Fair When Sparse, Accurate When Rich: RL-Gated Conversational Fusion for Temporally Equitable Knowledge Tracing

Fan Zhang
University of Florida
f.zhang1@ufl.edu

Rui Guo
University of Miami
rui.guo@miami.edu

Hai Li
University of Florida
li.ha@ufl.edu

ABSTRACT

Knowledge Tracing (KT) models often underperform for students with short interaction histories, creating a temporal equity gap between early- and late-stage learners. We tackle this issue by fusing behavioral traces with dialogue-derived signals using a lightweight reinforcement-learning (RL) gate that adaptively reweights evidence across the learning trajectory. To make temporal disparities explicit, we evaluate by student-level practice-volume cohorts (deciles from 10% to 100%) and report ACC/AUC within cohorts and overall. Across both backbones, the RL-gated variant consistently improves aggregate accuracy and AUC while reducing across-cohort variance, yielding flatter, more reliable cohort curves relative to a behavioral-only baseline and a naïve feature-augmented model without gating. These findings suggest a practical design principle for next-generation knowledge tracing systems: when interaction histories are sparse, semantically rich auxiliary signals can stabilize prediction, and as histories accumulate, the model can shift back toward behavioral evidence. We discuss limitations particularly the reliance on rich interaction data, and outline directions for testing on public datasets with sufficient depth or proxy signals, as well as for exploring stronger equity objectives and interpretable gating.

Keywords

Knowledge Tracing; Temporal Equity; Reinforcement Learning; Conversational Features; LLM; DKT; SAKT.

1. INTRODUCTION

Knowledge Tracing (KT) has become a cornerstone of modern educational technology [10], especially in the development of intelligent tutoring systems [34]. KT models predict students' knowledge states based on their learning interactions [10]. Computer-supported education systems rely on KT to personalize learning experiences and optimize instructional decisions [17, 21]. However, recent research reveals a concerning pattern: existing KT models, including classical Bayesian Knowledge Tracing (BKT) and Deep Knowledge Tracing (DKT), often fail to achieve equitable tutoring across different student populations, exhibiting systematic bias against certain groups [13, 29]. While early-stage prediction methods have shown effectiveness in some domains via statistical approaches such as survival analysis [14], KT presents temporal fairness challenges that extend beyond traditional demographic disparities. A critical, under-explored dimension of fairness involves the differential treatment of early-stage learners with limited interaction histories versus experienced learners with extensive practice records. This temporal bias manifests as less accurate predictions for early-stage students due to sparse sequential data, leading to suboptimal recommendations during crucial initial phases; this phenomenon corresponds to the well-known *cold-start* problem in both knowledge tracing and recommender systems [7, 9]. Conversely, experienced learners benefit from rich histories that enable more precise knowledge estimation, creating a "rich-get-richer" effect in which students who need the most support receive the least accurate guidance.

To address temporal fairness in KT, we present a three–tier comparative framework evaluated on two standard backbones (DKT and SAKT). The Baseline relies on behavioral traces only (skill, correctness, optionally difficulty/response time). The +Features model augments the baseline with time–aligned conversation–derived signals (dialogue–rubric vectors and unit–level mastery indicators) designed to avoid temporal leakage. The +Features+RL model adds a gating mechanism trained with policy–gradient updates to adaptively weight behavioral versus conversational streams across a learner's trajectory. We operationalize temporal equity via *percent–progress bins* (10%, 20%, . . . , 100%), explicitly targeting more reliable predictions for early–stage learners while preserving overall performance [5, 10, 20, 21, 26]. This framing is especially relevant as tutoring systems become more conversational. In newer AI-mediated environments, dialogue is part of the instructional process itself, which makes dialogue-integrated KT increasingly practical even if it remains inapplicable to many legacy benchmark datasets [2, 11, 15].

Our contributions are threefold. (1) We formulate a fairness–aware optimization for KT that treats disparity across progress bins as an explicit signal, enabling adaptive fusion of heterogeneous evidence through an RL gate. (2) We integrate conversation–derived features with behavioral traces using leak–conscious, step–aligned fusion, offering a

Fan Zhang, Rui Guo, and Hai Li. Fair When Sparse, Accurate When Rich: RL-Gated Conversational Fusion for Temporally Equitable Knowledge Tracing. In Anthony Botelho, Maria Mercedes T. Rodrigo, Adish Singla, Hiroaki Ogata, Hyojeong So, and Young Hoan Cho (eds.) Proceedings of the 19th International Conference on Educational Data Mining, Seoul, Republic of Korea, June, 2026, pp. 446–452. International Educational Data Mining Society (2026).

https://doi.org/10.5281/zenodo.21039877

practical path to richer KT without compromising temporal validity. (3) We provide a systematic comparison across backbones and variants and visualize progress–aligned performance with cohort–size overlays, demonstrating qualitatively flatter, more equitable curves for the RL–gated model while maintaining competitive aggregate accuracy and AUC. Detailed metrics and variant comparisons are reported in the Results section.

