medRxiv preprint doi: https://doi.org/10.1101/2024.05.14.24307373; this version posted May 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Delving into PubMed Records: Some Terms in Medical Writing Have Drastically Changed after the Arrival of ChatGPT

Kentaro Matsui[1,2*]

[1]Department of Clinical Laboratory, National Center Hospital, National Center of Neurology and Psychiatry, 4-1-1 Ogawa-Higashi, Kodaira-shi, Tokyo, 187-8551, Japan

[2]Department of Sleep-Wake Disorders, National Institute of Mental Health, National Center of Neurology and Psychiatry, 4-1-1 Ogawa-Higashi, Kodaira-shi, Tokyo, 187-8553, Japan

***Corresponding author:**

Kentaro Matsui, MD PhD, Department of Clinical Laboratory, National Center Hospital, National Center of Neurology and Psychiatry, 4-1-1 Ogawa-Higashi, Kodaira-shi, Tokyo, 187-8551, Japan, matsui.kentaro@ncnp.go.jp

Funding: KM is supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI (grant number 22K15778).

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.14.24307373; this version posted May 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Abstract

It is estimated that ChatGPT is already widely used in academic paper writing. This study aims to investigate whether the usage of specific terminologies has increased, focusing on words and phrases frequently reported as overused by ChatGPT. The list of 118 potentially AI-influenced terms was curated based on posts and comments from anonymous ChatGPT users, and 75 common academic phrases were used as controls. PubMed records from 2000 to 2024 (until April) were analyzed to track the frequency of these terms. Usage trends were normalized using a modified Z-score transformation. A linear mixed-effects model was used to compare the usage of potentially AI-influenced terms to common academic phrases over time. A total of 26,403,493 PubMed records were investigated. Among the potentially AI-influenced terms, 75 displayed a meaningful increase (modified Z-score ≥ 3.5) in usage in 2024. The linear mixed-effects model showed a significant effect of potentially AI-influenced terms on usage frequency compared to common academic phrases ($p < 0.001$). The usage of potentially AI-influenced terms showed a noticeable increase starting in 2020. This study revealed that certain words and phrases, such as "delve," "underscore," "meticulous," and "commendable," have been used more frequently in medical and biological fields since the introduction of ChatGPT. The usage rate of these words/phrases has been increasing for several years before the release of ChatGPT, suggesting that ChatGPT might have accelerated the popularity of scientific expressions that were already gaining traction. The identified terms in this study can provide valuable insights for both LLM users, educators, and supervisors in these fields.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.14.24307373; this version posted May 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Author Summary

Artificial intelligence systems have rapidly integrated into academic writing, particularly in the medical and biological fields. This study investigates changes in the frequency of specific terminologies reported as overused by ChatGPT. By analyzing PubMed records from 2000 to 2024, we tracked 118 potentially AI-influenced terms and compared them with 75 common academic phrases. The study's findings reveal that terms such as 'delve,' 'underscore,' 'meticulous,' and 'commendable' saw a marked increase in usage in 2024. However, this trend actually began around 2020. This suggests that while some of these terms were already gaining popularity before the release of ChatGPT, the large language model may have accelerated their adoption in scientific literature. Furthermore, the analysis highlights that the impact of ChatGPT extends beyond new terminologies to altering the frequency and style of commonly used academic phrases. Understanding these trends can help researchers and educators see how AI tools are shaping academic writing.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.14.24307373; this version posted May 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Introduction

ChatGPT rapidly achieved widespread global use after its launch on November 30, 2022. Trained on a vast corpus of text data, the large language model (LLM) including ChatGPT generates natural language with remarkable fluency. Shortly after its release, ChatGPT's applicability for scientific writing in medical and biological fields became evident [1, 2]. Due to the fervor surrounding its capabilities, it was credited as an author on several papers, igniting considerable debate (currently, AI is not acknowledged as an author in scholarly publications [3]). There were even opinions that the use of ChatGPT in paper writing was plagiarism [4], but in reality, LLMs such as ChatGPT, Gemini, and Claude are already being used in paper writing. The use of LLMs can be applied in various ways in academic writing [1, 5] and is also important for the research activities of non-native researchers whose first language is not English [6-8]. Presently, a framework has been established that permits the use of LLMs in writing, provided their involvement is adequately acknowledged [3].

