

Prompting Strategies for Maximizing Language Model Accuracy on Graduate-Level Science Questions

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: What prompting strategies maximize language model accuracy on graduate-level science questions v20. 10 claims were extracted from source literature; 5 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Practical and ethical challenges of large language models in education: A systematic scoping review. Research question: What prompting strategies maximize language model accuracy on graduate-level science questions v20.

2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.7/10.

3 Results

11 papers retrieved. 10 claims extracted; 5 independently verified. Quality review score: 6.7/10.

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 study conducted a systematic scoping review of 118 peer-reviewed papers published since 2017.	✓	0.26
The review identified 53 use cases for LLMs in automating education tasks.	✓	0.21
The identified use cases are categorized into nine main categories: profiling/labelling, detection, grading, teaching su	✓	0.31
The review identified low technological readiness as a practical or ethical challenge.	×	0.15
The review identified a lack of replicability and transparency as a practical or ethical challenge.	×	0.15
The review identified insufficient privacy and beneficence considerations as a practical or ethical challenge.	✓	0.17
The study produced three recommendations for future studies.	×	0.09
One recommendation is to update existing innovations with state-of-the-art models such as GPT-3 or GPT-4.	✓	0.19
One recommendation is to embrace the initiative of open-sourcing models or systems.	×	0.14
One recommendation is to adopt a human-centred approach throughout the development process.	×	0.09

References

- <https://doi.org/10.1007/s10956-024-10177-x>
- <https://doi.org/10.1111/bjet.13370>
- <https://doi.org/10.1007/s11704-024-40555-y>