2. RELATED WORK

2.1 Knowledge Tracing

Classical KT models, most notably BKT, represent mastery as a latent state updated from response histories with parameters for initialization, learning, slip, and guess [10]. Subsequent extensions account for learner heterogeneity and temporal dynamics, including individualized BKT [31] and time-aware variants that model spacing and forgetting [22, 23]. Neural KT reframes KT as sequence modeling: DKT uses recurrent encoders to predict next-step correctness from interaction sequences [21], while attention-based architectures such as SAKT capture relevance among past items and skills through query–key attention [20, 33].

Two practical challenges motivate our work. First, modern KT often relies primarily on correctness patterns and student–item interactions, which can underutilize rich side information such as problem text, hints, or conversational evidence [1]. Second, directly injecting auxiliary signals risks temporal leakage if features are not carefully aligned to the prediction horizon (e.g., using future turns of a dialogue to predict the present step) [12, 24]. Our setting exposes both issues: we have post-problem dialogues that are potentially informative, but any benefit hinges on leak-conscious alignment and on a mechanism that can adapt how much the model should rely on these signals at different stages of a learner's trajectory. Moreover, KT systems face a form of the cold-start problem analogous to recommender systems—predictions are less reliable for students with sparse histories [7, 9]. We therefore complement standard overall metrics with a progress-aware analysis to make such temporal effects explicit.

2.2 Fairness in Educational AI and KT

A growing literature examines disparate outcomes in educational AI [8, 32]. In KT specifically, studies have surfaced systematic gaps across student subgroups and contexts, raising fairness concerns for data-driven personalization [13, 29]. Some work explores modeling choices aimed at mitigating disparities (e.g., time-augmented BKT) [6], while others document construct-related effects such as readability gaps in math KT or platform/socioeconomic biases in second-language KT [25, 27].

Our perspective focuses on temporal equity: comparable predictive performance for learners at early vs. late stages. In conventional training, aggregate objectives naturally overweight data-rich trajectories; learners with short histories precisely those who need the most reliable guidance—can receive noisier predictions. We operationalize this dimension by grouping students into deciles based on total practice volume (10%–100%) and reporting performance by cohort. This differs from demographic fairness (which we do not analyze here) but is practically salient for tutorial decision making: reliable early-stage estimation can prevent compounding errors, while later stages may benefit more from accumulated behavioral evidence. Our equity lens thus aims at flattening cohort-wise performance curves, not enforcing strict parity at the expense of accuracy, and it aligns with the cold-start framing noted above [7, 9].

2.3 Reinforcement Learning in Educational AI

Reinforcement learning (RL) has been widely studied for sequencing and recommendation problems in education, treating personalization as a sequential decision process [3, 19]. Applications include curriculum sequencing and learning-path optimization as well as broader decision-making frameworks [4, 18, 28]. Much of this work, however, optimizes a single objective (e.g., learning gain, engagement) and gives limited attention to fairness-aware architectural adaptation.

In contrast, we use RL not to select content but to adapt model fusion. Concretely, a lightweight gate produces a convex weight that fuses behavioral and conversation/unit streams, with a policy-gradient objective [26] that balances predictive utility and a temporal-equity penalty defined over practice-volume cohorts. This framing has three practical advantages. First, it is model-agnostic: the gate can sit atop standard KT backbones (e.g., DKT, SAKT) without architectural overhaul. Second, it directly targets equity at the level of the evaluation protocol (deciles by practice volume), avoiding a mismatch between training goals and assessment. Third, it supports history-adaptive reliance on auxiliary signals: conversational features can carry more weight where behavioral evidence is scarce and recede when long histories suffice. Our empirical results (Section Results) suggest that such equity-aware fusion can reduce across-cohort variance while maintaining overall accuracy, providing a cautious but actionable path toward temporally robust KT.

