Comparative Analysis of AI Tools for Academic Research and Technical Report Writing

Author Name: **Ms.K. Abinaya**,Assistant Professor/CSE. Kings College of Engineering,Punalkulam.

Co-Author Name: **Shanmathi.S, Ramakavi.S, Sandhiya.A, Sandhiya.J**

Abstract

Students and researchers now have access to a wide range of artificial intelligence (AI) tools that claim to help with finding literature, reading papers, drafting text, editing language and managing references. These tools differ greatly in how they work, how reliable they are and how safely they can be used in academic work. This paper presents a comparative analysis of AI tools used for academic research and technical report writing. It groups the tools into four categories (general-purpose chat assistants, AI research and literature tools, writing and editing assistants, and reference managers), proposes a set of seven evaluation criteria, compares the categories against these criteria, and recommends a responsible workflow for students. The paper is conceptual and based on a review of published literature and the publicly documented functions of the tools. It does not report experimental results, and the ratings given are qualitative and illustrative. It concludes that no single tool is best for all tasks, that tools which are grounded in real sources are safer for literature work than tools which only generate text, and that human verification and honest disclosure remain essential.

Keywords: artificial intelligence, large language models, academic research, technical report writing, literature review, research tools, academic integrity, comparative analysis

1. Introduction

The release of large language model (LLM) chatbots has changed how students and researchers search, read and write. Transformer-based language models (Vaswani et al., 2017) trained on very large text collections can summarise articles, explain concepts, draft paragraphs and suggest code (Brown et al., 2020). Reviews of this development note both large opportunities for teaching, research and productivity and serious risks to accuracy and integrity (Kasneci et al., 2023; Dwivedi et al., 2023; Lund et al., 2023).

A student writing a technical report or a literature review today can choose from dozens of tools, and new ones appear every few months. Choosing among them is difficult because the tools solve different problems. A chat assistant that writes fluent text is not the same as a tool that searches a database of real papers, and neither is the same as a tool that formats citations. Using the wrong tool for a task can produce fabricated references, weak arguments or unintended plagiarism.

The objectives of this paper are: (i) to classify the main kinds of AI tools used in academic research and report writing; (ii) to propose criteria for comparing them; (iii) to compare the categories against these criteria and map them to stages of the research and writing process; and (iv) to discuss ethical issues and recommend a responsible workflow.

2. Background

2.1 Large Language Models in Academic Work

LLM-based assistants generate text by predicting likely words from patterns learned in training data. This makes them fluent and flexible, but it also means they can produce statements that sound correct yet are false, a problem known as hallucination (Ji et al., 2023). In scholarly settings this appears as invented references, wrong page numbers, or confident summaries of papers that do not say what is claimed (Alkaissi and McFarlane, 2023). Bender et al. (2021) further caution that such models reproduce patterns without understanding and can reflect biases in their training data.

2.2 Retrieval-Augmented Tools

One way to reduce this problem is retrieval-augmented generation, in which the system first retrieves relevant documents and then generates an answer based on them (Lewis et al., 2020). Many research-oriented tools use this idea together with scholarly databases, for example the open data platform described by Kinney et al. (2023). Because the answer is linked to retrievable sources, the user can check the evidence, although the summary itself can still be wrong.

2.3 Academic Integrity and Policy

Major publishers and journals have stated that AI tools cannot be listed as authors because they cannot take responsibility for the work, and that their use should be disclosed (Thorp, 2023). Detection software is not a dependable solution: Liang et al. (2023) found that GPT detectors wrongly flagged writing by non-native English speakers as AI-generated. Institutional rules differ, so students must check the policy of their own institution and course.

3. Categories of AI Tools

For this paper, tools are grouped by the main job they perform. Specific products change quickly, so the categories matter more than the product names. Examples are given only to make the categories concrete.