While LLMs can produce natural writing, their output also exhibits certain characteristics [9, 10]. Recently, it became a topic of discussion on X (formerly Twitter) and Reddit that ChatGPT frequently outputs the word 'delve' (https://www.reddit.com/r/mildlyinfuriating/comments/1bzvgqj/apparently_using_the_word_delve_is_a_sign_of_the/ [Accessed 2024, April 12]). In addition, recent reports focusing on detecting text generated by LLMs have identified several frequently used words, such as 'commendable,' 'meticulous,' 'intricate,' and 'realm' [11-14]. The extraction of these characteristic keywords of LLMs in these previous reports was performed by comparing human-generated text with ChatGPT-generated text [11, 13, 14]. While this approach revealed ChatGPT's characteristics among the words commonly used by both humans and ChatGPT, it had methodological limitations in extracting words with low usage frequencies. Clarifying the word expressions that LLMs tend to use in medical and biological papers is crucial for designing academic writing support and medical education programs [15].

medRxiv preprint doi: https://doi.org/10.1101/2024.05.14.24307373; this version posted May 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Moreover, revealing the extent of ChatGPT's impact on papers in the medical and biological fields is essential for maintaining the fairness and reliability of academic research and from the perspective of research ethics [16]. However, the existing literature lacks a thorough investigation of the specific ways in which ChatGPT has transformed academic writing practices in the medical and biological disciplines, necessitating further research.

As the usefulness of LLMs becomes more evident, the number of researchers using LLMs for writing papers has been gradually increasing [11, 13, 17]. It would logically follow that there has been an increase in the number of research reports featuring specific expressions unique to LLMs. This study, therefore, tests the hypothesis that the adoption of certain scientific terminologies has risen following the advent of ChatGPT. Focusing on words and phrases frequently reported as used by ChatGPT, I investigated PubMed records from 2000 onwards and performed a comparison using phrases commonly used in academia as a control. This analysis aims to empirically explore the influence of LLMs on the lexicon of medical literature.

Methods

Search for Records

Unlike earlier studies [12-14], this research, drawing insights from various anonymous end-users, extracted potentially AI-influenced terms from Reddit, X (formerly Twitter), blogs, and forums, focusing on words and phrases frequently produced by LLMs. The selection of these terms was carried out through a rigorous manual curation process from April 12 to May 24, 2024, identifying 118 potentially AI-influenced terms. In addition, as a control group, I used the top 100 collocations identified as characteristic of the academic corpus in a previous study [18]. Phrases that could be searched on PubMed as two consecutive words were included (for example, the collocation "between and" is used in the form of "between A and B," so it was excluded as no records were found when searching for "between and [Text

medRxiv preprint doi: https://doi.org/10.1101/2024.05.14.24307373; this version posted May 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Word]"). In the end, 75 common academic phrases were chosen for verification in this study. The list of these phrases appears in Table 1.

Table 1. Words and phrases examined for usage rates

Potentially AI-influenced terms	
Verbs	
	align, bolster, burgeon, captivate, catalyze, compel, delve, demystify, dive, elevate, elucidate, embark, embrace, employ, encounter, endeavor, enhance, enrich, explore, facilitate, foster, harness, leverage, mitigate, navigate, nuance, optimize, pave, pioneer, resonate, revolutionize, safeguard, shed light, showcase, streamline, underscore, unleash, unlock, unravel, unveil, utilize
Adjectives	
	aesthetic, commendable, comprehensive, crucial, cutting-edge, disruptive, dynamic, essential, ever-evolving, fresh, groundbreaking, holistic, illustrious, imperative, innovative, invaluable, intricate, keen, meticulous, methodical, multifaceted, notable, noteworthy, overarching, paramount, pivotal, poised, potent, robust, seamless, tailored, transformative, unparalleled, unprecedented, unwavering, valuable, versatile, vibrant, vital
Adverbs	
	additionally, aptly, effectively, excellently, impressively, ingeniously, moreover, reportedly, scholarly, strategically, ultimately, undoubtedly
Nouns	
	adventure, beacon, capability, complexity, cornerstone, digital world, driving force, enigma, game-changer, hurdle, interplay, intersection, journey, kaleidoscope, landscape, new era, paradigm shift, prowess, realm, spearhead, state-of-the-art, symphony, synergy, tapestry, testament, treasure trove
Common academic phrases (controls)	
	according to, are presented, are shown, associated with, association between, based on, between groups, can be, changes in, characterized by, compared to, compared with, consistent with, control group, correlated with, correlation between, current study, data set, decrease in, defined as, defined by, determined by, differences between, due to, effect on, exposure to, follow up, have shown, higher than, in addition, in contrast, increase in, indicate that, interaction between, is consistent, no significant, number of, obtained from, occurrence of, our results, our study, over time, participants were, percentage of, present study, prevalence of, previous studies, related to, relationship between, respect to, results obtained, sample size, show that, significant between, significant difference, significantly different, significantly higher, statistically significant, subset of, suggest that, these findings, these results, this article, this paper, this study, to determine, to evaluate, was calculated, was measured, was performed, was used, were collected, were determined, were significantly, with respect