3. METHOD

3.1 Data Description

We analyze logs from a deployed middle–school mathematics platform in which each problem–solving attempt may be followed by an optional chatbot conversation. The corpus spans approximately one academic term (about eight months) and contains 2,678 students, 1,605 items (three difficulty bands), and 10,877 problem–solving instances linked to up to 31 skills/competencies, with an overall correctness of about 47%. Each step records an item/skill identifier, binary correctness, and elapsed time; when present, the post–problem dialogue consists of multiple turns between the student and the chatbot (on average, about 1.47 post–problem conversations per attempt and about 7.97 turns). These characteristics provide the empirical context and are consistent with prior descriptions of the same platform.

The data analyzed in this study were derived from the ALTER-Math project [30]. The underlying human-subjects study was reviewed and approved by the Institutional Review Board (IRB) at the University of Florida (IRB202301838). All research procedures involving human participants in the source study were conducted in accordance with the ethical standards of the University, and informed consent was obtained from all participants prior to their inclusion in the study.

For the present analysis, only de-identified learner log data and post-problem dialogue data were used.

To assess how performance varies with the amount of prior history, we group *students* by their total number of attempted problems and form ten learner cohorts at the deciles: 10%, 20%, . . . , 100%. Unless otherwise noted, we report Accuracy (ACC) and AUC *within each cohort* and overall. This cohorting emphasizes reliability for learners with *sparse* vs. *rich* practice histories and is used consistently across all experiments.

3.2 LLM–Derived Dialogue Features

Beyond behavioral traces, we incorporate features distilled from the post–problem conversation. Feature extraction is conducted *offline* using an instruction–tuned LLaMA–family model, Llama-3-8B-Instruct, with decoding parameters temperature = 0, top-p = 0.9, and max_tokens = 128, running on a single A100–80GB GPU [16]. The resulting numeric features are then *time–aligned* to the KT sequence so that features used at step t reflect only information available up to (and including) t, preventing temporal leakage. We use two groups of dialogue features:

1. Rubric vector (fixed–length): aggregated indicators that summarize the conversation signals most relevant to KT—e.g., knowledge–point mentions/coverage, clarification/help–seeking, confusion/uncertainty cues, correctness–related evidence, and engagement tone—computed per solved item from the dialogue that follows that item.
2. Unit–level mastery (scalar per step): an LLM–based verdict on whether the learner evidences mastery of the curricular unit associated with the current item.

Both groups are exported as numeric features and concatenated as an auxiliary stream; missing dialogues at a step are zero–imputed for the rubric vector with a neutral default for mastery.

3.3 Models

We instantiate two standard knowledge–tracing backbones and build three variants on each:

DKT encodes the interaction sequence with a recurrent network and outputs a behavioral logit $z_t^{(\text{beh})}$ for step t [21]. SAKT aggregates relevant past items via self–attention to produce $z_t^{(\text{beh})}$ [20].

1. **Baseline (behavioral only)** uses the behavioral stream (skill×correctness) with available covariates (difficulty, response time).
2. **+Features** concatenates the *rubric vector* and the unit-mastery scalar as an auxiliary stream $X_t^{(\text{conv})}$ that is strictly step–aligned (forward–only).
3. **+Features+RL** adds a lightweight *gating* mechanism that produces a convex weight $\alpha_t \in [0, 1]$ each step and fuses streams:

$$z_t \;=\; \alpha_t \, z_t^{(\text{beh})} \;+\; \big(1-\alpha_t\big)\, z_t^{(\text{conv})}, \qquad \hat{p}_t \;=\; \sigma(z_t).$$

3.4 Fairness–Aware Reinforcement–Learning Gate

The gate is trained with policy–gradient updates [26] to preserve overall predictive utility while discouraging steep disparities across learner cohorts (10–100% by practice volume). Concretely, we maximize the objective

$$\begin{aligned}\mathcal{J} = \;& \underbrace{\mathbb{E}\big[\mathrm{Perf}(\hat{p}_t, r_t)\big]}_{\text{overall utility}} \\ & - \lambda_{\text{fair}} \cdot \underbrace{\Phi\big(\{\mathrm{Perf}_c\}_{c=1}^{10}\big)}_{\text{temporal–equity penalty}} \\ & + \lambda_{\text{ent}} \cdot \underbrace{\mathbb{H}[\alpha_t]}_{\text{entropy}} \,.\end{aligned}$$

where Perf denotes a differentiable surrogate aligned with ACC/AUC, Φ penalizes large performance gaps/variance across the ten practice–volume cohorts, and $\mathbb{H}$ regularizes the gate to avoid premature collapse. Intuitively, the gate can rely more on conversational evidence for learners with short histories and gradually shift toward behavioral evidence as history accumulates, aiming for flatter, more equitable cohort curves without sacrificing aggregate accuracy.

