A Digital Safety Net: Detecting AI-Generated Hallucinations With the CAM Software Solution

Christopher A. Clemons
Associate Professor
Auburn University
Agricultural Education, Leadership, and Communication
345 West Samford Ave.
College of Education
Alabama, 36849
Cac01323@auburn.edu

A Digital Safety Net: Detecting AI-Generated Hallucinations With the CAM Software Solution

Introduction/Need for Innovation or Idea

The adoption of generative artificial intelligence has the potential to threaten the academic integrity of research and manuscript preparation (Grace et al., 2024; Zhu et al., 2025). The creation of hallucinated or entirely fabricated citations has existed prior to the introduction of AI; however, the advent of AI has perpetuated the perceived reliability of information. These fallacies undermine the foundation of scholarly research, which relies on a verifiable chain of evidence to ensure trustworthiness and integrity. Existing software tools are largely ill-equipped to solve this problem. Reference management software is proactive yet fails to externally validate a user-curated library, assuming the source data is valid. Style checkers and sentence pattern recognition programs can confirm a correctly formatted reference while simultaneously failing to validate the factual existence of the citation. Often, manually verifying every source is prohibitively time-consuming. This creates an unmet need for a simple and reactive proofreading tool that researchers can easily use (Davis, 1985) at the end of the writing process for a final integrity check. To address this gap, the Citation Assisted Management (CAM) software package was developed as a standalone, user-friendly application to support academics in their research and writing.

How it Works/Methods

The conceptual framework for this innovation leverages a large language model (LLM) as an intelligent assistant to perform multi-factor analysis of a scholarly manuscript. This approach moves beyond rigid, rule-based decision tree structures to understand the nuances and context of academic writing. The framework is built on the concept of AI-directed external data validation, where the AI is instructed to use a key data point within a reference, the digital object identifier (DOI), to cross-reference the citation components, such as author and title, against official and externally retrieved publication metadata.

A standalone desktop application, known as Citation Assisted Management (CAM), was developed to serve as an intelligent safety net for researchers. The user-friendly interface is built using Python with its native Tkinter library, ensuring broad compatibility. The application's core logic communicates with Google's Gemini Application Programming Interface (API), which serves as the backend engine. A user then uploads their entire manuscript (e.g., .docx, .pdf), and the CAM program calculates a weighted similarity analysis to determine common and unique issues in author-created research manuscripts, e.g., citation mismatches, APA 5, 6, and 7 formatting errors, and AI hallucination detection with preliminary testing indicating a high degree of accuracy This innovative process is based on a novel method for which a patent pending application is pending (Clemons, 2025). The program then generates a single, actionable report that categorizes all potential mismatches and formatting errors, allowing the author to serve as the expert-in-the-loop, ensuring user control and final assessment. The CAM application demonstrates a bilateral process of creating and successfully implementing a deterministic and heuristic classification model that employs a linear decision boundary algorithm. To establish reliable AI hallucination detection, the author developed a method that projects parentheticals and references onto a single dimension by calculating a weighted linear combination of semantic similarity scores. The classification is then performed by a threshold activation function on the

probability score analysis between the authenticity of the citation and the predetermined value of the activation function.

Results to Date/Implications

A functional prototype has been successfully developed and extensively tested prior to alpha/beta level testing, patent submission, and end-user evaluation. The CAM program has undergone an iterative development process, evolving from an initial prototype (Version 1.01, May 2025) to the current operating system (3.0.1, September 2025), with enhanced features and reliability. The initial version established the core functionality of AI-powered cross-referencing. A major usability upgrade was introduced in version 2.0, transitioning the system from manual text entry of the manuscript to a seamless file (.docx and .pdf) support system. Major updates were added to include analysis for multiple APA formatting versions and live reference verification using external search coding to address potential errors, such as broken digital object identifiers (DOIs). Reliability and stability of the processing code were addressed by a self-correcting (recursive) coding pathway to handle malformed data from the API, prevent software errors, and prevent "crashing." The software has been optimized to handle large documents while minimizing the likelihood of timeout errors when CAM accesses the API.

Future Planning/Advice to Others

This innovation has the potential to positively impact the academic integrity of research-based manuscript submissions while saving researchers valuable time. For the profession, CAM streamlines the manuscript preparation, allowing graduate students and faculty to focus more on their research programs. Addressing the future of the CAM, the Auburn University Office of Technology Transfer will incorporate the software into the internal graduate software suite, student writing center, and system-wide libraries' production tools, while developing promising large-scale distribution licensing with external private educational software developers. This would provide a high-value academic tool to the campus community and reinforce Auburn University's commitment to evolving AI policies, potential, and academic integrity. Upon completion of testing, the CAM software will be made available to members of the American Association of Agricultural Education at no cost for download, subsidized by the developer's technology transfer credit offsets through Auburn University. Authors and members of AAAE will be the only individuals to be granted no charge access as this tool was conceptualized and developed with the intent to improve the efficiency of our field, provide a direct and tangible product to researchers, and to fulfill the "giving back" ethos of the developer in response to the professional and intellectual growth AAAE has provided to the author.

Costs/Resources Needed

The development of the Citation Assisted Management (CAM) software has been highly efficient by leveraging existing institutional resources and the developer's intellectual property capital during the pre-release stage. Capital outlay this budget covers final validation, faculty development time, a three-month release candidate test with participation incentives, and the initial Auburn University-wide deployment. This estimate also includes Phase Two allocations, such as a 12-month campus-wide deployment and funding allocation for one course release. The direct technology costs for API usage are negligible (approximately $0.08 per analysis).

References

Clemons, C. A. (2025). *System and method for AI-driven detection of fabricated scholarly references* (U.S. Patent Application No. 63/914,699). U.S. Patent and Trademark Office (USPTO).

Davis, F. D. (1985). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* (Doctoral dissertation, Massachusetts Institute of Technology).

Grace, K., Stewart, H., Sandkühler, J. F., Thomas, S., Weinstein-Raun, B., & Brauner, J. (2024). Thousands of AI authors on the future of AI. https://doi.org/10.48550/arXiv.2401.02843

Zhu, Y., Huang, Z., Dou, Z., & Wen, J. R. (2025). One token can help! Learning scalable and pluggable virtual tokens for retrieval-augmented large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence, 39*(24), 26166–26174. https://doi.org/10.1609/aaai.v39i24.34813