Full document analysis

Combined report

Document: paper.pdf · **References:** 15 · **Sentences analyzed:** 68 · **Date:** 19 July 2026

5
/ 100
Document score

Needs work

Fix the flagged items before you submit

15 critical issues to fix · 13% of analyzed text flagged as AI

Citations

15 of 15 need attention

● **0** verified ● **0** outdated ● **15** not found

AI writing

13% AI-generated

● **59%** human ● **13%** AI

01 Citation findings

15 references checked against 8 academic databases

Critical: must fix before submitting (15)

#1 Critical **venues, unresolved identifiers, author-name variation, improbable report numbers, or mismatched metadata**

venues, unresolved identifiers, author-name variation, improbable report numbers, or mismatched metadata

Not found. Citation not found in any enabled database (arxiv, crossref, dblp, google_books, open_library, openalex, pubmed, semantic_scholar)

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

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

#2 Critical **Relevance failure in machine-generated evidence chains**

M Alberon et al. (2021) · Quarterly Review of Synthetic Scholarship

Not found. Citation not found in any enabled database (arxiv, crossref, dblp, google_books, open_library, openalex, pubmed, semantic_scholar)

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

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

#3 Critical

The credibility premium of DOI-shaped strings in automated literature reviews

P Denholm et al. (2024) · Journal of Computational Epistemics

Not found. Citation not found in any enabled database (arxiv, crossref, dblp, google_books, open_library, openalex, pubmed, semantic_scholar)

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

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

#4 Critical

Auditing citations produced by language systems: A practitioner handbook

L Harrowfield (2022) · Auditing citations produced by language systems: A practitioner handbook

Not found. Citation not found in any enabled database (arxiv, crossref, dblp, google_books, open_library, openalex, pubmed, semantic_scholar)

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

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

#5 Critical

Citation drift and the imitation of scholarly authority

A Kestrel et al. (2022) · Archives of Machine-Assisted Inquiry

Not found. Citation not found in any enabled database (arxiv, crossref, dblp, google_books, open_library, openalex, pubmed, semantic_scholar)

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

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

#6 Critical

Resolver confidence is not source confidence

J Merrin et al. (2023) · Proceedings of the 18th Symposium on Reference Integrity

Not found. Citation not found in any enabled database (arxiv, crossref, dblp, google_books, open_library, openalex, pubmed, semantic_scholar)

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

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

#7 Critical **Metadata recombination in generative research assistants**

P Norwick et al. (2023) · International Journal of Bibliographic Systems

Not found. Citation not found in any enabled database (arxiv, crossref, dblp, google_books, open_library, openalex, pubmed, semantic_scholar)

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

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

#8 Critical **Cascaded verification for high-volume reference screening**

B Osterley et al. (2022) · Transactions on Evidence Infrastructure

Not found. Citation not found in any enabled database (arxiv, crossref, dblp, google_books, open_library, openalex, pubmed, semantic_scholar)

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

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

#9 Critical **Four-stage validation of scholarly identifiers**

R Quellman et al. (2020) · Review of Digital Knowledge Assurance

Not found. Citation not found in any enabled database (arxiv, crossref, dblp, google_books, open_library, openalex, pubmed, semantic_scholar)

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

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

#10 Critical **Beyond resolution: Testing whether cited sources support generated claims**

Y Sato-Meridian (2024) · Journal of Evidence Alignment

Not found. Citation not found in any enabled database (arxiv, crossref, dblp, google_books, open_library, openalex, pubmed, semantic_scholar)

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

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

#11 Critical **The anatomy of a believable false reference**

C Vale et al. (2023) · Review of Synthetic Scholarship

Not found. Citation not found in any enabled database (arxiv, crossref, dblp, google_books, open_library, openalex, pubmed, semantic_scholar)

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

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

#12 Critical **Reference reliability in neural text generation**

H Wexler (2021) · Reference reliability in neural text generation

Not found. Citation not found in any enabled database (arxiv, crossref, dblp, google_books, open_library, openalex, pubmed, semantic_scholar)

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

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

#13 Critical **Traceable verification interfaces for machine-assisted scholarship**

E Yarrow et al. (2023) · Computational Provenance Letters

Not found. Citation not found in any enabled database (arxiv, crossref, dblp, google_books, open_library, openalex, pubmed, semantic_scholar)

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

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

#14 Critical **Citation completeness under constrained prompting**

K Zorin et al. (2022) · Annals of Automated Literature Practice

Not found. Citation not found in any enabled database (arxiv, crossref, dblp, google_books, open_library, openalex, pubmed, semantic_scholar)

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

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

#15 Critical **Citation completeness under constrained prompting**

K Zorin et al. (2022) · Journal of Generative Research Methods

Not found. Citation not found in any enabled database (arxiv, crossref, dblp, google_books, open_library, openalex, pubmed, semantic_scholar)

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

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

02 AI writing detection

59% estimated human-written · 9 passages flagged as likely AI-generated.