I used PubMed's advanced search feature (https://pubmed.ncbi.nlm.nih.gov/advanced/) to reveal the number of records in which these words were used by searching for "Text Word". To ensure comprehensive coverage of verb forms in English, the search query included the base form, third person singular present, present participle/progressive, past tense, and past participle. For nouns, both singular and plural forms were incorporated. Considering the

medRxiv preprint doi: https://doi.org/10.1101/2024.05.14.24307373; this version posted May 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

daily increase in records indexed in PubMed, the search conditions were standardized from January 1, 2000, to April 30, 2024. The search formulas for all words/phrases are shown in S1 Table.

Data Preparation

To investigate the usage trends of potentially AI-influenced terms in the PubMed database, we first calculated the usage frequency of each term by dividing the number of records containing the term by the total number of records in PubMed for each year from 2000 to 2024 (up to April 30, 2024). This process yielded a dataset with usage frequency for each term and year. Next, the modified Z-score transformation was used to normalize the usage frequency and facilitate comparisons across terms and years. For each term, the median and median absolute deviation (MAD) were calculated. The modified Z-score was computed by subtracting the median from each occurrence rate, dividing the result by the MAD, and multiplying by 0.6745. To identify significant deviations in term usage, we considered an absolute modified Z-score of 3.5 or higher as indicative of a meaningful increase or decrease [19]. The resulting dataset, containing the modified Z-scores for each term and year, was then used for further statistical analysis.

Statistical Analysis

A linear mixed-effects model was used to compare the usage of potentially AI-influenced terms and common academic phrases from 2000 to 2024. The data, consisting of modified Z-scores for each word or phrase, were obtained and reshaped into a long format. The model, constructed using the 'lme' function from the 'nlme' package in R, included the modified Z-

medRxiv preprint doi: https://doi.org/10.1101/2024.05.14.24307373; this version posted May 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

scores as the dependent variable, the group (potentially AI-influenced terms or common academic phrases) as a fixed effect, and a random intercept for each word or phrase to account for repeated measures. The model's summary was generated to assess the significance of the fixed effect of the group on term usage. A line plot with 95% confidence intervals was created using the 'ggplot2' package to visualize the trends in mean usage for each group from 2000 to 2024. The significance level for all statistical tests was set at 0.05. The analysis was performed using R version 4.3.2.

Results

A total of 26,403,493 records between January 1, 2000, and April 30, 2024 were extracted from PubMed. The frequency rates of each word/phrase were determined using the annual total number of records as the denominator, followed by the calculation of the modified Z-score. The Modified Z-score for all the words and phrases across all periods is shown in S2 Table.

In this study, among the 118 potentially AI-influenced terms verified, 75 words/phrases (listed in descending order: 'delve,' 'underscore,' 'meticulous,' 'commendable,' 'showcase,' 'intricate,' 'tapestry,' 'symphony,' 'impressively,' 'realm,' 'cutting-edge,' 'prowess,' 'captivate,' 'noteworthy,' 'groundbreaking,' 'unlock,' 'compel,' 'leverage,' 'notable,' 'unveil,' 'ingeniously,' 'pivotal,' 'bolster,' 'holistic,' 'safeguards,' 'elevate,' 'unwavering,' 'transformative,' 'pioneer,' 'enigma,' 'embark,' 'invaluable,' 'testament,' 'nuance,' 'mitigate,' 'game-changer,' 'valuable,' 'endeavor,' 'imperative,' 'crucial,' 'revolutionize,' 'unleash,' 'effectively,' 'employ,' 'digital world,' 'foster,' 'demystified,' 'multifaceted,' 'navigate,' 'unravel,' 'ever-evolving,' 'streamline,' 'intersection,' 'utilize,' 'harness,' 'shed light,' 'strategically,' 'seamless,' 'encounter,' 'essential,' 'align,' 'additionally,' 'pave,' 'poised,' 'innovative,' 'synergy,' 'comprehensive,' 'burgeon,' 'aptly,' 'dive,' 'unparalleled,' 'ultimately,' 'vital,' 'journey,' 'enhance') displayed a modified Z-score

medRxiv preprint doi: https://doi.org/10.1101/2024.05.14.24307373; this version posted May 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

exceeding 3.5 in 2024. While the majority of the 75 common academic phrases (controls) displayed no significant deviations in usage rates, phrases such as 'occurrence of,' 'these findings,' 'have shown,' 'interaction between,' and 'characterized by' surpassed a modified Z-score of 3.5 in 2024. On the other hand, the phrases 'percentage of,' 'was measured,' 'number of,' 'with respect,' 'respect to,' and 'to determine' registered modified Z-scores below -3.5 in the same year (Fig 1).

