More A than I: Testing for Large Language Model Plagiarism in Political Science

Robert Keener, *The University of Tennessee, Knoxville, USA*

ABSTRACT This article shows how the sudden introduction of large language models (LLMs) has allowed a sudden, significant increase in the ability of political science professionals to plagiarize their articles by prompting LLMs to write for them. Evidence of this is shown through a brief overview of the limitations of LLMs and by searching for words that are disproportionately used by the most popular LLM, ChatGPT, in peer-reviewed articles. What is found is a rapid spike in the use of words that are unremarkable except for their popularity in ChatGPT's output as determined by an AI professional. This shows that this method can be used to indicate the likelihood of plagiarism in a given article. It then concludes with the limitations of this keyword detection method and recommendations for limiting LLM plagiarism in the field of political science as a whole.

INTRODUCTION

There has yet to be a full tally of exactly how seismic the effects were of November 30, 2022, the first day ChatGPT entered public use. ChatGPT is a large language model (LLM) that uses AI to generate normally accurate text. This almost immediately became a concern in that dishonest political science professionals could now automate one of the most difficult and characteristic parts of academia—publishing generative peer-reviewed articles (Michels 2023). This article will give background on how LLMs work, what their current limitations are, and how these limitations can be exploited to detect AI-generated articles. Then, this article shows this "keyword detection method" in action as it is used to show the sudden uptick of suspicious manuscripts between the years 2022 and 2023—as ChatGPT, the most popular LLM, reached public use.

Using this keyword detection method, this article shows that the spike in AI-generated manuscripts has almost certainly already compromised the profession and risks the credibility of all academics if it is not regulated. This is accomplished by four models, going from the most general to the most specific, showing that AI plagiarism is already a problem among peer-reviewed articles. Furthermore, this article ends with a discussion on the limits of the keyword detection method and how it is not intended to be a catch-all solution but a tool in a belt of other AI detection methods. Any reviewer who takes their job and their field seriously has the responsibility to adapt to these large changes, so the final goal of this article is to help them accomplish this task.

THEORY

Technology advances more rapidly over time, and it is incumbent on academic institutions and professions to monitor these developments to prevent dishonest practices. Major developments that have been vehicles for dishonest practices have normally been used not among professionals but among students including the introduction of smartphones, the use of Wiki platforms when writing papers, and the use of programable calculators. Smartphones became efficient ways to portably retrieve and organize data but had to be limited to times outside of test-taking. Wiki platforms are often used as fast ways to gain a background understanding of countless subjects, but it is understood that all the information needs to be confirmed in more reputable platforms, and programable calculators are still fantastic ways to easily compute bespoke formulas but must be screened to prevent cheating when test-taking.

The consistent pattern is that a new technology is introduced, its characteristics enhance both honest and dishonest production

Robert Paul Keener *is a PhD candidate in political science at the University of Tennessee and an adjunct professor at Mars Hill University. He can be reached at rkeener4@vols.utk.edu.*

© The Author(s), 2025. Published by Cambridge University Press on behalf of American Political Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

doi:10.1017/S1049096525000368

https://doi.org/10.1017/S1049096525000368 Published online by Cambridge University Press

of material, and over time academia adapts the technology in ways that allow for honest production while limiting dishonest production. The use of LLMs in academic institutions is no different except that it presents a way for professionals in political science to cheat and remains a task that academia must work through. LLMs provide a great helping hand, able to identify cases that meet certain criteria that can enhance a study, as well as assisting in brainstorming through problems/roadblocks in the material.

Although many academic institutions have different definitions of plagiarism, when one defines it as simply "to steal and pass off (the ideas or words of another) as one's own: use (another's production) without crediting the source" (Merriam-Webster 2019) the line between plagiarism and acceptable use of LLMs is clear. Like Wiki sites, LLMs can provide good background information and brainstorming, but should not be used for direct text development, nor should its background information/brainstorming be used without verification from more reputable sources. The best first step to doing this is to rid the field of inappropriate use of the technology, in this case, that use being LLM plagiarism.

