Investigating Linguistic and Content Accuracy of Artificial Intelligence-Assisted Tools for English Writing: A Case Study of English Popular Press Articles

Cholthicha Sudmuk
Department of Linguistics, Graduate School, Suan Sunandha Rajabhat University, Thailand
E-mail: cholthicha.su@ssru.ac.th
Suwaree Yordchim
Department of Linguistics, Graduate School, Suan Sunandha Rajabhat University, Thailand
E-mail: suwaree.yo@ssru.ac.th
Sahar Aghaei
Department of Psychology, Undergraduate student, Michigan State University, USA
E-mail: aghaeisa@msu.edu

Abstract

This study investigated the linguistic and content accuracy of AI-assisted tools in summarizing and evaluating popular press articles that report empirical research. A qualitative content analysis design was employed to examine the performance of AI in summarizing and evaluating English-language popular press articles. The findings showed that AI-generated summaries and evaluations have a high level of linguistic and factual accuracy. Although the AI-generated evaluations also exhibited substantial factual accuracy, they lacked sufficient supporting evidence from the original press articles. These findings are consistent with previous research suggesting that AI-assisted tools primarily rely on natural language processing algorithms to enhance linguistic accuracy rather than to produce comprehensive evidence-based judgments. Consequently, the AI-generated evaluations frequently omitted the detailed supporting information required for rigorous evidence-based reporting.

Keywords: Linguistic Accuracy, Artificial Intelligence-Assisted Tools, English Writing, Popular Press Articles

Introduction

The rise of artificial intelligence (AI) has transformed the production and distribution of written content. Breakthroughs in large language models (LLMs) and natural language processing (NLP) have led to AI-assisted writing tools that mimic human writing with fluency and coherence (Brown et al., 2020; OpenAI, 2023). These tools offer immediate feedback to enhance linguistic accuracy in academic writing, making them valuable in higher education (Kasneci et al., 2023). Consequently, tools like ChatGPT and Grammarly have become essential for students, writers, journalists, and content creators by improving linguistic accuracy and streamlining the writing process.

Linguistic accuracy is essential for effective writing, reflecting adherence to grammatical rules, sentence structure, appropriate word choice, and proper mechanics such as spelling and punctuation. In English as a Foreign Language (EFL), many learners struggle with linguistic accuracy in writing tasks due to limited exposure and gaps in their knowledge of English grammar (Hyland, 2019; Nation, 2001). Common errors, like subject–verb agreement and article misuse, can obscure meaning and reduce the quality of writing, affecting both credibility and reader comprehension (Bunnarong et al., 2023; Ferris, 2011; Hinkel, 2004). Thus, achieving linguistic accuracy should be a fundamental goal in developing English writing skills.

The emergence of AI-supported writing tools has introduced new possibilities for addressing persistent difficulties of linguistic accuracy by offering essential automated feedback on grammar, vocabulary, sentence structure, and mechanics. Previous studies demonstrate that such assistance can not only refine writers' revision practices but also lead to notable enhancements in linguistic precision and overall writing performance (Escalante et al., 2023). While much of the existing research has focused on the use of AI-assisted tools for improving student writing and language learning, comparatively little attention has been paid to their effectiveness in evaluating journalistic texts that communicate scientific findings to the public.

A popular press article, such as a news website or magazine, that covers an empirical study translates primary, data-driven research into accessible, everyday language. It helps the public understand complex scientific findings without requiring academic training. These kinds of articles play a significant role in translating complicated empirical research into accessible language for non-specialist audiences, thereby influencing public understanding of science and research-based knowledge (Schäfer, 2017). However, the process of simplifying research findings may introduce issues related to interpretation and content accuracy, making critical evaluation essential.

The rapid advancement of AI technologies has changed how written texts are analyzed in educational and professional settings. Generative AI tools now offer automated feedback on writing quality, including content organization and coherence (Alpar, 2025; Kasneci et al., 2023). While studies show these systems can provide valuable insights and revision strategies, there are variations in the specificity of their evaluations (Alpar, 2025; Cotton et al., 2024). Concerns remain about the reliability of AI assessments for science communication texts. These subtle inaccuracies can affect the quality and credibility of information in popular press articles, impacting public understanding of scientific issues.

As AI technologies become more prevalent in educational and research environments, this study aims to investigate their effectiveness in summarizing and evaluating popular press articles reporting on empirical studies, with an emphasis on both linguistic and content accuracy. This examination will enhance research on AI-assisted writing tools and offer educators insights into effectively integrating AI tools into the teaching of writing skills.

