Synthetic Data Quality Effects on Small Language Model Inference Efficiency

Assignee Research

June 1, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: What is the impact of synthetic data source quality on the inference efficiency and throughput of small language models trained for natural language inference tasks. The evolution of Generative Pre-trained Transformer (GPT) models has led to significant advancements in various natural language processing applications, particularly in legal textual entailment. We present an analysis of GPT-3.5 (ChatGPT) and GPT-4 performances on COLIEE Task 4. 7 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Black-Box Analysis: GPTs Across Time in Legal Textual Entailment Task. Research question: What is the impact of synthetic data source quality on the inference efficiency and throughput of small language models trained for natural language inference tasks?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.3/10.

3 Results

12 papers retrieved. 7 claims extracted; 1 independently verified. Quality review score: 4.3/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
The COLIEE Task 4 dataset contains questions along with their associated relevant articles from Japanese statute law spa	✓	0.17
The number of questions in the COLIEE Task 4 dataset increases in recent years.	×	0.04
GPT-3.5 (English) achieved an accuracy of 0.7222 in the year H18 for the COLIEE Task 4 dataset.	×	0.06
GPT-4 (English) achieved an accuracy of 0.6944 in the year H18 for the COLIEE Task 4 dataset.	×	0.06
GPT-3.5 (Japanese) achieved an accuracy of 0.6667 in the year H18 for the COLIEE Task 4 dataset.	×	0.07
GPT-4 (Japanese) achieved an accuracy of 0.6944 in the year H18 for the COLIEE Task 4 dataset.	×	0.07
GPT-4 has demonstrated exceptional performance in passing the Uniform Bar Examination (UBE), achieving a score that surp	×	0.04

References

- http://arxiv.org/abs/2010.03813v2
- http://arxiv.org/abs/2411.11001v1
- http://arxiv.org/abs/2309.05501v1