Currently, there is no 100% reliable test for the detection of AI in articles, as the entire purpose of AI is to mimic human behavior to a degree that seems natural, raising the question, "Why should we bother to read something that no one bothered to write?" This does not mean that AI-written articles are without flaws. How ChatGPT and other LLMs work is that they are presented with large amounts of data and then, through pattern recognition, use that data to generate similar outputs on novel inputs. In the case of peer-reviewed political science research, it has been taught what an article looks like but not how to create one without quite a bit of guesswork (Stöffelbauer 2023).

The outputs of ChatGPT have been accurate to a degree. ChatGPT 3.5 was able to pass an MBA exam from Wharton School of Finance (Rosenblatt 2023), passed a CPA exam (Steinhardt 2023), and can translate very well (White 2022). A prior study has shown that without the use of detection methods, LLM-generated articles can easily be published in high-impact medical journals (Khlaif 2023). In some cases, the only reason that undergraduate students were caught using ChatGPT was because their answers were too well-written for their level (Huang 2023). For these reasons, it is safe to say that LLM accuracy is already established, but the detection of AI is still being developed. Turnitin is a popular AI detection tool that many academics use, with 98% of higher education institutions having purchased access. In one study, researchers tested the efficacy of Turnitin, finding it to be inconsistent and unreliable, often determining AI influence in only 54% of the 100% AI-generated work. Furthermore, there are issues with false positives, where work written before the release of ChatGPT is still identified as having AI influence (Perkins et. al. 2023). Although the outputs are inconsistent, this does not mean that Turnitin is a bad tool, only not enough on its own—which is where the benefits of the keyword detection method can be introduced.

Accuracy issues in AI are easily spotted by diligent reviewers normally due to consistency issues, "miraging" citations, and inaccurate data (Howell, Baker, and Stylianopoulos 2023), but what has not been addressed, and what this article seeks to demonstrate, is the use of keywords to find articles that are plagiarized, in the sense that they claim to be written by professionals but are in fact AI-generated. This is now easily detected, but without a high degree of certainty. Similar to how later iterations of AI will exaggerate the remaining flaws in other AI-generated content, the most popular iteration of ChatGPT, ChatGPT 3.5, has flaws that can be used to detect likely AI-generated content. These flaws are words that were overrepresented in the data that "fed" ChatGPT 3.5. If an article uses an abundance of words that are not as commonly used outside of AI, there is enough suspicion to cast doubt on the integrity of the work itself.

Using this keyword detection method, this article shows that the spike in AI-generated manuscripts has almost certainly already compromised the profession and risks the credibility of all academics if it is not regulated.

When LLMs like ChatGPT are "fed" these vast amounts of data, they act as a references when generating new content. For instance, if ChatGPT were fed data mainly from the United Kingdom, even content produced by someone living and working in the United States would include spellings like "colour," "honour," and "mum." Therefore, in this hypothetical world, if an American English speaker were to submit an article for publication that was written in British English, it would be a strong indicator that the work is likely AI-generated and therefore plagiarized.

The keyword detection method used in this article applies the same logic but to a wider degree. The information that was "fed" into the LLM has an overrepresentation of certain keywords and phrases, which leads to the use of these words being overrepresented in articles that were written during the first year that ChatGPT was widely used, 2023, than during the previous 2022. The top five most overrepresented words by ChatGPT writing, relative to regular human writing, are "delve," "tapestry," "vibrant," "landscape," and "realm" (Li 2024)

This introduces the hypothesis:

H1: After 2022, there will be a substantive increase in the amount of political science writing that uses keywords that are overrepresented in AI use.

