



## Master's Thesis Proposal

**Topic:** Mitigating Temporal Bias in Pretrained Language Models via Fine-Tuning Strategies

**Supervisor:** ZeinabSadat Taghavi

**Examiner:** Hinrich Schutze

**Level:** MSc

**Summary:** Introduction

LLMs exhibit temporal bias due to imbalanced pretraining data, leading to poor performance on historical/future time periods. This thesis designs and evaluates fine-tuning strategies to reduce this bias, enabling robust temporal reasoning across diverse timeframes.

### Research Questions:

1. Can task-specific fine-tuning (e.g., temporal span extraction) improve LLMs' temporal reasoning accuracy on underrepresented time periods?
2. How do adversarial training or data augmentation strategies affect model robustness to temporal distribution shifts?

### Methodology:

- Develop a fine-tuning framework using temporal tasks (e.g., date-infilled QA, event ordering) and datasets like COMPLEXTEMPQA.
- Experiment with few-shot learning, adversarial training, and synthetic data augmentation for temporal adaptation.
- Compare model performance against baselines using robustness metrics (e.g., accuracy drop across time periods).

### Expected Contribution:

A novel fine-tuning framework to mitigate temporal bias, with empirical validation on downstream tasks and analysis of generalization limits.

**Requirements:** Huggingface, PyTorch, advanced NLP/Transformer knowledge, experience with model fine-tuning.

- Qingyu Tan, Hwee Tou Ng, and Lidong Bing (2023). *Towards Benchmarking and Improving the Temporal Reasoning Capability of Large Language Models*. DOI: 10.48550/ARXIV.2306.08952. URL: <https://arxiv.org/abs/2306.08952>
- Raphael Gruber et al. (2024). *ComplexTempQA: A Large-Scale Dataset for Complex Temporal Question Answering*. DOI: 10.48550/ARXIV.2406.04866. URL: <https://arxiv.org/abs/2406.04866>