[Insert Fig 1 here]

Fig 1. Scatter plot of word/phrase usage frequency vs. modified Z-Score in 2024.

Fig 1 illustrates the relationship between the frequency of use and the modified Z-scores for words and phrases with absolute modified Z-scores exceeding 3.5 in 2024. Red circles represent potentially AI-influenced terms, while grey circles represent common academic phrases (controls). The x-axis shows the number of total records using the words/phrases on a logarithmic scale, and the y-axis displays the modified Z-score for usage frequency.

The linear mixed-effects model revealed a significant effect of the group (potentially AI-influenced terms vs. common academic phrases) on the usage frequency. The model showed that the usage of potentially AI-influenced terms was significantly higher than that of common academic phrases ($\beta = 0.552$, $SE = 0.079$, $t(191) = 6.958$, $p < 0.001$). The line plot (Fig 2) illustrates the trends in mean frequency for potentially AI-influenced terms and common academic phrases from 2000 to 2024. While the frequency of the control group remains relatively stable, the potentially AI-influenced terms begin to show an increase around 2016, with a notable and steep upward trajectory starting in 2020 that becomes particularly pronounced in 2023 and 2024.

[Insert Fig 2 here]

medRxiv preprint doi: https://doi.org/10.1101/2024.05.14.24307373; this version posted May 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Fig 2. Mean usage (modified Z-scores) of potentially AI-influenced terms and common academic phrases from 2000 to 2024. Shaded areas represent 95% confidence intervals.

Discussion

This study demonstrated that, in the fields of medicine and biology, a number of specific words and phrases, led by "delve," "underscore," "meticulous," and "commendable," have come to be used more frequently following the advent of ChatGPT. The increasing trend in the usage rates of these words/phrases was more pronounced in 2024 than in 2023 in almost all cases. This may reflect the generalization of LLM use among researchers in the fields of medicine and biology, as shown in previous findings [13]. The list of overused terms suggested in this study will help those writing with LLMs centered around ChatGPT.

It has been observed that medical texts generated by ChatGPT, while fluent and logical, tend to include less specific information and more generalized expressions compared to those authored by humans, which feature a richer and more diverse content [10]. In general papers, it has been noted that ChatGPT tends to 1) use the same style and expressions repeatedly, 2) show a decrease in the frequency of basic verbs like 'is' and 'are,' and 3) frequently use adjectives and adverbs [11]. Particularly for adjectives and adverbs, numerous words that ChatGPT frequently uses have been pointed out [14]. Because this study only counted the records where specific words or phrases occurred, it did not evaluate the weight of terms appearing multiple times. In the current study, several words that were previously identified as frequently used by ChatGPT did not exhibit a notable increase in usage; yet, ChatGPT may actually overuse these words more than suggested by the results of this study. Similarly, frequently used verbs such as 'enhance', 'elevate', and 'utilize' may also have been overused by LLMs more than suggested by this study.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.14.24307373; this version posted May 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

A previous reports' limitation lies in their lack of focus on the specific words or terms overused by ChatGPT, thus failing to comprehensively explore characteristic terms. As extensively debated online (https://www.reddit.com/r/mildlyinfuriating/comments/1bzvgqj/apparently_using_the_word_delve_is_a_sign_of_the/ [Accessed 2024, April 12]), the increased usage of the word 'delve' was incredibly pronounced compared to other words or phrases, with a modified Z-score of around 100. Despite its overwhelming presence, previous papers compared texts created by humans and ChatGPT [12-14] did not mention 'delve,' highlighting a strength of this study's methodology. This study cannot conclusively establish the connection between the frequent use of 'delve' and the emergence of ChatGPT, although its impact is highly suspected. The frequent use of the term 'delve' by ChatGPT could be attributed to its prominence in the training data, possibly resulting from common instructions during the reinforcement learning from human feedback phase, or as a feature of large language models designed to project authority; however, these hypotheses remain speculative and unconfirmed.

