Full document analysis

Combined report

Document: sample-thesis-draft.pdf · **References:** 32 · **Sentences analyzed:** 164 · **Date:** 06 October 2026

45
/ 100
Document score

Needs work

Fix the flagged items before you submit

5 critical issues to fix · 5 to review · 11% of analyzed text flagged as AI

Citations

10 of 32 need attention

22 verified **3** outdated **2** unconfirmed **5** not found

AI writing

11% AI-generated

87% human **11%** AI

01 Citation findings

32 references checked against 8 academic databases

Critical: must fix before submitting (5)

#3 Critical

Grounded clinical answer synthesis with hierarchical evidence retrieval

H. Lindqvist et al. (2022) · Proceedings of ACL 2022

Not found. This reference could not be found in any of the 8 academic databases we searched.

What to do: Remove it or replace it with a real, verifiable source. Examiners flag fabricated citations.

Searched: arXiv, CrossRef, DBLP, Google Books, Open Library, OpenAlex, PubMed, Semantic Scholar. No match.

#11 Critical **Hallucination-aware decoding for medical question answering**

P. Marchetti et al. (2023) · Findings of EMNLP 2023

Not found. This reference could not be found in any of the 8 academic databases we searched.

What to do: Remove it or replace it with a real, verifiable source. Examiners flag fabricated citations.

Searched: arXiv, CrossRef, DBLP, Google Books, Open Library, OpenAlex, PubMed, Semantic Scholar. No match.

#17 Critical **ClinRetrieve: Dense retrieval for electronic health records at scale**

S. Varga et al. (2021) · Journal of Biomedical Informatics

Not found. This reference could not be found in any of the 8 academic databases we searched.

What to do: Remove it or replace it with a real, verifiable source. Examiners flag fabricated citations.

Searched: arXiv, CrossRef, DBLP, Google Books, Open Library, OpenAlex, PubMed, Semantic Scholar. No match.

#24 Critical **Calibrated uncertainty estimation for retrieval-augmented language models in healthcare**

A. Brennan et al. (2024) · Advances in Neural Information Processing Systems 37

Not found. This reference could not be found in any of the 8 academic databases we searched.

What to do: Remove it or replace it with a real, verifiable source. Examiners flag fabricated citations.

Searched: arXiv, CrossRef, DBLP, Google Books, Open Library, OpenAlex, PubMed, Semantic Scholar. No match.

#30 Critical **Evidence-weighted citation generation for clinical decision support**

M. Osei et al. (2023) · AMIA Annual Symposium Proceedings

Not found. This reference could not be found in any of the 8 academic databases we searched.

What to do: Remove it or replace it with a real, verifiable source. Examiners flag fabricated citations.

Searched: arXiv, CrossRef, DBLP, Google Books, Open Library, OpenAlex, PubMed, Semantic Scholar. No match.

Review: should be checked (5)

#6 Review **Training language models to follow instructions with human feedback**

Long Ouyang et al. (2022) · arXiv preprint arXiv:2203.02155

Outdated. This paper was cited as an arXiv preprint but has been published at Advances in Neural Information Processing Systems 35 (NeurIPS 2022).

What to do: Update the citation to the latest published version.

Closest match: Semantic Scholar

#9 Review **Exploring the limits of transfer learning with a unified text-to-text transformer**

Colin Raffel et al. (2019) · Proceedings of ICML 2019

Unconfirmed match. The closest database match is low-confidence and could not be confirmed.

What to do: Check the author, year and venue against the original source by hand.

#15 Review **REALM: Retrieval-augmented language model pre-training**

Kelvin Guu et al. (2020) · arXiv preprint arXiv:2002.08909

Outdated. This paper was cited as an arXiv preprint but has been published at the 37th International Conference on Machine Learning (ICML 2020).

What to do: Update the citation to the latest published version.

Closest match: arXiv

#21 Review **Dense passage retrieval for clinical question answering**

Y. Chen et al. (2021) · Proceedings of ACL 2021

Unconfirmed match. The closest database match is low-confidence and could not be confirmed.