3.5 Training and Evaluation

All variants minimize masked cross–entropy over valid steps. Unless otherwise noted, we keep the same data splits, batch size, and training epochs across backbones/variants to enable controlled comparison.[1] We report Accuracy (ACC) and AUC *within each learner cohort* (10–100% by practice volume) and overall. Plots overlay a dashed right axis indicating the number of students per cohort to contextualize variance across groups with fewer vs. more interactions. Results are reported for DKT and SAKT under identical cohort definitions and evaluation code.

4. RESULTS

4.1 Setup for Temporal-Equity Evaluation

All models are evaluated by percent progress (10%, 20%, . . . , 100%) computed per learner from the normalized position t/T. For each bin and overall we report Accuracy (ACC) and AUC. Plots include a dashed right axis showing cohort size per bin (students when available; otherwise interactions), enabling visual inspection of stability vs. noise across the trajectory.

4.2 Aggregate Trends Across Backbones

Figures 1 a-d show per-bin AUC/ACC for DKT and SAKT. In both backbones, +Features+RL yields the strongest curves overall and particularly in early-to-mid progress bins where histories are short. The dashed right axes indicate that bins with fewer data points (or fewer students) have more volatile performance; even there, +Features+RL maintains a consistent advantage.

Table 1 summarizes percent-bin statistics (mean, SD, min, max, median) by model/variant. On DKT, +Features+RL improves mean AUC from 0.5985 (baseline) to 0.6094 and mean ACC from 0.5738 to 0.5877. On SAKT, +Features+RL

---

[1] In our runs we used a batch size of 64 and trained for 12 epochs per setting.

Table 1: Percent-progress bin statistics (mean ± SD, with min/max and median) by backbone and variant.

Backbone	Variant	ACC_{mean}	ACC_{SD}	ACC_{min}	ACC_{max}	ACC_{med}
DKT	baseline	0.5738	0.0253	0.5184	0.6064	0.5777
DKT	+features	0.5777	0.0242	0.5285	0.6165	0.5749
DKT	+features+RL	*0.5877*	*0.0205*	*0.5609*	*0.6212*	*0.5877*
SAKT	baseline	0.5596	0.0321	0.4887	0.6013	0.5715
SAKT	+features	0.5677	0.0291	0.5000	0.6005	0.5744
SAKT	+features+RL	*0.5872*	*0.0231*	*0.5527*	*0.6171*	*0.5880*

Backbone	Variant	AUC_{mean}	AUC_{SD}	AUC_{min}	AUC_{max}	AUC_{med}
DKT	baseline	0.5985	0.0320	0.5398	0.6445	0.6031
DKT	+features	0.5921	0.0272	0.5444	0.6312	0.5930
DKT	+features+RL	*0.6094*	*0.0261*	*0.5664*	*0.6507*	*0.6059*
SAKT	baseline	0.5792	0.0396	0.4829	0.6155	0.5969
SAKT	+features	0.5754	0.0315	0.5159	0.6113	0.5900
SAKT	+features+RL	*0.6027*	*0.0225*	*0.5666*	*0.6349*	*0.6066*

Notes. (i) Statistics are computed over ten student-level practice-volume cohorts (deciles: 10%, 20%, . . . , 100%). For each cohort we evaluate ACC and AUC, then summarize across cohorts (mean, SD, min, max, median). (ii) *Underlined italics* mark the best value within each row group (per backbone across variants): for *means/medians/max*, higher is better; for *SD*, lower is better (flatter, more equitable across cohorts); for *min*, higher indicates stronger worst-cohort performance. (iii) The `+features+RL` variant consistently achieves the highest means and lowest SDs across both backbones, suggesting the RL gate boosts overall quality while reducing dispersion across practice-volume cohorts. (iv) Min/max arise from different cohorts (early vs. late); interpret alongside SD and cohort sizes (shown in the figures' dashed right axes) to understand variability under sparse vs. rich histories.

raises mean AUC from 0.5792 (baseline) to 0.6027 and mean ACC from 0.5596 to 0.5872. Notably, +Features (without RL) shows mixed effects on AUC (slightly below baseline on both backbones), indicating that naive fusion of conversational/unit features can be suboptimal; the RL gate appears necessary to realize consistent gains.

