# Master Internship Proposal

## General information

**Key-terms:** Machine Learning; Knowledge Tracing ; Education ; large-scale public datasets.
**Title**: Sequential and cognitive state modeling for classroom Knowledge Tracing using public Mathematics education datasets
**Work environment**: UCBL Lyon1, LIRIS laboratory, UMR 5205 CNRS, Lyon, France
**Duration and period**: 5 months (M2, Bac+5 from April to August 2026: ) or 3 months (M1, Bac+4 from June to August 2026).
**Supervisors**: Dr M. Essaid KHANOUCHE
**Application**: Send **CV** and **academic transcripts** of 2024/25 and 2025-26 to mohamed-essaid.khanouche@univ-lyon1.fr

## Description

Knowledge Tracing (KT) aims to predict a student's response to future questions using historical interaction sequences. Early probabilistic formulations relied on fixed transition assumptions, while neural models such as Deep Knowledge Tracing (DKT) and later attention-based approaches inferred sequential dependencies directly from correctness data [1]. However, subsequent analyses have demonstrated that many modern architectures still focus on correlation patterns rather than explicitly modeling the learner's internal state [2], [3].

This internship focuses on large-scale, publicly available offline datasets, with a primary emphasis on mathematics benchmarks such as ASSISTments2017 and Junyi2019. These datasets are widely used in recent benchmarking studies and provide a common evaluation ground for sequential modeling approaches [1], [4], [5]. The objective is to design a sequential or mathematical knowledge tracing approach that outperforms strong baselines on these datasets using Accuracy and Area Under the Curve as evaluation metrics.

**State-of-the-Art Context:**

Several recent works integrate item difficulty as a conditioning factor within attention or state update mechanisms. These approaches show that difficulty-aware modeling improves response prediction by aligning item challenge with learner proficiency dynamics [6], [7]. Other studies introduce explicit forgetting mechanisms through temporal decay or memory attenuation functions, confirming the importance of time-dependent effects in knowledge evolution [8]. However, these mechanisms have been extensively explored and show diminishing marginal gains when treated as the primary modeling contribution [8].

**Proposed Approach:**

Recent research reframes knowledge tracing as a latent state estimation problem rather than a simple sequence labeling task. State-aware and interpretable models demonstrate that explicit internal state representations provide clearer learning dynamics and improved predictive behavior [5], [9], [10]. Parallel work introduces uncertainty-aware transitions and cognitive fluctuation modeling, suggesting that learner responses depend on internal cognitive conditions rather than accumulated correctness alone [11], [12].

This internship builds on these findings by exploring novel student state evolution mechanisms inspired by cognitive and biologically motivated learning processes.

- **Item Difficulty:** Retained as an interaction factor that modulates state transitions [6].
- **Forgetting:** Incorporated as a baseline component rather than the central source of novelty [8].
- **Novelty:** The core research question examines whether alternative state transition formulations, grounded in cognitive or biological learning principles (e.g., reinforcement, consolidation), can improve response prediction beyond existing approaches [13].

The internship targets the design of a sequential architecture or mathematical formulation that integrates these state dynamics within recurrent or attention-based frameworks.

## Planning

The internship follows a structured research plan divided into four phases:

- **Phase 1: Literature Review**
  Review recent work on sequential, state-aware, and cognitively motivated knowledge tracing models published between 2021 and 2025. This review analyzes attention-based, interpretable, and uncertainty-aware formulations to identify limitations in existing student state update rules [2], [5], [9], [11].
- **Phase 2: Dataset Preparation**
  Preprocess ASSISTments2017 and Junyi2019 into standardized sequential formats used in prior studies [1], [4], [6]. This stage includes computing item difficulty estimates from population-level correctness statistics following established methodologies [6], [7].
- **Phase 3: Model Development**
  Implement strong baseline architectures to establish reference performance [1], [2], [5]. Subsequently, develop the novel student state transition mechanism inspired by cognitive or biological learning processes, informed by uncertainty-aware and cognitive fluctuation studies [11], [12], [13].
- **Phase 4: Evaluation and Ablation**

  Evaluate the proposed approach using Accuracy and AUC. This phase compares the proposed model against baseline methods and conducts targeted ablation experiments to isolate the contribution of the proposed state transition formulation.

## Required Profile

The position targets a Master's or Engineering student in Computer Science, Data Science, or Artificial Intelligence.

- **Technical Skills:** Proficiency in PyTorch or TensorFlow; familiarity with large-scale sequential datasets using pandas and numpy.
- **Scientific Background:** Experience with sequential modeling (Recurrent Networks, Transformers, etc.) and the ability to interpret mathematical formulations and reason about model behavior during training.

## References

*[1] Z. Liu et al., "Revisiting knowledge tracing: A simple and powerful model," Proc. ACM Multimedia, 2024.*

*[2] Y. Zhou et al., "Tracing knowledge instead of patterns: Stable knowledge tracing with diagnostic transformer," Proc. ACM Web Conf., 2023.*

*[3] S. Shen et al., "Learning process-consistent knowledge tracing," Proc. ACM SIGKDD, 2021.*

*[4] Z. Liu et al., "Enhancing deep knowledge tracing with auxiliary tasks," Proc. ACM Web Conf., 2023.*

*[5] Z. Liu and W. Luo, "Interpretable knowledge tracing with multiscale state representation," Proc. ACM Web Conf., 2024.*

*[6] Z. Liu et al., "Assessing student dynamic knowledge state by exploring the question difficulty effect," Proc. ACM SIGIR, 2022.*

*[7] S. Shen et al., "MLC-DKT: A multi-layer context-aware deep knowledge tracing model," Knowledge-Based Systems, vol. 303, 2024.*

*[8] Y. Bai et al., "Rethinking and improving student learning and forgetting processes for attention-based knowledge tracing models," Proc. AAAI, 2025.*

*[9] X. Li et al., "Extending context window of attention-based knowledge tracing models via length extrapolation," Proc. ECAI, 2024.*

*[10] S. Huang et al., "Towards robust knowledge tracing models via k-sparse attention," Proc. ACM SIGIR, 2023.*

*[11] W. Cheng et al., "Uncertainty-aware knowledge tracing," Proc. AAAI, 2025.*

*[12] M. Hou et al., "Cognitive fluctuations enhanced attention network for knowledge tracing," Proc. AAAI, 2025.*

*[13] S. Shen et al., "Tracing knowledge state with individual cognition and acquisition estimation," Proc. ACM SIGIR, 2021.*