What to do: Check the author, year and venue against the original source by hand.

#28 Review **REPLUG: Retrieval-augmented black-box language models**

Weijia Shi et al. (2023) · arXiv preprint arXiv:2301.12652

Outdated. This paper was cited as an arXiv preprint but has been published at the 2024 Conference of the North American Chapter of the ACL (NAACL 2024).

What to do: Update the citation to the latest published version.

Closest match: Semantic Scholar

Verified: matched a database record (22)

#	Reference	Source	Status
1	**Attention is all you need** Ashish Vaswani et al. (2017) · Advances in Neural Information Processing Systems 30	Semantic Scholar	Verified
2	**BERT: Pre-training of deep bidirectional transformers for language understanding** Jacob Devlin et al. (2019) · Proceedings of NAACL-HLT 2019	CrossRef	Verified
4	**Retrieval-augmented generation for knowledge-intensive NLP tasks** Patrick Lewis et al. (2020) · Advances in Neural Information Processing Systems 33	OpenAlex	Verified
5	**Dense passage retrieval for open-domain question answering** Vladimir Karpukhin et al. (2020) · Proceedings of EMNLP 2020	CrossRef	Verified
7	**Language models are few-shot learners** Tom Brown et al. (2020) · Advances in Neural Information Processing Systems 33	Semantic Scholar	Verified
8	**BioBERT: a pre-trained biomedical language representation model for biomedical text mining** Jinhyuk Lee et al. (2020) · Bioinformatics	PubMed	Verified
10	**What disease does this patient have? A large-scale open domain question answering dataset from medical exams** Di Jin et al. (2021) · Applied Sciences	OpenAlex	Verified
12	**PubMedQA: A dataset for biomedical research question answering** Qiao Jin et al. (2019) · Proceedings of EMNLP-IJCNLP 2019	CrossRef	Verified
13	**Large language models encode clinical knowledge** Karan Singhal et al. (2023) · Nature	CrossRef	Verified
14	**Survey of hallucination in natural language generation** Ziwei Ji et al. (2023) · ACM Computing Surveys	CrossRef	Verified
16	**The probabilistic relevance framework: BM25 and beyond** Stephen Robertson et al. (2009) · Foundations and Trends in Information Retrieval	CrossRef	Verified
18	**Enabling large language models to generate text with citations** Tianyu Gao et al. (2023) · Proceedings of EMNLP 2023	DBLP	Verified
19	**Self-RAG: Learning to retrieve, generate, and critique through self-reflection** Akari Asai et al. (2024) · International Conference on Learning Representations (ICLR 2024)	OpenAlex	Verified

20	**Leveraging passage retrieval with generative models for open domain question answering** Gautier Izacard et al. (2021) · Proceedings of EACL 2021	DBLP	Verified
22	**ROUGE: A package for automatic evaluation of summaries** Chin-Yew Lin (2004) · Text Summarization Branches Out	DBLP	Verified
23	**BLEU: a method for automatic evaluation of machine translation** Kishore Papineni et al. (2002) · Proceedings of ACL 2002	CrossRef	Verified
25	**LoRA: Low-rank adaptation of large language models** Edward J. Hu et al. (2022) · International Conference on Learning Representations (ICLR 2022)	OpenAlex	Verified
26	**Llama 2: Open foundation and fine-tuned chat models** Hugo Touvron et al. (2023) · arXiv preprint arXiv:2307.09288	arXiv	Verified
27	**Chain-of-thought prompting elicits reasoning in large language models** Jason Wei et al. (2022) · Advances in Neural Information Processing Systems 35	Semantic Scholar	Verified
29	**MIMIC-III, a freely accessible critical care database** Alistair E. W. Johnson et al. (2016) · Scientific Data	PubMed	Verified
31	**FActScore: Fine-grained atomic evaluation of factual precision in long form text generation** Sewon Min et al. (2023) · Proceedings of EMNLP 2023	OpenAlex	Verified
32	**BERTScore: Evaluating text generation with BERT** Tianyi Zhang et al. (2020) · International Conference on Learning Representations (ICLR 2020)	Semantic Scholar	Verified

02 AI writing detection

87% estimated human-written · 18 passages flagged as likely AI-generated.