4.3 Variance Reduction and Curve Flattening

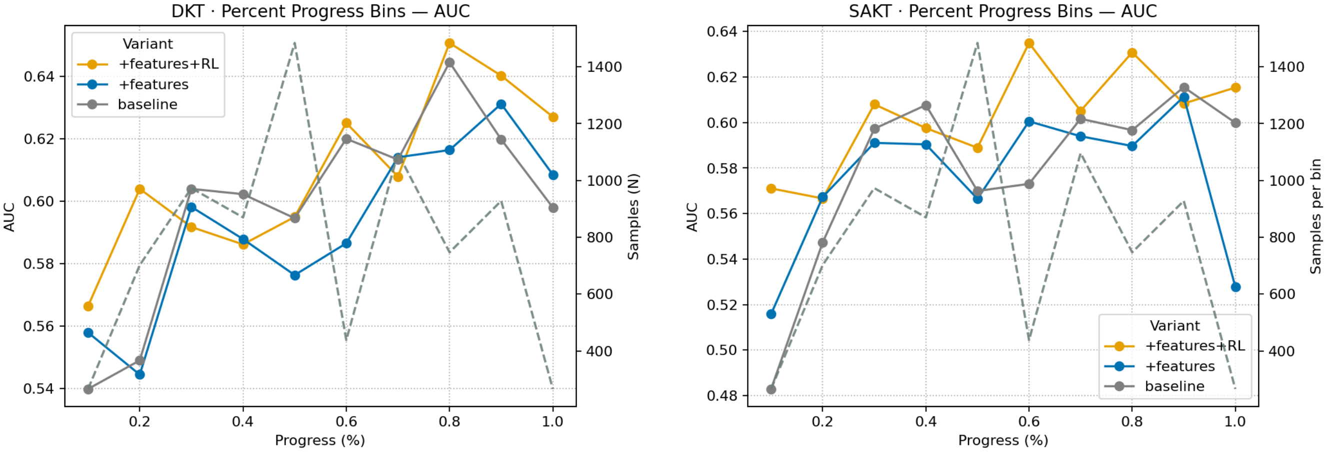
Beyond mean improvements, +Features+RL consistently reduces across-bin variance, suggesting flatter, more equitable trajectories. For DKT, ACC SD decreases from 0.0253 (baseline) to 0.0205 (+Features+RL), and AUC SD from 0.0320 to 0.0261. For SAKT, ACC SD drops from 0.0321 to 0.0231, and AUC SD from 0.0396 to 0.0225. This aligns with the design goal: attenuate the early/late performance gap by adapting the reliance on conversational vs. behavioral evidence over time.

5. CONCLUSION AND DISCUSSION

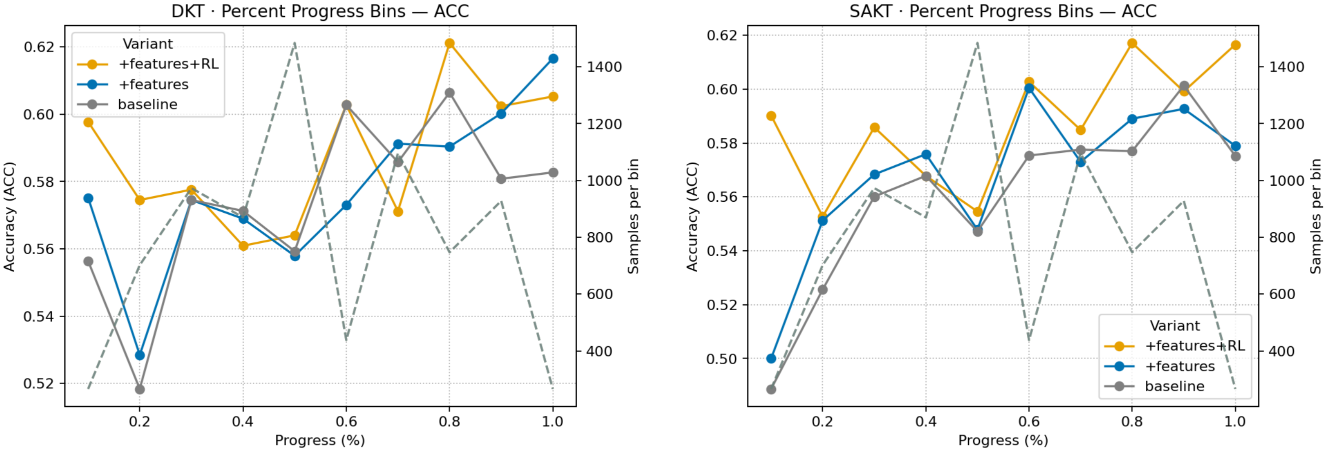
This work examined temporal equity in KT by augmenting behavioral inputs with dialogue-derived features and by learning an adaptive fusion policy via a lightweight RL gate. Under two standard backbones (DKT, SAKT) and a cohorting scheme that groups students by practice volume (10–100% deciles), the RL-gated variant consistently yielded higher aggregate performance (ACC/AUC) and lower dispersion across cohorts, producing flatter cohort-wise curves. Taken together, these results suggest a practical route to reduce performance variability between sparse- and rich-history learners without discarding established KT architectures. At the same time, these findings should be interpreted in relation to prior work on fairness in KT rather than as a standalone fairness claim: prior studies have shown that KT systems can yield inequitable performance across learners and contexts [1], and recent work has explored mitigating such disparities by modifying temporal assumptions within the KT model itself, such as time-augmented BKT [6]. Our results extend this line of work from a different angle. Rather than redesigning the KT backbone itself, we show that a lightweight prediction-stage fusion mechanism can improve reliability when learner histories are sparse while remaining compatible with established KT architectures.