Notably, the frequency of use for the potentially AI-influenced terms investigated in this study had already diverged markedly even before ChatGPT released in November 2022. In particular, 'delve' was used exceptionally often in 2024, but it had also been used at an exceptionally high frequency in academic writing since 2021. One hypothesis arises from this observation—the increasing popularity of these terms in scientific writing may have influenced the output of ChatGPT. Also, ChatGPT's training process might have further reinforced their usage, possibly creating a bidirectional causal relationship. In other words, ChatGPT may have accelerated the inevitable temporal changes in writing in research. However, this hypothesis would be difficult to verify, since we cannot observe a parallel universe where ChatGPT does not exist.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.14.24307373; this version posted May 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Interestingly, some of the common academic phrases used as controls also deviated in their proportion of use in 2024. The four phrases 'occurrence of', 'these findings', 'have shown', and 'interaction between' significantly increased in frequency of use in 2024, but since they are all very commonly used expressions in academic writing, it would be difficult for us humans to recognize that their frequency has increased. Conversely, the six phrases 'percentage of', 'was measured', 'number of', 'with respect', 'respect to', and 'to determine' notably decreased in usage in 2024. When interpreting these results, we must remember that the language used in papers naturally evolves over time [20]; many phrases that decreased in frequency had already been declining even before the introduction of ChatGPT. However, the two phrases 'to determine' and 'number of' did not show a noticeable decrease in frequency of use before 2022, and their frequency of use appears to have decreased significantly after 2023 (see S2 Table). While this result may be coincidental, it could also indicate that the proliferation of ChatGPT may subtly lead to the decreased use of certain words or phrases without us recognizing it.

This study has some limitations. The most important limitation is that the terms potentially influenced by LLMs in this study were identified through manual inspection rather than being extracted in an objective and systematic manner. Therefore, there may be words or phrases that were not included in this study yet have seen a significant increase in usage post-ChatGPT. Additionally, temporal shifts in the frequency of word or phrase use could have been influenced by external factors such as evolving research trends and shifts in the style of scientific communication, factors not accounted for in this study. Another limitation is the application of the Modified Z-score to the time series data. As each word or phrase has only 25 data points (one for each year from 2000 to 2024), the Modified Z-score, which is primarily designed for detecting outliers, may have limited applicability in this context. The small number of data points also makes it challenging to apply certain statistical methods for trend analysis or to compare the usage frequencies between specific years (e.g.,

medRxiv preprint doi: https://doi.org/10.1101/2024.05.14.24307373; this version posted May 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

2023 and 2024). Furthermore, the study may have been influenced by the presence of certain proper nouns. For example, there is a service called "Microsoft Delve." A PubMed search for ("microsoft"[Text Word]) AND ("delve"[Text Word]) yielded only three hits (Accessed 2024, May 24), suggesting that the impact of "Microsoft Delve" on the results is likely to be minimal. However, there remains a possibility that other unforeseen proper nouns could have influenced the usage frequency of some terms. Lastly, the absence of long-term trend analysis limits our ability to fully assess the impact of AI on language usage. Particularly, since the data for 2024 is limited to April, it cannot be denied that the results may fluctuate when looking at the whole year

Conclusion

This study highlights the overuse of specific words and phrases that have become more prevalent since the introduction of ChatGPT. The list of selected terms discussed in this study can be advantageous for both users employing LLMs for writing purposes and for individuals in educational and supervisory capacities within the fields of medicine and biology. However, the changes in academic writing suggested by this study may be temporary and specific to 2024; as LLMs improve, distinguishing between human and AI-generated text may become more challenging [21]. Thus, further studies on future changes in academic terminology are warranted. Many researchers are expected to continue using LLMs for their writing—needless to say, adhering to ethical aspects and taking responsibility for the final output is crucial for the authors when using these tools.

Acknowledgements

During the preparation of this work, the author used GPT-4, GPT-4o and Claude 3 Opus for drafting R code, proofreading the manuscript, and improving the readability of the text.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.14.24307373; this version posted May 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

After using these services, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References

1. Biswas S. ChatGPT and the Future of Medical Writing. Radiology. 2023;307(2):e223312. Epub 2023/02/03. doi: 10.1148/radiol.223312. PubMed PMID: 36728748.

2. Salvagno M, Taccone FS, Gerli AG. Can artificial intelligence help for scientific writing? Critical care (London, England). 2023;27(1):75. Epub 2023/02/26. doi: 10.1186/s13054-023-04380-2. PubMed PMID: 36841840; PubMed Central PMCID: PMCPMC9960412.

3. Lin Z. Towards an AI policy framework in scholarly publishing. Trends Cogn Sci. 2024;28(2):85-8. Epub 2024/01/10. doi: 10.1016/j.tics.2023.12.002. PubMed PMID: 38195365.

4. Thorp HH. ChatGPT is fun, but not an author. Science (New York, NY). 2023;379(6630):313. Epub 2023/01/27. doi: 10.1126/science.adg7879. PubMed PMID: 36701446.

