Cross-Domain Fine-Tuning of Baichuan-2: Legal vs. Biomedical FactCC and TruthfulQA Performance

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: How does cross-domain fine-tuning of Baichuan-2 on legal vs. biomedical text compare in terms of FactCC benchmark scores when evaluated with varying TruthfulQA misalignment thresholds. In the era of digital healthcare, the huge volumes of textual information generated every day in hospitals constitute an essential but underused asset that could be exploited with task-specific, fine-tuned biomedical language representation models, improving patient care and. 15 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Localising In-Domain Adaptation of Transformer-Based Biomedical Language Models. Research question: How does cross-domain fine-tuning of Baichuan-2 on legal vs. biomedical text compare in terms of FactCC benchmark scores when evaluated with varying TruthfulQA misalignment thresholds?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.7/10.

3 Results

14 papers retrieved. 15 claims extracted; 0 independently verified. Quality review score: 3.7/10.

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 BaseBIT model was pretrained on various Italian corpora of generic text, including a recent Wikipedia dump and vario	×	0.03
The BioBIT model was derived from machine-translated biomedical abstracts using Google's neural machine translation (NMT	×	0.07
The MedBIT model was obtained using a corpus of selected medical texts natively written in Italian.	×	0.09
The model was initialized with a monolingual Italian version of BERT, obtained from a recent Wikipedia dump and various	×	0.03
The BaseBIT model was used as the baseline for the study.	×	0.03
Over 12 thousand BERT-based models are hosted in the Huggingface model repository, covering more than 20 different non-E	×	0.04
The BioBIT model was trained using machine-translated PubMed abstracts.	×	0.04
Google's neural machine translation (NMT) system was used to translate PubMed abstracts into Italian.	×	0.06
The NMT system used is a combination of transformers and recurrent neural networks (RNNs).	×	0.04
The NMT system has been shown to work well in clinical settings, such as for translating abstractions of clinical trials	×	0.06
The overall process involves starting from a general-purpose, Italian checkpoint and deriving several biomedical adaptat	×	0.07
Each model has been evaluated on MLM and the best-performing models on a battery of popular biomedical downstream tasks.	×	0.06
The pretraining objective is a combination of MLM (Masked Language Modeling) and next sentence prediction (NSP).	×	0.03
The MLM objective is based on randomly masking 15% of the input sequence and trying to predict the missing tokens.	×	0.02
For the NSP objective, the model is given a couple of sentences and has to guess if the second comes after the first in	×	0.03

References

- http://arxiv.org/abs/2212.10422v3
- http://arxiv.org/abs/2311.18580v2
- http://arxiv.org/abs/2310.05276v1