The immediate broader impact of this architecture is therefore not that every legacy ITS can adopt post-problem dialogue features tomorrow, but that it provides a deployable fusion template for settings in which learning evidence extends beyond correctness sequences alone. This is consistent with broader KT research showing that student–item correctness traces often underutilize richer side information such as hints, problem text, or conversational evidence [1]. From a fairness perspective, this also matters because more information does not automatically translate into more equitable modeling; the educational AI literature has repeatedly emphasized that fairness depends on how evidence is represented and used, not merely on whether additional signals are available [5]. In that sense, the present architecture should be read as a practical mechanism for selectively using auxiliary evidence under conditions of evidence scarcity, rather than as a universal retrofit for all KT deployments.

Beyond aggregate metrics, the cohort view offered diagnostic value. We observed that adding dialogue features naïvely (without gating) did not uniformly help—in some cases AUC was unchanged or slightly reduced—indicating that conversational evidence is heterogeneous and history-dependent. This point is important in light of prior fairness work: just as inequities in KT do not disappear automatically through



(a) DKT — AUC (cohort size dashed on right axis) (b) SAKT — AUC (cohort size dashed on right axis)



(c) DKT — ACC (cohort size dashed on right axis) (d) SAKT — ACC (cohort size dashed on right axis)

Figure 1: Percent–progress evaluation across learner cohorts formed by student-level practice volume (deciles: 10%, 20%, . . . , 100%). X-axis: cohort by practice volume (10% → 100%). Left Y-axis: performance metric (AUC in panels (a,b); ACC in panels (c,d); range 0–1). Right Y-axis (dashed line): the number of valid prediction samples per bin. (Number of students). Curves compare Baseline, +Features, and +Features+RL; reading them together with the dashed cohort-size trace clarifies reliability under sparse (lower-decile) vs. rich (upper-decile) histories.

model complexity alone [13, 29], our results suggest that they also do not disappear simply by appending more features. Instead, auxiliary conversational evidence appears useful only when it is carefully aligned and selectively weighted. The RL gate appeared to address this by modulating reliance on conversational vs. behavioral signals along the trajectory; early cohorts (where histories are short) tended to benefit more from the additional stream, whereas later cohorts did not require as strong a contribution. In that sense, the equity-aware fusion objective functioned less like a global regularizer and more like a schedule for when and how to use each information source.

We view these findings as pointing to a possible path rather than a definitive solution. In particular, the approach remains intentionally simple: a standard backbone, a compact set of dialogue features (rubric + unit mastery), and a single gating mechanism with a fairness term. This simplicity made the behavior easier to reason about and the gains more attributable, but it also leaves room for complementary improvements (e.g., calibration, per-cohort operating points, or interpretable gating analyses). It also distinguishes our contribution from prior efforts that target fairness by modifying the KT model itself [6]. As such, our study is best interpreted as evidence that adaptive, equity-aware fusion is a viable design principle for KT, not as a claim that any particular architecture is optimal. Accordingly, the present contribution should be read less as a dataset-specific feature recipe and more as a transferable fusion principle for modern tutoring environments that record richer interaction traces than correctness sequences alone [1].