5. Lin Z. Techniques for supercharging academic writing with generative AI. Nat Biomed Eng. 2024. Epub 2024/03/19. doi: 10.1038/s41551-024-01185-8. PubMed PMID: 38499642.

6. Berdejo-Espinola V, Amano T. AI tools can improve equity in science. Science (New York, NY). 2023;379(6636):991. Epub 2023/03/10. doi: 10.1126/science.adg9714. PubMed PMID: 36893248.

7. Hwang SI, Lim JS, Lee RW, Matsui Y, Iguchi T, Hiraki T, et al. Is ChatGPT a "Fire of Prometheus" for Non-Native English-Speaking Researchers in Academic Writing? Korean J Radiol. 2023;24(10):952-9. Epub 2023/10/05. doi: 10.3348/kjr.2023.0773. PubMed PMID: 37793668; PubMed Central PMCID: PMCPMC10550740.

8. Matsui K, Koda M, Yoshida K. Implications of Nonhuman "Authors". Jama. 2023;330(6):566. Epub 2023/08/08. doi: 10.1001/jama.2023.10568. PubMed PMID: 37552501.

9. AlAfnan MA, MohdZuki SF. Do artificial intelligence chatbots have a writing style? An investigation into the stylistic features of ChatGPT-4. Journal of Artificial intelligence and technology. 2023;3(3):85-94.

10. Liao W, Liu Z, Dai H, Xu S, Wu Z, Zhang Y, et al. Differentiating ChatGPT-Generated and Human-Written Medical Texts: Quantitative Study. JMIR Med Educ. 2023;9:e48904. Epub 2023/12/28. doi: 10.2196/48904. PubMed PMID: 38153785; PubMed Central PMCID: PMCPMC10784984.

11. Geng M, Trotta R. Is ChatGPT Transforming Academics' Writing Style? arXiv preprint arXiv:240408627. 2024.

12. Gray A. ChatGPT" contamination": estimating the prevalence of LLMs in the scholarly literature. arXiv preprint arXiv:240316887. 2024.

13. Liang W, Zhang Y, Wu Z, Lepp H, Ji W, Zhao X, et al. Mapping the increasing use of llms in scientific papers. arXiv preprint arXiv:240401268. 2024.

14. Liang W, Izzo Z, Zhang Y, Lepp H, Cao H, Zhao X, et al. Monitoring ai-modified content at scale: A case study on the impact of chatgpt on ai conference peer reviews. arXiv preprint arXiv:240307183. 2024.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.14.24307373; this version posted May 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

15. Song C, Song Y. Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. Frontiers in psychology. 2023;14:1260843. Epub 2024/01/02. doi: 10.3389/fpsyg.2023.1260843. PubMed PMID: 38162975; PubMed Central PMCID: PMCPMC10754989.

16. Jeyaraman M, Ramasubramanian S, Balaji S, Jeyaraman N, Nallakumarasamy A, Sharma S. ChatGPT in action: Harnessing artificial intelligence potential and addressing ethical challenges in medicine, education, and scientific research. World J Methodol. 2023;13(4):170-8. Epub 2023/09/29. doi: 10.5662/wjm.v13.i4.170. PubMed PMID: 37771867; PubMed Central PMCID: PMCPMC10523250.

17. Cheng H, Sheng B, Lee A, Chaudhary V, Atanasov AG, Liu N, et al. Have AI-Generated Texts from LLM Infiltrated the Realm of Scientific Writing? A Large-Scale Analysis of Preprint Platforms. bioRxiv. 2024:2024-03.

18. Durrant P. Investigating the viability of a collocation list for students of English for academic purposes. English for Specific Purposes. 2009;28(3):157-69.

19. Crosby T. How to detect and handle outliers. Taylor & Francis; 1994.

20. Park M, Leahey E, Funk RJ. Papers and patents are becoming less disruptive over time. Nature. 2023;613(7942):138-44. Epub 2023/01/05. doi: 10.1038/s41586-022-05543-x. PubMed PMID: 36600070.

21. Herbold S, Hautli-Janisz A, Heuer U, Kikteva Z, Trautsch A. A large-scale comparison of human-written versus ChatGPT-generated essays. Scientific reports. 2023;13(1):18617. Epub 2023/10/31. doi: 10.1038/s41598-023-45644-9. PubMed PMID: 37903836; PubMed Central PMCID: PMCPMC10616290.