This is comparable to the sudden spike of articles using the word "terrorism" after the September 11 Attacks, shown in figure 1. After a dramatic, and tragic, event occurred, the study of terrorism was quickly a concern for many academics. Unlike the sudden use of keywords in large language models, after September 11, 2001, the study of terrorism was quickly relevant to the real world and there was a market demand for understanding the phenomenon of terrorism. The word most overrepresented in the most popular LLM is "delve" (Li 2024). As shown in figure 2, the use of the word "delve" increased by roughly 2.7 times between the years 2022 and 2023. For reference, the use of the word "terrorism" increased by less than 2.25 times, despite "terrorism" being relevant to the real world and "delve" only being a word that ChatGPT uses more average than humans.

https://doi.org/10.1017/S1049096525000368 Published online by Cambridge University Press

Figure 1

Annual Use of the Word "Terrorism" in Academic Articles

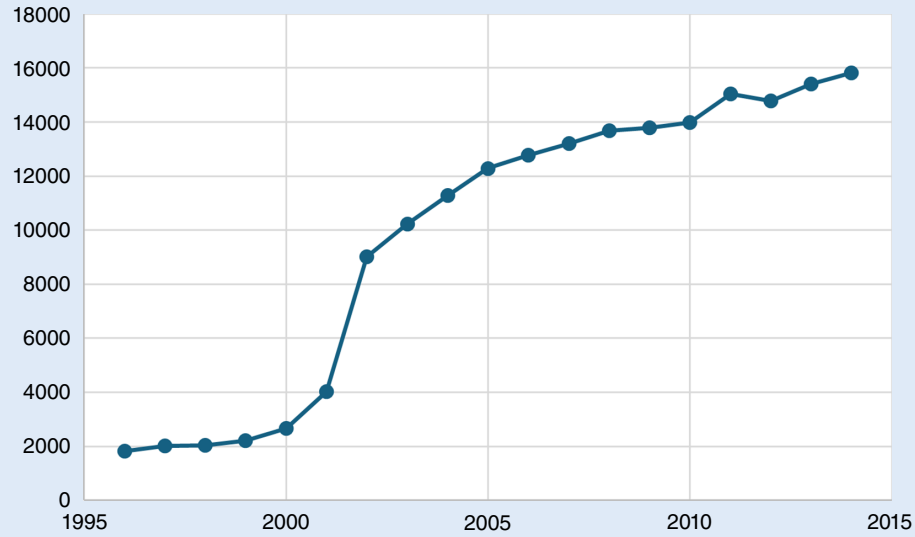
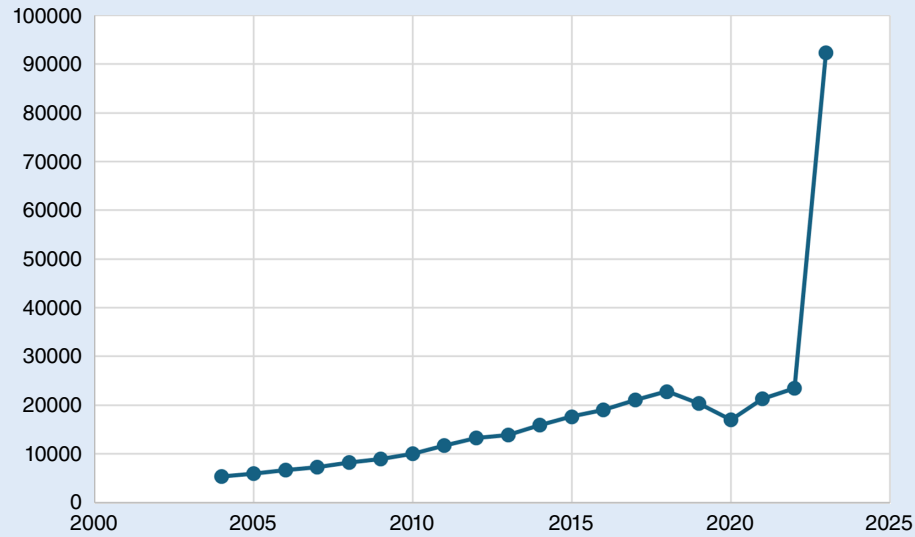


