



Bachelor's Thesis Proposal

Topic: Analyzing the Impact of Date Formats on Temporal Reasoning in Language Models

Supervisor: ZeinabSadat Taghavi

Examiner: Hinrich Schutze

Level: BSc

Summary: Introduction

Pretrained language models (LLMs) like GPT and T5 struggle with temporal reasoning tasks, particularly when handling time-dependent facts. This thesis investigates whether the format of dates (e.g., "January 2020" vs. "2020-01") influences model performance.

Research Questions:

1. How do different date formats affect LLMs' accuracy on temporal reasoning benchmarks (e.g., TEMPReason)?
2. Is there a correlation between pretraining data token distributions (e.g., frequency of historical vs. modern dates) and model performance?

Methodology:

- Evaluate baseline performance of LLMs (e.g., T5, GPT-2) on TEMPReason with controlled date-format variations.
- Statistical comparison of model accuracy across date formats and time periods.

Expected Contribution:

A systematic analysis of date formatting's impact on temporal reasoning, highlighting limitations of LLMs in handling time-sensitive information.

Requirements: Huggingface, familiarity with Transformer models, basic Python/data analysis.

- 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>