11% of analyzed sentences read as AI-generated, 5% as mixed, 84% as human.

Flagged passages

> In today's rapidly evolving healthcare landscape, large language models have emerged as a transformative force with the potential to revolutionize clinical decision-making. **97%**

> By seamlessly integrating retrieval mechanisms with generative capabilities, retrieval-augmented generation offers a robust and scalable solution to the persistent challenge of hallucination. **95%**

> It is important to note that the reliability of generated answers is paramount in high-stakes medical

settings, where even minor inaccuracies can have significant consequences. **94%**

This multifaceted approach not only enhances factual accuracy but also fosters greater trust among clinicians and patients alike. **93%**

Furthermore, the proposed framework leverages state-of-the-art dense retrieval techniques to ensure that every answer is grounded in verifiable evidence. **92%**

Overall, these findings underscore the critical importance of evidence grounding in the development of trustworthy clinical AI systems. **91%**

In conclusion, this thesis paves the way for future research at the intersection of information retrieval, natural language processing and clinical medicine. **90%**

Moreover, the integration of domain-specific knowledge bases plays a pivotal role in bridging the gap between general-purpose language models and specialized medical expertise. **89%**

As the field continues to advance at an unprecedented pace, it is crucial to establish rigorous evaluation protocols that capture the nuances of clinical reasoning. **88%**

These results highlight the delicate balance between retrieval depth and answer conciseness that practitioners must navigate. **87%**

Notably, the model demonstrates a remarkable ability to synthesize information from multiple sources into coherent and contextually relevant responses. **86%**

Ultimately, the goal is to empower healthcare professionals with tools that augment, rather than replace, their clinical judgment. **85%**

This chapter delves into the intricacies of the retrieval pipeline, shedding light on the design choices that underpin its performance. **84%**

Taken together, the ablation studies provide compelling evidence that each component contributes meaningfully to the overall system. **83%**

In the realm of medical question answering, the stakes could not be higher. **82%**

Such advancements hold immense promise for improving patient outcomes and streamlining clinical workflows. **81%**

The comprehensive evaluation presented herein offers valuable insights for researchers and practitioners navigating this complex domain. **80%**

By fostering transparency through explicit citations, the system aligns with the broader push toward explainable and accountable artificial intelligence. **79%**

We retrieve the top-k passages from a PubMed index and pass them to the generator, which ensures a strong foundation for the final answer. **68%**

Compared to the BM25 baseline, dense retrieval improves recall@20 by a notable margin, underscoring its effectiveness in the clinical setting. **66%**

The fine-tuning stage uses LoRA adapters, which offer a lightweight yet powerful alternative to full fine-tuning. **64%**

Error analysis reveals that most failures stem from outdated guidelines in the retrieval corpus, a challenge that warrants further investigation. **62%**

Annotators rated each answer for faithfulness and completeness, providing a nuanced picture of model behaviour. **60%**

These limitations notwithstanding, the approach shows promise for deployment in real-world clinical environments. **58%**

We hypothesize that larger retrieval corpora would further enhance performance across all question categories. **57%**

This aligns with prior work suggesting that grounding significantly reduces unsupported claims. **55%**

How to read this. Scores estimate the likelihood that a passage was generated by an AI language model; they are probabilistic, not proof. Use them to locate sections worth rewriting in your own voice. Citation verification cross-checks every reference against eight academic databases and the cited paper's full text where available.

Verified against CrossRef, OpenAlex, Semantic Scholar, PubMed and more. This report reflects automated checks at the time of generation and is provided to help you review your work. Always confirm flagged items yourself. © Check Your Draft.