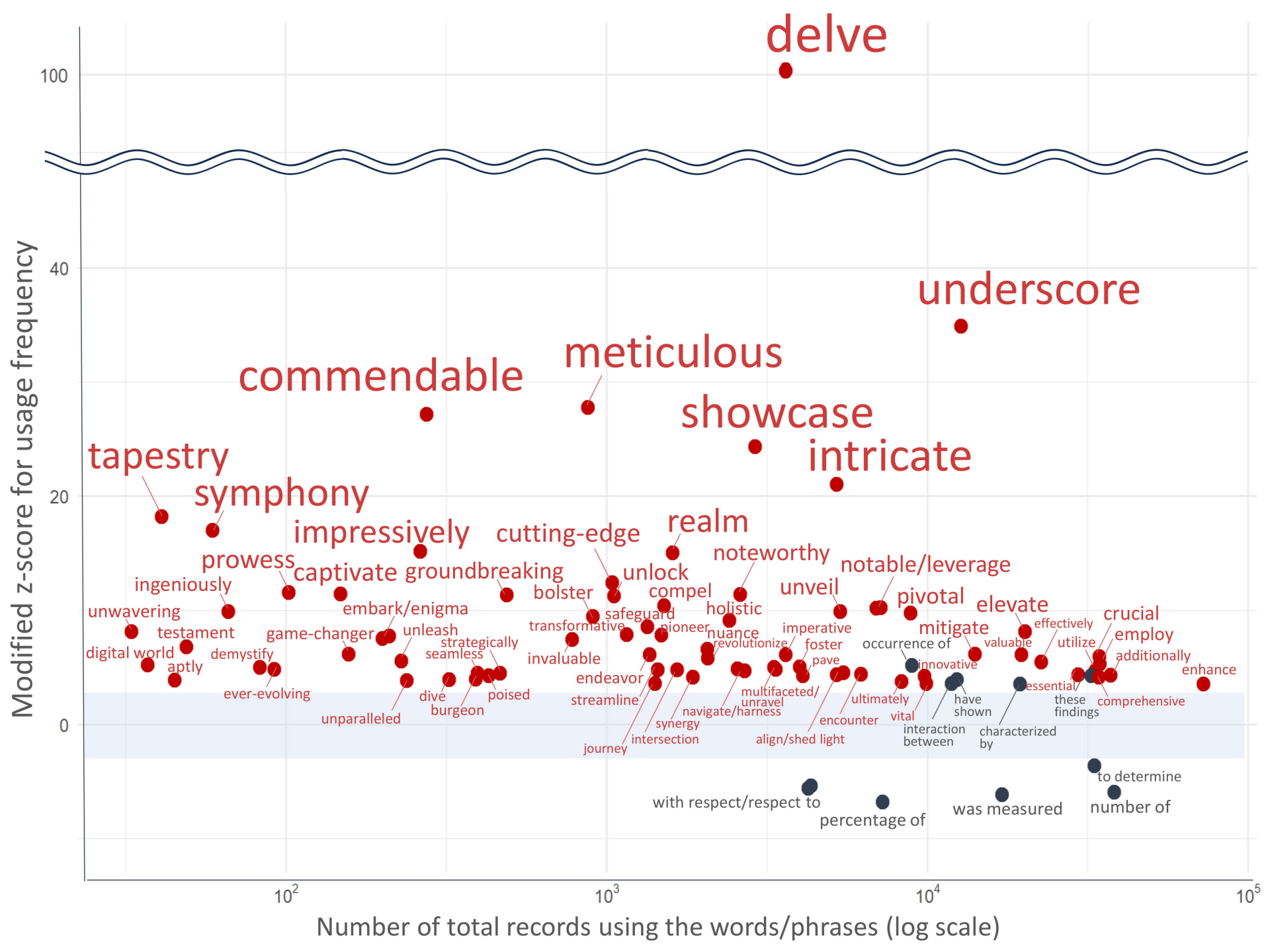
medRxiv preprint doi: https://doi.org/10.1101/2024.05.14.24307373; this version posted May 24, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