Research Objectives

To investigate the linguistic and content accuracy of AI-assisted tools' summary and evaluation of popular press articles that report empirical studies.

Scope of Research

This paper investigates the linguistic and content accuracy of ChatGPT, which is the first AI technology to make complex, human-like interactions highly accessible and user-friendly for the public. According to Digital-adoption.com, ChatGPT is one of the 50 most-visited websites in the world. The data consisted of three English-language popular-press articles reporting empirical research findings in the fields of education, health, and social sciences.

Literature Review

Recent advancements in AI have greatly impacted language evaluation and writing support. Kasneci et al. (2023) indicate that Generative AI systems like ChatGPT, Google Gemini, and Microsoft Copilot use LLMs trained on extensive text data to create and revise

written content. These tools identify grammatical errors, suggest improvements, enhance coherence, and offer feedback on writing quality. According to Dwivedi et al (2023), the growing availability of AI has prompted educators and researchers to incorporate AI-assisted tools into writing instruction and assessment.

Research on AI-assisted writing tools has shown positive results. Alpar (2025) found that these tools effectively supported EFL students in writing opinion essays, with ChatGPT-4 providing notably better feedback. Similarly, Kasneci et al. (2023) highlighted the benefits of AI models for personalized feedback and immediate language support, indicating a strong potential for improving writing skills.

Most research has concentrated on student writing and academic essays, while less focus has been given to evaluating journalistic texts such as popular press articles that communicate empirical research. As AI becomes more integrated into education and professional practices, it is important to assess how effectively these tools achieve linguistic and content accuracy in science-related media texts.

Research Methodology

This study employed a qualitative content analysis design to investigate the linguistic and content accuracy of ChatGPT's fifth-generation model in summarizing and evaluating English-language popular press articles reporting on empirical research findings. The study focused on comparing the summary and evaluation created by ChatGPT and assessing their accuracy against established linguistic criteria.

Data Selection

The data consisted of three English-language popular press articles reporting empirical research findings in education, health, and the social sciences. The articles were randomly selected from reputable online news outlets that regularly publish science-related content for general audiences. Three popular press articles are as follows:

1. The popular press article in the field of education is "*Why do some kids learn to talk earlier than others?*" by Christy DeSmith (Harvard Gazette, 2024), which discusses an empirical paper, titled "Everyday Language input and production in 1,001 children from six continents," published in *the Proceedings of the National Academy of Sciences (PNAS)* by developmental psychologist Elika Bergelson and colleagues.
2. The popular press article in the field of health is *"Large study presents evidence for behavioral sciences in policymaking"* by researchers at Columbia University's Mailman School of Public Health (2023), which discusses an empirical paper, titled "A Synthesis of Evidence for Policy from Behavioral Science during COVID-19," published in *Nature Human Behavior*, 625(7993), 134–147, by Ruggeri et al. (2024).
3. The popular press article in the field of the social sciences is "*Friendships that Bridge Wealth Divides Help Social Mobility, Study Finds*" by Heather Stewart (The Guardian, 2025), which discusses an empirical paper titled "Social Capital in the United Kingdom: Evidence from Six Billion Friendships", OSF Preprints. https://doi.org/10.31235/osf.io/kb7dy by Harris et al. (2025).

Linguistic Accuracy Criteria

The linguistic accuracy criteria used in this study were adapted from the framework for analyzing learner language proposed by Ellis and Barkhuizen (2005). This framework combines traditional measures of linguistic accuracy with discourse-level evaluations of

textual effectiveness. These criteria emphasize examining language performance through multiple dimensions of accuracy and textual quality rather than focusing solely on grammatical errors.

There are five grammatical accuracies in the criteria:

1. Lexical Accuracy: concerns the appropriate and precise use of vocabulary to convey intended meanings. There are four indicators: appropriate word choice, correct collocations, accurate use of technical terms, and avoidance of semantic errors.

2. Syntactic Accuracy: refers to the correctness and sophistication of sentence structures. There are four indicators: well-formed sentences, appropriate clause combinations, correct use of subordination and coordination, and absence of sentence fragments and run-ons.

3. Mechanical Accuracy: involves the correct application of writing conventions. There are four indicators: spelling, punctuation, capitalization, and formatting.

4. Cohesion: refers to the linguistic devices that connect sentences and ideas within a text. There are five indicators: conjunctions, reference words, lexical repetition, substitution, and ellipsis.

5. Coherence: refers to the overall logical organization and meaningful progression of ideas throughout a text. There are five indicators: clear topic development, logical sequencing, consistent focus, well-structured paragraphs, and clear argumentation.