Figure 2

Annual Use of the Word "Delve" in Academic Articles



METHOD

This study uses the keyword method on peer-reviewed political science articles. These articles are sourced from the database OpenAlex (n.d.), an open-access catalog of over 209 million scientific documents, allowing for keyword searches in said documents to chart the popularity of trends over time (Keener 2025). As demonstrated in figure 2, one can see how it charts the popularity of articles on a subject and how the popularity of said subject can grow rapidly after major events. The most popular LLM, ChatGPT, was made available for public use on November 30, 2022, giving most studies written and published in 2022 little to no time to use ChatGPT, and it can be argued that 2022 was the last year where LLM plagiarism was not a concern. This leads to the importance of the years 2022 and 2023 on the four charts shown below.

The results are shown in four charts, the first one being a general baseline as to what the output of political science articles are, simply searching for "political science" articles on OpenAlex. The second chart captures the increase of a single keyword. To

https://doi.org/10.1017/S1049096525000368 Published online by Cambridge University Press

remain as parsimonious as possible, it only shows the increase of the most commonly overused word in ChatGPT, "delve," and how often it is used in political science articles. The final two models show the use of the top three words most overused by ChatGPT ("delve," "tapestry," and "vibrant") in political science articles, and the final model does the same, but with the addition of the fourth and fifth most common keywords ("landscape" and "realm"). There is no formal consensus as to why these are the most overused words in ChatGPT, but almost as soon as ChatGPT was released different hobbyists and bloggers have made lists of words and phrases they find particularly overused, and these lists have been aggregated by Dan Li, the CEO of an AI company, PlusDocs, to make his own list of AI "watchwords," as his own platform has the incentive to avoid resembling ChatGPT as best as he and his employees can (Li 2024).

It is important to note that unlike the use of the word "terrorism" after September 11, 2001, these keywords do not have any significance in major scholarly work in the field of political science, nor are they words that are trendy to humans in other circumstances, such as slang or cultural references, only that they are overrepresented in ChatGPT's outputs relative to human-generated writing.

RESULTS

As expected, figure 3 is necessary to this study precisely for its lack of significance. If there had been a sudden similar increase in articles that referenced political science, a dramatic increase in the keywords would be expected, as it would not mean they are overrepresented in the literature but only growing alongside it. As demonstrated in figure 3, the amount of political science articles increased slightly, meaning that if there was no plagiarism and all things constant, there should be a similar increase in the keywords, the same rate of usage, but drawn from a larger well.

Figure 4 shows that this is not the case and that between the years 2022 and 2023 there was an increase in articles using the words "political science" and "delve" despite there being fewer overall "political science" articles that year. Keyword use accelerates even more quickly when combining keywords, as shown in figures 5 and 6, where despite more keywords being added, there is little difference in the sudden acceleration of articles combining keywords, only a smaller overall sample size as more keywords are introduced.

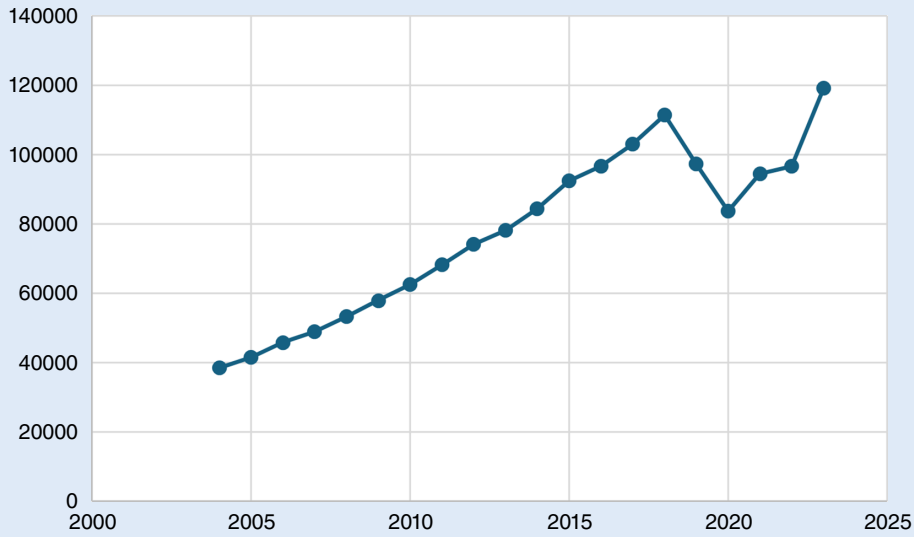
These results show us that there is a high likelihood that political science professionals are using ChatGPT in peer-reviewed work. The use of these keyword searches demonstrates how suspicious work can now be detected, and as later editions of ChatGPT are introduced, new keywords will likely be introduced as well. Although the method remains the same, easily adaptable inputs will change.