Category	Typical examples	Main function
General-purpose chat assistants	ChatGPT, Claude, Gemini, Copilot	Explain concepts, brainstorm, outline, draft and rewrite text, help with code and data, analyse uploaded documents
AI research and literature tools	Elicit, Consensus, Perplexity, Semantic Scholar, Scite	Search scholarly sources, summarise papers, extract findings into tables, show how papers cite one another
Writing and editing assistants	Grammarly, QuillBot, Writefull	Correct grammar, improve clarity and tone, paraphrase, suggest academic phrasing
Category	Typical examples	Main function
Reference and knowledge managers	Zotero, Mendeley (with plug-ins)	Store papers, format citations and bibliographies, organise notes

Table 1. Categories of AI tools used in research and report writing

4. Evaluation Criteria

The following seven criteria were chosen because they relate directly to the quality and safety of academic work.

- **Source grounding and verifiability:** does the tool link its claims to real, checkable sources?
- **Accuracy and hallucination risk:** how likely is it to produce false or invented content?
- **Writing support:** how well does it help with structure, clarity and technical style?
- **Research support:** how well does it help find, screen and synthesise literature?
- **Citation and reference handling:** does it produce correct, usable references?
- **Privacy and data control:** what happens to uploaded text and personal data?
- **Cost and accessibility:** are the main functions available free or at low cost, and are they easy to use?

5. Comparative Analysis

Table 2 compares the four categories using a three-level scale (High, Medium, Low). The ratings reflect the general design of each category and the issues reported in the literature. They are qualitative judgements for guidance, not measured scores, and individual products within a category may differ. For hallucination risk, a higher rating means a higher risk.

Criterion	Chat assistants	Research and literature tools	Writing and editing assistants	Reference managers
Source grounding	Low to Medium (depends on whether web or document access is used)	High	Not applicable	High (user-supplied sources)
Hallucination risk	High for references and facts	Medium (sources are linked, summaries may still err)	Low (edits existing text)	Low
Writing support	High	Low	High for language, Medium for structure	Low
Criterion	Chat assistants	Research and literature tools	Writing and editing assistants	Reference managers
Research support	Medium	High	Low	Medium (organising only)
Citation handling	Low unless checked by the user	Medium to High	Low	High
Privacy control	Varies by plan and settings	Varies	Varies, text is often sent to servers	Medium to High (local storage possible)
Cost and accessibility	Free tiers available, advanced features paid	Free tiers with limits	Free basic version, advanced features paid	Mostly free or low cost

Table 2. Qualitative comparison of tool categories (illustrative)

5.1 Key Observations

First, general-purpose chat assistants are the most flexible tools and are strongest for brainstorming, explaining difficult ideas, structuring a report and improving drafts. However, they carry the highest risk when used to produce facts or references without checking, since fluent text can hide errors (Alkaissi and McFarlane, 2023).

Second, research and literature tools are better suited to finding and reviewing evidence because their answers are tied to retrievable papers. Their weakness is coverage and depth: they may miss relevant work, depend on abstracts rather than full texts, or summarise findings too simply.

Third, writing and editing assistants improve readability and are especially helpful to non-native English writers, but

they do not check whether the content is correct. Heavy paraphrasing can also change the meaning of technical statements.

Fourth, reference managers are the most dependable for citation accuracy because the data comes from the user's own library. They do not write or search intelligently, so they complement rather than replace the other categories.

5.2 Mapping Tools to Stages of Work

Stage	Suitable tools	Human responsibility
Topic selection and scoping	Chat assistants	Judge whether the topic is original and feasible
Literature search	Research and literature tools	Read the actual papers and check coverage
Stage	Suitable tools	Human responsibility
Reading and note-taking	Chat assistants (with uploaded papers), research tools	Verify every summary against the source
Outlining and drafting	Chat assistants	Own the argument, write the final text
Language editing	Writing and editing assistants	Confirm meaning has not changed
Citations and bibliography	Reference managers	Check every entry for accuracy
Final review	None to replace the author	Check facts, originality and compliance with policy

Table 3. Suggested use of tools across the research and writing process

6. Illustrative Scenario

The following hypothetical example shows how the comparison can guide tool choice. Ananya, a third-year computer science student, must write a technical report on lightweight machine-learning models for mobile devices. She uses a chat assistant to understand the topic and draft a possible outline. She then uses a literature tool to find recent papers, downloads the relevant ones, and reads them herself, noting which claims the tool's summaries got right and wrong. She stores the papers in a reference manager, which generates the citations. She writes the report in her own words, uses a chat assistant to suggest improvements to structure, and a writing assistant to correct grammar. Before submitting, she checks every reference against the source and adds a statement describing how AI tools were used. In this workflow each tool does the job it is best at, and the student remains responsible for the content.