Content Accuracy Criteria:

Content accuracy criteria are standards used to evaluate whether the information presented in a text accurately represents the source, facts, research findings, or subject matter. In this study, the criteria are synthesized from the work of Schwitzer (2008), Sumner et al. (2014), and Trench and Bucchi (2021), as these studies specifically examine how accurately media reports represent scientific research.

There are eight content accuracies in the criteria:

1. Accuracy of Research Purpose: Does the article accurately explain the purpose or research question of the original study?

2. Accuracy of Methodology: Does the article accurately describe how the study was conducted?

3. Accuracy of Findings: Are the study results reported correctly?

4. Accuracy of Interpretation: Does the article interpret the findings appropriately?

5. Statistical Accuracy: Are quantitative findings represented correctly?

6. Source Attribution Accuracy: Is the source properly identified?

7. Terminological Accuracy: Are scientific concepts and terminology used correctly?

8. Balance and Objectivity: Is the reporting balanced and free from sensationalism?

Data Analysis Procedures

1. The selected articles were submitted individually to ChatGPT's fifth-generation model, using a standardized prompt designed to summarize and evaluate the press articles. To ensure consistency, all prompts were administered under the same conditions and within a similar timeframe.
2. The AI summary and evaluation were separately collected, transcribed, and organized as data for analysis.
3. The AI summary and evaluation were examined with linguistic accuracy criteria.
4. The AI summary and evaluation were examined with the content accuracy criteria.

- The AI summary of the popular press article: Were there any factual inaccuracies?
- The AI evaluation of the popular press article: Were there any pieces of information that went beyond what is directly supported by the results

Ethical Considerations

All data analyzed in this study were obtained from publicly available sources. No human participants were involved. Therefore, informed consent was not required. The study adhered to principles of academic integrity by accurately reporting AI-generated outputs and properly acknowledging all sources used in the analysis.

Results

The results of the analysis of the three selected popular press articles' the AI summary and evaluation are divided into two parts: Part 1: The linguistic accuracy of the AI summary and evaluation, and Part 2: The content accuracy of the AI summary and evaluation.

Part 1: The linguistic accuracy of the AI summary and evaluation

The results of the analysis of the linguistic accuracy of both the AI summary and evaluation by the linguistic accuracy framework of Ellis and Barkhuizen (2005) are as follows:

1. Grammatical Accuracy

All AI-generated summaries and evaluations of the three press articles exhibit excellent grammatical accuracy. Sentences are well-formed and follow standard English grammar conventions. Subject-verb agreement, tense consistency, article usage, and pronoun reference are accurate throughout the text. Example: *The article also quotes the lead researcher directly, allowing readers to understand the researchers' interpretations of the findings rather than relying solely on journalistic paraphrasing* (DeSmith's 2024 AI-generated summary). This sentence demonstrates correct subject-verb agreement ("*The article quotes*"), appropriate parallel structure, and accurate clause construction.

2. Lexical Choice

Technical terms are used accurately without excessive jargon in all AI-generated summaries and evaluations of the three press articles, allowing accessibility while maintaining scientific rigor. Example: *Technical concepts such as "behavioral science," "policy recommendations," and "evidence synthesis" are explained through practical examples related to COVID-19 public health policies* (ScienceDaily's 2023 AI-generated evaluation). The vocabulary is precise, discipline-specific, and appropriate for communicating behavioral science research to a general audience.

3. Syntactic Complexity and Accuracy

All AI-generated summaries and evaluations of the three press articles contain a variety of sentence structures, including simple, compound, and complex constructions. Complex sentences are effectively controlled and contribute to informational density without reducing readability. Example: *The article demonstrates a generally balanced approach. While emphasizing the benefits of cross-class friendships, it avoids claiming that social connections alone determine social mobility* (Stewart's 2025 AI-generated evaluation). The articles utilize a variety of sentence structures. Complex information is often condensed into syntactically sophisticated yet readable constructions.

4. Mechanics (Spelling, Punctuation, and Capitalization)

Mechanical accuracy is consistently high throughout all AI-generated summaries and evaluations of the three press articles. Specialized terminologies, institutional names, and

researcher affiliations are correctly capitalized and punctuated. Example: *DeSmith correctly explains that the strongest predictors of children's language production were age, clinical factors (e.g., prematurity and developmental conditions), and the amount of speech children hear in their environment.* (DeSmith's 2024 AI-generated summary).