These results show us that there is a high likelihood that political science professionals are using ChatGPT in peer-reviewed work.

Side by side, the data are even more concerning, as the percentage of increase in the use of these words is greater than that of "terrorism" in 2002 in almost all cases, as shown in figure 1. When more keywords are added, the sample sizes begin to shrink rapidly, but these are only the most egregious overuses of keywords, and therefore the most suspicious. Other combinations can also be tried to find keywords, such as limiting a search to "Political Science" and three keywords randomly chosen from the top five, but not necessarily the top three—as this would likely create

Figure 3

Annual Use of the Phrase "Political Science" in Academic Articles



https://doi.org/10.1017/S1049096525000368 Published online by Cambridge University Press

Figure 4

Annual Use of the Phrase "Political Science" and the Word "Delve" in Academic Articles

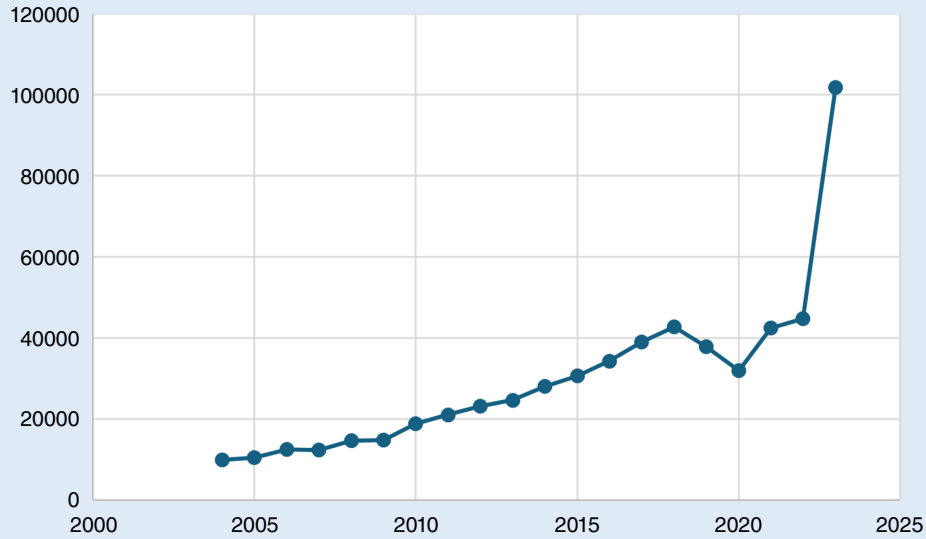
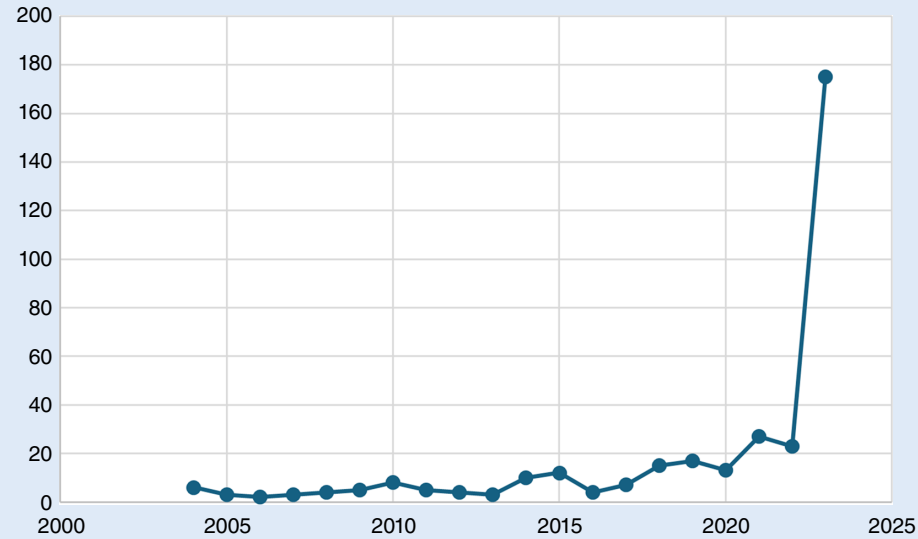