7. Proposed Evaluation Plan for Future Studies

Because this paper is conceptual, the ratings in Table 2 should be tested in later work. A mixed-methods approach is recommended.

- **Reference accuracy test:** ask each tool for references on the same set of topics and check what percentage exist and are correctly cited.
- **Summary fidelity test:** compare tool-generated summaries of sample papers with summaries written by experts.
- **Writing quality:** have independent reviewers score reports drafted with different tools on clarity, structure and technical correctness.
- **Usability:** measure perceived ease of use and usefulness (Davis, 1989) and use the System Usability Scale (Brooke, 1996).

- **Efficiency:** record time taken for literature search and drafting with and without tools.
- **Privacy review:** compare data-use and retention policies of the tools.

8. Ethical Considerations

Responsible use of AI in academic work depends on a few principles, which relate to the broader ethical framework proposed by Floridi et al. (2018).

- **Honesty and disclosure:** state clearly which tools were used and for what purpose, following institutional rules (Thorp, 2023).
- **Accountability:** the author is responsible for every claim, whatever tool helped produce it.
- **Verification:** never cite a source that has not been opened and read.
- **Privacy and confidentiality:** do not upload unpublished data, confidential material or personal information to tools without understanding how that data is stored and used.
- **Fairness:** be aware of bias in outputs, and of the unreliability of AI-detection software, which can unfairly affect non-native English writers (Liang et al., 2023).
- **Learning integrity:** use tools to support understanding, not to avoid the thinking that the assignment is meant to develop.

9. Recommendations

1. Choose tools by task, not by popularity: use research tools for finding sources, chat assistants for explanation and structure, writing tools for language, and reference managers for citations.
2. Treat all AI output as a draft that needs checking, especially facts, numbers, quotations and references.
3. Prefer tools that show their sources, and open those sources yourself.
4. Keep records of prompts and tool use so that disclosure is accurate.
5. Read and follow the AI policy of the institution, course or journal.
6. Protect data: use privacy settings and avoid sharing sensitive material.

10. Limitations

This analysis has several limitations. AI tools change quickly, so features, prices and quality may differ from those described here. The categories are simplified, and many products now combine functions from more than one category. The ratings are qualitative and not based on experiments. Finally, the paper does not cover discipline-specific tools such as those for data analysis or code generation in depth. Empirical testing as described in Section 7 is needed to confirm the findings.

11. Conclusion

AI tools can make research and report writing faster and can help students understand difficult material, but their value depends on choosing the right tool for each task and checking what it produces. General-purpose chat assistants are flexible but risky for facts and references; research tools are better grounded in real literature; writing assistants improve language but not correctness; and reference managers provide reliable citation handling. A combined workflow with clear human responsibility, verification and honest disclosure gives the best balance of benefit and risk. Future work should test these categories experimentally and update the comparison as the tools continue to develop.

References

Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus, 15*(2), e35179.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). ACM.

Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan et al. (Eds.), *Usability evaluation in industry* (pp. 189-194). Taylor & Francis.

Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877-1901.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340.

Dwivedi, Y. K., Kshetri, N., Hughes, L., et al. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, 102642.

Floridi, L., Cowls, J., Beltrametti, M., et al. (2018). AI4People: An ethical framework for a good AI society. *Minds and Machines, 28*(4), 689-707.

Ji, Z., Lee, N., Frieske, R., et al. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), Article 248.

Kasneci, E., Sessler, K., Küchemann, S., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274.

Kinney, R., Anastasiades, C., Authur, R., et al. (2023). The Semantic Scholar Open Data Platform. *arXiv preprint* arXiv:2301.10140.

Lewis, P., Perez, E., Piktus, A., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems, 33*, 9459-9474.

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns, 4*(7), 100779.

Lund, B. D., Wang, T., Mannuru, N. R., et al. (2023). ChatGPT and a new academic reality: Artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology, 74*(5), 570-581.

Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science, 379*(6630), 313.

Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998-6008.