An overview of learner modeling techniques," *User Modeling and User-Adapted Interaction*, vol. 27, nos. 3–5, pp. 313–350, 2017. https://doi.org/10.1007/s11257-017-9193-2

[13] M. V. Yudelson, K. R. Koedinger, and G. J. Gordon, "Individualized bayesian knowledge tracing models," in *International Conference on Artificial Intelligence in Education*, Cham: Springer International Publishing, 2013, pp. 171–180. https://doi.org/10.1007/978-3-642-39112-5_18

[14] V. Aleven, B. McLaren, I. Roll, and K. Koedinger, "Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor," *International Journal of Artificial Intelligence in Education*, vol. 16, no. 2, pp. 101–128, 2006. https://doi.org/10.3233/irg-2006-16(2)02

[15] A. Duval and F. D. Malliaros, "GraphSVX: Shapley value explanations for graph neural networks," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Bilbao, Spain, 2021.

[16] Z. C. Zacharia, "Examining whether touch sensory feedback is necessary for science learning through experimentation: A literature review of two different lines of research across K-16," *Educational Research Review*, vol. 16, pp. 116–137, 2015. https://doi.org/10.1016/j.edurev.2015.10.001

[17] T. de Jong, M. C. Linn, and Z. C. Zacharia, "Physical and virtual laboratories in science and engineering education," *Science*, vol. 340, no. 6130, pp. 305–308, 2013. https://doi.org/10.1126/science.1230579

[18] N. T. Heffernan and C. L. Heffernan, "The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching," *International Journal of Artificial Intelligence in Education*, vol. 24, no. 4, pp. 470–497, 2014. https://doi.org/10.1007/s40593-014-0024-x

[19] C. Conati, A. Gertner, and K. VanLehn, "Using bayesian networks to manage uncertainty in student modeling," *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 371–417, 2002. https://doi.org/10.1023/a:1021258506583

[20] M. Chi, K. VanLehn, D. Litman, and P. Jordan, "Empirically evaluating the application of reinforcement learning to the induction of effective and adaptive pedagogical strategies," *User Modeling and User-Adapted Interaction*, vol. 21, nos. 1–2, pp. 137–180, 2011. https://doi.org/10.1007/s11257-010-9093-1

[21] A. Vaswani *et al.*, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, pp. 5998–6008, 2017.

[22] F. Dong, Y. Zhang, and J. Yang, "Attention-based recurrent convolutional neural network for automatic essay scoring," in *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, Vancouver, Canada, 2017, pp. 153–162. https://doi.org/10.18653/v1/K17-1017

[23] Y. Choi *et al.*, "Towards an appropriate query, key, and value computation for knowledge tracing," in *L@S '20: Seventh (2020) ACM Conference on Learning @ Scale*, Virtual Event USA, 2020, pp. 341–344. https://doi.org/10.1145/3386527.3405945

[24] Z. A. Pardos and N. T. Heffernan, "Modeling individualization in a bayesian networks implementation of knowledge tracing," in *International Conference on User Modeling, Adaptation, and Personalization*, Berlin, Heidelberg: Springer International Publishing, 2010, pp. 255–266. https://doi.org/10.1007/978-3-642-13470-8_24

10 AUTHORS

Lei Su studied in the Department of Physics, Anhui Normal University and received her bachelor's degree in 2004. From 2004 to 2006, She worked at Anhui Light Industry Technician College. From 2006 to 2009, She studied at Guangxi Normal University and received her master's degree in 2009. Currently, she works at Anhui Sanlian University. She has published eight papers, one of which has been indexed by SCI. Her research interests include theoretical physics (E-mail: suleileisu2026@163.com).

Changyong Yu studied in the Department of Physics, Anhui Normal University and received his bachelor's degree in 2000. Currently, he works at Anhui Sanlian University. He has published two papers, one of which has been indexed by SCI (E-mail: chyyu0403@163.com).

Ping Wang studied in the Department of Physics, Anhui University and received her bachelor's degree in 2001. From 2001 to 2006, she worked at the PLA Artillery Academy. From 2006 to 2009, she studied at the PLA Army Academy and received her master's degree in 2009. Currently, she works at Anhui Sanlian University. She has published six papers, two of which have been indexed by EI, one of which has been indexed by SCI. Her research interests include physics teaching and optical detection (E-mail: mcwangping1978@163.com).