Figure 5

Annual Use of the Phrase "Political Science" and the Top Three Keywords in Academic Articles



another example of the worrying trend in the sudden increase in words preferred by ChatGPT (see table 1).

DISCUSSION

This is not to say that there is no place for AI in political science. In the same sense that undisclosed AI will be a major issue for institutions that do not monitor for it, those who do not use disclosed AI will also be limiting their opportunities for research. New work is being developed that uses AI as a new method for testing theories (Martineau 2021), in the early stages of writing articles or lesson planning, and in brainstorming ideas (White 2022), and LLMs like ChatGPT are powerful translation tools (White 2022). This only becomes an issue when it is generating work that has not been disclosed as having been AI-generated. "Plagiarism" is in itself not plagiarism if the work is being cited (Childers and Burton 2015), and this leads to a wider discussion on what is to be professionally accepted. What complicates the issue greatly is that LLMs like ChatGPT do not disclose where they source their information during the "feeding" process mentioned earlier. This leads to a distinct possibility that ChatGPT itself is

https://doi.org/10.1017/S1049096525000368 Published online by Cambridge University Press

Figure 6

Annual Use of the Phrase "Political Science" and the Top Five Keywords in Academic Articles

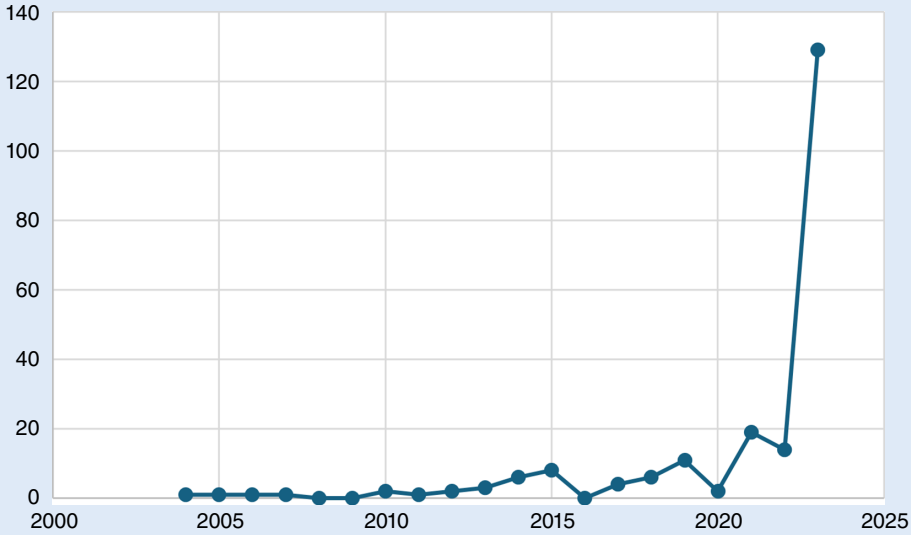


Table 1

Summarized Keyword Increases over Important Years

Measure	Year T	Year T+1	n at T	n at T+1	Percentage of increase
"Terrorism"	2001	2002	4017	9010	124.27%
"Political Science"	2022	2023	96579	119057	23.28%
"Political Science" and "Delve"	2022	2023	4474	10172	127.37%
"Political Science" and top 3 keywords	2022	2023	23	175	660.87%
"Political Science" and top 5 keywords	2022	2023	14	129	821.43%

unintentionally plagiarizing human work. This would mean that even disclosed AI use from human users is still not professionally acceptable because it could easily be the work of others that ChatGPT itself had copied without citing, putting a problematic air over all LLM-generated papers.