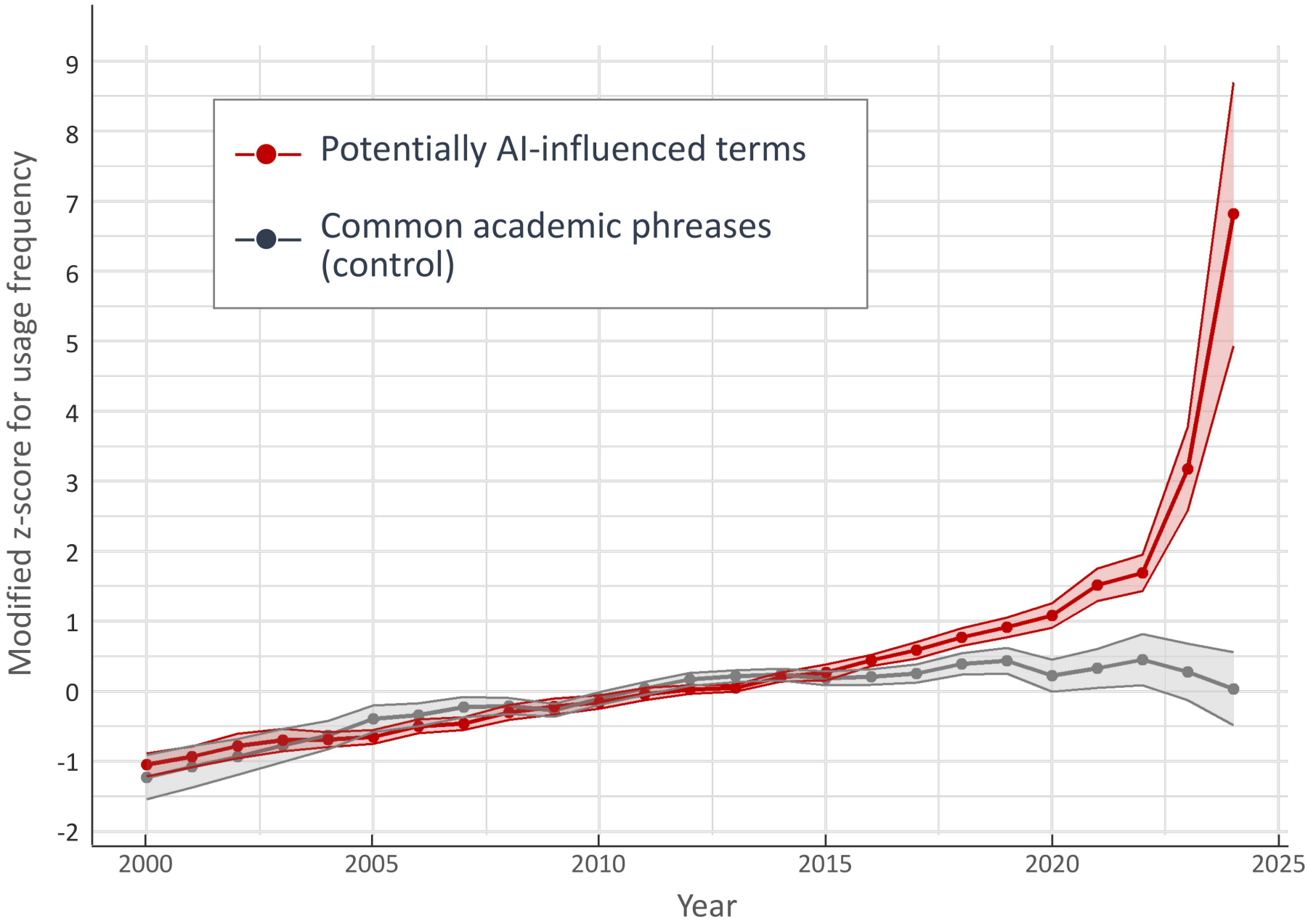
Supporting information

S1 Table. PubMed search formulas for examined words and phrases

S2 Table. Modified Z-score for usage frequency of each word/phrase from 2000 to 2024

S3 Data. The list of the total number of records per year, the number of records in which each word or phrase is used, and the code for identifying potentially AI-influenced terms.


delve
underscore
meticulous
commendable
showcase
intricate
tapestry
symphony
impressively
cutting-edge
realm
noteworthy
notable/leverage
prowess
captivate
groundbreaking
unlock
ingeniously
bolster
compel
unveil
pivotal
elevate
unwavering
embark/enigma
holistic
crucial
testament
game-changer
unleash
transformative
safeguard
pioneer
nuance
imperative
mitigate
effectively
employ
digital world
demystify
strategically
seamless
invaluable
revolutionize
foster
occurrence of
valuable
utilize
additionally
aptly
pave
innovative
endeavor
enhance
ever-evolving
dive
poised
streamline
multifaceted/
unravel
ultimately
essential/
these
findings
comprehensive
unparalleled
burgeon
navigate/harness
have
shown
vital
synergy
encounter
interaction
between
characterized
by
intersection
journey
align/shed light
to determine
with respect/respect to
percentage of
was measured
number of
100
40
20
0
Modified z-score for usage frequency
10^2
10^3
10^4
10^5
Number of total records using the words/phrases (log scale)


Potentially AI-influenced terms
Common academic phreases (control)
Modified z-score for usage frequency
9
8
7
6
5
4
3
2
1
0
-1
-2
2000
2005
2010
2015
2020
2025
Year