5. Cohesion

All AI-generated summaries and evaluations of the three press articles use cohesion as a linguistic device to connect sentences and ideas. Example: *One of the strengths of the article is its accessibility to non-specialist readers. Technical concepts such as "behavioral science," "policy recommendations," and "evidence synthesis" are explained through practical examples related to COVID-19 public health policies.* (ScienceDaily's 2003 AI-generated evaluation). The articles demonstrate strong cohesion, using reference devices and transitional expressions. Key concepts such as *behavioral science*, *evidence*, *policy*, and *COVID-19* are repeated strategically to maintain thematic continuity.

6. Coherence

Each section of the AI-generated summaries and evaluations of the three press articles contributes directly to the article's primary purpose, resulting in a coherent narrative structure. For example, Stewart's 2025 AI-generated summary and evaluation are highly coherent, as every section contributes directly to the central argument that friendships across socioeconomic groups can improve economic outcomes and social mobility. Their progression is logical and enables readers to understand both the evidence and its broader significance.

In sum, the linguistic accuracy framework of Ellis and Barkhuizen (2005) demonstrates a very high level of linguistic accuracy of all AI-generated summaries and evaluations of the three press articles. All of them successfully combine grammatical precision, appropriate vocabulary, varied sentence structures, and strong cohesion and coherence. Consequently, it validates the linguistic accuracy of AI-assisted tools when generating or evaluating popular press articles that cover an empirical study.

Part 2: The content accuracy of the AI summary and evaluation

The results of this part are divided into two categories: AI summaries of popular press articles and AI evaluations of popular press articles.

1. AI summaries of popular press articles

The analysis of the AI summaries reveals their content accuracy. DeSmith's 2024 summary accurately outlines findings from Bergelson et al. (2024) on language development in 1,001 children across 12 countries, highlighting key predictors like age, clinical factors, and environmental speech exposure. ScienceDaily's 2003 summary correctly reports Ruggeri et al. (2023), which found empirical support for 18 out of 19 COVID-19 policy recommendations from a review of 747 studies, noting some insufficient evidence. Stewart's 2025 summary effectively presents a study on socioeconomic backgrounds and social mobility, showing that low-income children in economically connected areas earn approximately £5,100 more as adults than peers from less connected communities.

In sum, the AI summary has factual accuracy of the purposes, methods, and findings with significant points from the press article as supporting evidence.

2. AI evaluations of popular press articles

The AI evaluation is a formative assessment that identifies strengths and weaknesses of each popular press article across five categories: accuracy of research reporting, clarity and accessibility, representation of evidence, balance and objectivity, linguistic quality, and educational value. In each category, the AI evaluation gives the overall

assessment without detailed information or supporting examples from popular press articles. For example, DeSmith's 2024 evaluation states that the press article has excellent factual accuracy and faithful representation of the original research, but it does not provide concrete examples from the press article to support the evaluation's claim. ScienceDaily's 2003 evaluation reports that the press article demonstrates strong evidence-based reporting by including quantitative information about the study design and findings, such as the number of experts involved, the number of countries represented, and the number of studies reviewed. However, there is no example from the press article as supporting evidence. Stewart's 2025 evaluation indicates that the press article employs effective examples, such as comparisons between Kingston upon Thames and Canterbury, to illustrate how levels of economic connectedness can vary between communities with similar socioeconomic profiles. However, no detailed information from the press article is referred to support this argument.

In sum, the analysis of AI evaluation indicates a high level of factual accuracy in the content, but it lacks supporting evidence from the press article. It does not include detailed information that serves as the basis for evidence-based reporting. Additionally, it fails to provide simplified methodological details for general readers, a key feature of popular press articles that discuss empirical studies.

Conclusion and Discussion

This study investigates the effectiveness of AI-assisted tools for summarizing and evaluating popular press articles reporting on empirical studies. It focuses on both linguistic and content accuracy. The findings reveal that the AI-generated summaries and evaluations demonstrate a high level of linguistic accuracy. In terms of content accuracy, both exhibit excellent factual correctness, but the AI evaluations lack detailed supporting information from the press articles.

The findings of this study on the high level of linguistic accuracy in AI-generated summaries and evaluations reinforce the results of many previous studies on the usefulness of AI-assisted tools for enhancing linguistic accuracy in academic writing (Escalante et al., 2023). AI-assisted tools utilize natural language processing algorithms to identify errors, suggest revisions, generate alternative expressions, and provide a mechanical writing check. Therefore, it enhances linguistic accuracy in summary and evaluation.