6. LIMITATIONS

First, our dataset contains substantial interaction volume per learner and includes post-problem dialogues from a deployed tutoring setting. This richness likely made it feasible to learn a useful fusion policy: the model could observe enough variation to adjust weights between behavioral and conversational signals. In contrast, many public KT datasets contain shorter sequences and no conversational channel, which limits how far our specific feature design and gating policy can be generalized as is. We therefore refrained from reporting results on standard public benchmarks that lack comparable signals, to avoid an apples-to-oranges comparison.

Second, our evaluation emphasizes temporal equity through practice-volume cohorts (student-level deciles). Although this targets a practically salient dimension of fairness (supporting learners with limited histories), it is only one slice. We did not analyze demographic fairness, course-level heterogeneity, or domain transfer, nor did we exhaust architectural variants of the gate (e.g., hierarchical or skill-conditional gating). The present results should therefore be read as context-constrained: they demonstrate that equity-aware fusion can stabilize cohort-wise performance on this dataset and setup, but external validity remains to be tested.

In the future, we plan to (i) explore public datasets with sufficient interaction depth or proxied contextual signals (e.g., textual rationales, hint requests) to stress-test the approach beyond our setting; (ii) examine calibration and thresholding per cohort to complement AUC/ACC; (iii) probe the learned gate (feature attributions, skill-conditional patterns) for interpretability; and (iv) evaluate alternative equity objectives (e.g., explicit worst-cohort optimization or variance-constrained training) and more expressive fusion architectures. These steps will clarify when equity-aware fusion provides the largest benefit and how it can be safely adopted in broader KT deployments.

References

[1] G. Abdelrahman, Q. Wang, and B. Nunes. Knowledge tracing: A survey. *ACM Computing Surveys*, 55(11):1–37, 2023.

[2] K. Academy. Meet khanmigo: Khan academy's ai-powered teaching assistant & tutor. *Khanmigo. ai*, 2024.

[3] M. M. Afsar, T. Crump, and B. Far. Reinforcement learning based recommender systems: A survey. *ACM Computing Surveys*, 55(7):1–38, 2022.

[4] S. Amin, M. I. Uddin, A. A. Alarood, W. K. Mashwani, A. Alzahrani, and A. O. Alzahrani. Smart e-learning framework for personalized adaptive learning and sequential path recommendations using reinforcement learning. *IEEE Access*, 11:89769–89790, 2023.

[5] R. S. Baker and A. Hawn. Algorithmic bias in education. *International journal of artificial intelligence in education*, 32(4):1052–1092, 2022.

[6] J. Barrett, A. Day, and K. Gal. Improving model fairness with time-augmented bayesian knowledge tracing. In *Proceedings of the 14th Learning Analytics and Knowledge Conference (LAK)*, pages 46–54, 2024.

[7] I. Bhattacharjee and C. Wayllace. Cold start problem: An experimental study of knowledge tracing models with new students. In *International Conference on Artificial Intelligence in Education*, pages 425–432, Cham, 2025. Springer Nature Switzerland.

[8] S. Bulathwela, M. Pérez-Ortiz, C. Holloway, M. Cukurova, and J. Shawe-Taylor. Artificial intelligence alone will not democratise education: On educational inequality, techno-solutionism and inclusive tools. *Sustainability*, 16(2):781, 2024.

[9] D. Cai, S. Qian, Q. Fang, J. Hu, and C. Xu. User cold-start recommendation via inductive heterogeneous graph neural network. *ACM Transactions on Information Systems*, 41(3):1–27, 2023.

[10] A. T. Corbett and J. R. Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4):253–278, 1994.

[11] M. Cukurova, M. Mavrikis, R. Luckin, J. Clark, and C. Crawford. Interaction analysis in online maths human tutoring: The case of third space learning. In *international conference on artificial intelligence in education*, pages 636–643. Springer, 2017.

[12] A. Darvishi, H. Khosravi, S. Sadiq, and D. Gašević. Incorporating ai and learning analytics to build trustworthy peer assessment systems. *British Journal of Educational Technology*, 53(4):844–875, 2022.