Furthermore, some writers use ChatGPT only for outlines, and some professionals use it only for translation. Fortunately, with AI plagiarism checkers and this keyword detection method, finding LLM plagiarism is not difficult. However, it is incumbent to police for this, as editors and reviewers should diligently check for LLM

This would mean that even disclosed AI use from human users is still not professionally acceptable because it could easily be the work of others that ChatGPT itself had copied without citing, putting a problematic air over all LLM-generated papers.

Additionally, there needs to be a large consensus on *how much* AI is too much. In 2021 Microsoft Word introduced text prediction, which suggests sentence completions as they are being written using pattern recognition technology. Spell checks in word processors search for words not in the dictionary, and when a word is misspelled it guesses what word the author was trying to write. Grammarly, a free add-on, allows for even more advanced spell-checking and suggestions, also often finishing sentences as authors write them. Both Microsoft Word and Grammarly are professionally accepted and not considered to be plagiarism. plagiarism before publication to mitigate the need for retraction.

Currently, most professional organizations like APSA have no guidelines for AI plagiarism. APSA's most recent ethics guide was written in February 2022, several months before the open launch of ChatGPT, and does not mention large language models or artificial intelligence (American Political Science Association, 2022). Members of these organizations can work together to define "AI plagiarism" in a way the field can accept, likely distinguishing between LLM plagiarism and nongenerative AIs like Grammarly or Microsoft Word's Text Predictor. In short, LLMs generate new

https://doi.org/10.1017/S1049096525000368 Published online by Cambridge University Press

text, whereas Grammarly and Text Predictor build off what text is already written and offer suggestions.

After professional organizations develop explicit guidelines and definitions of plagiarism, it is incumbent on institutions to apply them to maintain their credibility. When AI plagiarism is detected, institutions are likely to follow their methods for all types of plagiarism. This can mean internal reviews, retractions of articles, greater scrutiny of the offender's entire body of work in the past and future, and punishment spanning from a formal reprimand to termination. As of the time of writing, the credibility of entire institutions is endangered by undisclosed AI-generated content. If the field is not held to a high standard, where AI-generated articles are not only dismissed but there is stigmatization on those who produce them, anyone who claims to be a political scientist could suffer a loss of credibility. Failure to enforce regulations on how AI is used in political science, especially LLMs, weakens the field which hurts all professionals in some way.

CONCLUSION

This article has shown that AI-generated articles have already infested the field, leading to a sudden outbreak of plagiarism that can harm the credibility of the entire field of political science. As mentioned earlier, this is a highly speculative and catastrophized conclusion. What is far more likely to happen is that AI detection methods will be developed and higher education institutions that do not embrace them will suffer and those that do embrace them will thrive. In a sense, reviewers using new, innovative methods to try and detect AI is as important as authors including a "works cited" section.

The keyword detection method of finding suspicious AI work allows reviewers to keep up with these trends and hold authors to the same standards as earlier generations of authors who did not have the option to plagiarize in this way. Holding authors to a high standard, or at least the same standard that has been in place for generations, is only for their and the profession's benefit. There is a substantive demand for professional organizations like APSA to establish guidelines and best practices for defining and identifying AI plagiarism and establishing norms around how it is addressed and punished. Ultimately the field as a whole must pay one of two expenses, either the cost of adapting to the new landscape or the cost of its credibility.