The AI-generated summary demonstrates strong factual accuracy, supported by key points from the press article. This aligns with recent advancements in AI tools that provide automated feedback on writing quality and organization (Alpar, 2025; Kasneci et al., 2023). However, while the AI evaluation also shows high factual accuracy, it lacks detailed evidence from the press article. Studies indicate that generative AI can produce inaccurate feedback and overlook contextual nuances (Cotton et al., 2024; Dwivedi et al., 2023). This may stem from AI algorithms that imitate or compile content rather than make evidence-based judgments. Since the press article covers an empirical study, the complexity of data might affect the overall content accuracy of the AI-generated evaluation that AI algorithms can achieve.

References

Alpar, Ö. (2025). Evaluating generative AI tools for improving English writing skills: A preliminary comparison of ChatGPT-4, Google Gemini, and Microsoft Copilot. *European Journal of Educational Research, 14*(4), 1291–1308. https://doi.org/10.12973/eu-jer.14.4.1291

Bergelson, E., Soderstrom, M., Schwarz, I.-C., Rowland, C. F., Ramirez-Esparza, N., Hamrick, L. R., Marklund, E., ... & Cristia, A. (2024). Socioeconomic status does not reliably predict young children's language experience and vocalizations: Evidence from a large-scale, cross-cultural dataset. *Proceedings of the National Academy of Sciences*, *121*(1), Article e2300671121. https://doi.org/10.1073/pnas.2300671121

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Bunnarong, L., Chomyong, S., & Seedee, A. (2023). An analysis of English subject-verb-agreement and verb inflections by morpho-syntactic approach for EFL teaching at Rajamangala University of Technology, Thailand. *Academic Journal of Buriram Rajabhat University*, *15*(1), 27–54.

Columbia University's Mailman School of Public Health. (2023, December 13). *Large study presents evidence for behavioral sciences in policymaking*. ScienceDaily. sciencedaily.com.re

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating? Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, *61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Dwivedi, Y. K., Kshetri, N., Hughes, L., Ribeiro-Navarrete, S., Malcom, A. K., Kar, A. K., ... & Loureiro, S. M. C. (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*, Article 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford University Press.

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, *20*(1), Article 57. https://doi.org/10.1186/s41239-023-00425-2

Ferris, D. R. (2011). *Treatment of error in second language student writing* (2nd ed.). University of Michigan Press.

Hinkel, E. (2004). *Teaching academic ESL writing: Practical techniques in vocabulary and grammar*. Lawrence Erlbaum Associates.

Hyland, K. (2019). *Second language writing* (2nd ed.). Cambridge University Press.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poguet, O., Sailer, M., Schmidt, A., Stadler, M., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

OpenAI. (2023). *GPT-4 technical report*. arXiv. https://doi.org/10.48550/arXiv.2303.08774

Ruggeri, K., Stock, F., Haslam, S. A., Capraro, V., Boggio, P., Ellemers, N., Cichocka, A., Douglas, K. M., Rand, D. G., van der Linden, S., Cikara, M., Finkel, E. J., Druckman, J. N., Wohl, M. J. A., Petty, R. E., Tucker, J. A., Shariff, A., Gelfand, M., Packer, D., . . . Willer, R. (2024). A synthesis of evidence for policy from behavioural science during COVID-19. *Nature*, *625*(7993), 134–147. https://doi.org/10.1038/s41586-023-06840-9

Schäfer, M. S. (2017). How changing media structures are affecting science news coverage. In K. H. Jamieson, D. M. Kahan, & D. A. Scheufele (Eds.), *The Oxford handbook of the science of science communication* (pp. 51–59). Oxford University Press.

Schwitzer, G. (2008). How do US journalists cover treatments, tests, products, and procedures? An evaluation of 500 stories. *PLoS Medicine*, *5*(5), Article e95. doi.org

ScienceDaily. (2023, December 13). *Large study presents evidence for behavioral sciences in policymaking*. Columbia University's Mailman School of Public Health. https://www.sciencedaily.com/releases/2023/12/231213112502.htm

Stewart, H. (2025, March 23). Friendships that bridge wealth divides help social mobility, study finds. *The Guardian*. https://www.theguardian.com/society/2025/mar/24/friendships-that-bridge-wealth-divides-help-social-mobility-study-finds

Sumner, P., Vivian-Griffiths, S., Boivin, J., Williams, A., Venetis, C. A., Davies, A., Ogden, J., Whelan, L., Hughes, B., Dalton, B., Boy, F., & Chambers, C. D. (2014). The association between exaggeration in health-related science news and academic press releases. *BMJ*, *349*, Article g7015.

Trench, B., & Bucchi, M. (Eds.). (2021). *Routledge handbook of public communication of science and technology* (3rd ed.). Routledge.