[13] S. Doroudi and E. Brunskill. Fairer but not fair enough: On the equitability of knowledge tracing. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK)*, pages 335–339, 2019.

[14] M. J. Fard, P. Wang, S. Chawla, and C. K. Reddy. A bayesian perspective on early stage event prediction in longitudinal data. *IEEE Transactions on Knowledge and Data Engineering*, 28(12):3126–3139, 2016.

[15] A. Ganguly, N. Mehjabin, A. Malik, and A. Johri. Conversational ai agents in education: An umbrella review of current utilization, challenges, and future directions for ethical and responsible use. *AI and Ethics*, 6(1):72, 2026.

[16] A. Grattafiori, A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Vaughan, et al. The llama 3 herd of models, 2024.

[17] M. Khajah, R. Wing, R. V. Lindsey, and M. C. Mozer. Integrating latent-factor and knowledge-tracing models to predict individual differences in learning. In *Proceedings of the International Conference on Educational Data Mining (EDM)*, pages 99–106, 2014.

[18] P. Kulkarni. *Reinforcement and Systemic Machine Learning for Decision Making*. John Wiley & Sons, 2012.

[19] B. Memarian and T. Doleck. A scoping review of reinforcement learning in education. *Computers and Education Open*, 6:100175, 2024.

[20] S. Pandey and G. Karypis. A self-attentive model for knowledge tracing. In *12th International Conference on Educational Data Mining, EDM 2019*, pages 384–389. International Educational Data Mining Society, 2019.

[21] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, and J. Sohl-Dickstein. Deep knowledge tracing. *Advances in neural information processing systems*, 28, 2015.

[22] Y. Qiu, Y. Qi, H. Lu, Z. A. Pardos, and N. T. Heffernan. Does time matter? modeling the effect of time with bayesian knowledge tracing. In *Proceedings of the International Conference on Educational Data Mining (EDM)*, pages 139–148, 2011.

[23] I. Šarić-Grgić, A. Grubišić, and A. Gašpar. Twenty-five years of bayesian knowledge tracing: A systematic review. *User Modeling and User-Adapted Interaction*, 34(4):1127–1173, 2024.

[24] A. Shibani, S. Knight, and S. Buckingham Shum. Contextualizable learning analytics design: A generic model and writing analytics evaluations. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK)*, pages 210–219, 2019.

[25] F. Stinar, H. Lee, C. Belitz, N. Nasiar, S. Fancsali, S. Ritter, H. Almoubayyed, R. Baker, J. Ocumpaugh, and N. Bosch. Fairness of bayesian knowledge tracing for math learners of different reading ability. In *Proceedings of the 18th International Conference on Educational Data Mining (EDM)*, pages 170–181, 2025.

[26] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2nd edition, 2018.

[27] W. Tang, G. Chen, S. Zu, and J. Luo. Fair knowledge tracing in second language acquisition. *arXiv preprint arXiv:2412.18048*, 2024.

[28] X. Tang, Y. Chen, X. Li, J. Liu, and Z. Ying. A reinforcement learning approach to personalized learning recommendation systems. *British Journal of Mathematical and Statistical Psychology*, 72(1):108–135, 2019.

[29] S. Tschiatschek, M. Knobelsdorf, and A. Singla. Equity and fairness of bayesian knowledge tracing. *arXiv preprint arXiv:2205.02333*, 2022.

[30] W. Xing, Y. Song, C. Li, Z. Liu, W. Zhu, and H. Oh. Development of a generative ai-powered teachable agent for middle school mathematics learning: A design-based research study. *British Journal of Educational Technology*, 56(5):2043–2077, 2025.

[31] M. V. Yudelson, K. R. Koedinger, and G. J. Gordon. Individualized bayesian knowledge tracing models. In *International Conference on Artificial Intelligence in Education (AIED)*, pages 171–180. Springer, 2013.

[32] F. Zhang, W. Xing, C. Li, and Y. Jiang. Fair ai in educational predictions: A multi-group fairness approach using reinforcement learning. *The Internet and Higher Education*, page 101074, 2026.

[33] X. Zhang, J. Zhang, N. Lin, and X. Yang. Sequential self-attentive model for knowledge tracing. In *International Conference on Artificial Neural Networks*, pages 318–330. Springer, 2021.

[34] Y. Zhou, Z. Lv, S. Zhang, and J. Chen. Disentangled knowledge tracing for alleviating cognitive bias. In *Proceedings of the ACM Web Conference*, pages 2633–2645, 2025.