ACKNOWLEDGMENTS

I am grateful to God, and through Him, I am grateful to my wife, Mu Tong, my parents, my advisor Dr. Krista Wiegand, my colleagues Dana Abu Haltam, Matthew Millard, Michael McKoy, and Jeremiah Muhammad, as well as the rest of the political science department at the University of Tennessee.

DATA AVAILABILITY STATEMENT

Research documentation and data that support the findings of this study are openly available at the Harvard Dataverse at https://doi.org/10.7910/DVN/ZIB5NN.

CONFLICTS OF INTEREST

The author declares there are no ethical issues or conflicts of interest in this research. ■

REFERENCES

American Political Science Association. 2022. *A Guide to Professional Ethics in Political Science.*

Childers, Dan, and Sam Bruton. 2015. "'Should It Be Considered Plagiarism?' Student Perceptions of Complex Citation Issues." *Journal of Academic Ethics* 14 (1): 1–17.

Howell, C.W., Cal Baker, and Fayrah Stylianopoulos. 2023. "To Educate Students about AI, Make Them Use It." *Scientific American*, November 14, 2023.

Huang, Kalley. 2023. "Alarmed by A.I. Chatbots, Universities Start Revamping How They Teach." *New York Times*, January 16, 2023, sec. Technology.

Keener, Robert. 2025. "Replication Data for 'More A than I: Testing for Large Language Model Plagiarism in Political Science.'" *PS: Political Science & Politics.* Harvard Dataverse. DOI:10.7910/DVN/ZIB5NN.

Khlaif, Zuheir N., Allam Mousa, Muayad Kamal Hattab, Jamil Itmazi, Amjad A. Hassan, Mageswaran Sanmugam, and Abedalkarim Ayyoub. 2023."The Potential and Concerns of Using AI in Scientific Research: ChatGPT Performance Evaluation." *JMIR Medical Education* 9 (1): e47049. https://doi.org/10.2196/47049.

Li, Daniel. 2024. "The Most Overused ChatGPT Words - Plus." Www.plusdocs.com, April 21. https://www.plusdocs.com/blog/the-most-overused-chatgpt-words.

Martineau, Kim. 2021. "Generative AI Could Offer a Faster Way to Test Theories of How the Universe Works." IBM Research, February 9. https://research.ibm.com/blog/time-series-AI-transformers.

Merriam-Webster. 2019. "Definition of PLAGIARIZING." Merriam-Webster.com. https://www.merriam-webster.com/dictionary/plagiarizing.

Michels, Steven. (2023): Teaching (with) Artificial Intelligence: The Next Twenty Years. *Journal of Political Science Education*. DOI: 10.1080/15512169.2023.2266848

OpenAlex. n.d. Openalex.org. https://openalex.org/.

Perkins, Mike, Jasper Roe, Darius Postma, James McGaughran, and Don Hickerson. 2023. "Detection of GPT-4 Generated Text in Higher Education: Combining Academic Judgement and Software to Identify Generative AI Tool Misuse." *Journal of Academic Ethics* 22 (October).

Rosenblatt, Kalhan. 2023. "ChatGPT Passes MBA Exam given by a Wharton Professor." NBC News, January 23. https://www.nbcnews.com/tech/tech-news/chatgpt-passes-mba-exam-wharton-professor-rcna67036.

Steinhardt, S. J. 2023. "Latest Version of ChatGPT Passed a Practice CPA Exam." Www.nysscpa.org, May 23. https://www.nysscpa.org/article-content/latest-version-of-chatgpt-passed-a-practice-cpa-exam-052323#sthash.8wTF2hfa.dpbs.

Stöffelbauer, Andreas. 2023. "How Large Language Models Work." Data Science at Microsoft, October 24, 2023. https://medium.com/data-science-at-microsoft/how-large-language-models-work-91c362f5b78f.

White, Monica. 2022. "Top 10 Most Insane Things ChatGPT Has Done This Week." Springboard (Blog). December 9, 2022. https://www.springboard.com/blog/news/chatgpt-revolution/.

https://doi.org/10.1017/S1049096525000368 Published online by Cambridge University Press