13% of analyzed sentences read as AI-generated, 44% as mixed, 43% as human.

Flagged passages

This approach aligns with the traceable-verification framework described by Yarrow and Chen (2023). **91%**

These cases can generate false alarms even when the underlying source is authentic. **82%**

These classes can overlap. **81%**

A benchmark corpus of 480 short literature-review passages was generated from fictional research prompts in education, public health, software engineering, and environmental policy. **81%**

The findings illustrate why citation detection should not be reduced to URL availability. **79%**

The first two checks were automated. **79%**

Thirty-six fictional reviewer profiles were simulated for benchmark construction, each rating whether a reference looked credible before verification. **79%**

The third used normalized string comparison with a 0.88 similarity threshold. **69%**

The workflow follows the layered model proposed by Quellman and Dey (2020), with the additional requirement that a resolved identifier must correspond to the exact cited title. **66%**

Semantic mismatch occurs when a source may exist in some form but does not support the claim for which it is cited. **64%**

The experiment is deliberately framed as a synthetic benchmark rather than a claim about real-world prevalence. **63%**

Four checks were applied in sequence: (1) identifier syntax, (2) identifier resolution, (3) title-author-venue consistency, and (4) claim-source relevance. **62%**

Every element of this study is synthetic. **61%**

This synthetic study models that failure mode by generating 480 literature-review passages under four prompting conditions and evaluating the resulting references with manual verification, identifier resolution, and metadata-consistency checks. **60%**

References serve two distinct functions in scholarly prose: they connect claims to prior work and they signal that the author has entered an accountable evidence chain. **58%**

Ratings were based only on the formatted citation. **58%**

The benchmark also highlights an interface problem. **58%**

The fourth relied on two reviewers who classified each citation as supportive, tangential, or unsupported. **55%**

For example, a fabricated article may be assigned to an invented journal but given a DOI prefix that resembles a legitimate registry pattern. **52%**

The present study therefore asks three questions. **52%**

A reference was counted as falsely accepted when it received a credibility score of 4 or 5 on a five-point scale despite being constructed as unverifiable. **51%**

The fabricated bibliography below is intentionally constructed to support such testing. **51%**

Each passage contained between three and seven citations. **50%**

Detection systems should be evaluated not only on whether they flag a citation, but on whether they identify the failure mode and provide evidence that a user can inspect. **50%**

The strongest surface predictors were the presence of a DOI-shaped string, a journal title containing a disciplinary keyword, and a page range longer than six pages. **49%**

This final class is especially difficult because metadata checks alone cannot establish evidentiary relevance (Alberon & Pike, 2021;Sato-Meridian, 2024). **48%**

Plausible formatting is not evidence of bibliographic validity. **48%**

We distinguish four error classes. **47%**

This gap between rhetorical credibility and bibliographic reality is referred to here as citation drift. **46%**

A layered verification workflow reduced false acceptance from 31.8% to 4.6%, primarily by combining DOI resolution with title-author cross-checking and venue validation. **45%**

References containing realistic journal names, conventional page ranges, and syntactically valid digital object identifiers were rated as credible by reviewers even when no underlying source existed. **44%**

More useful tools provide an audit trail: which fields were checked, which sources were queried, what matched, what failed, and whether the final judgment depends on human reading. **44%**

A balanced corpus should therefore include references that fail at different layers and a smaller control set of fully verifiable citations. **43%**

Across all conditions, 31.8% of fabricated references were initially rated as credible. **43%**

Third, can a lightweight verification sequence reduce false acceptance without imposing excessive review effort? **41%**

The fabricated results suggest that surface plausibility and citation density are poor proxies for verifiability. **40%**

AI-assisted writing systems can produce citations that are linguistically credible yet bibliographically unsupported. **40%**

A resolver can confirm that an identifier points somewhere, but not that it points to the cited work or supports the associated claim. **38%**

Second, how do automated checks differ in their sensitivity to distinct hallucination patterns? **37%**

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.