Check for updates

*Learned Publishing*

LEARNED PUBLISHING Open Access

CASE STUDY OPEN ACCESS

# Can AI Express Sentiment Like Humans in Research Abstracts? Implications for Academic Publishing

Ying Wei[1] | Mingxin Yao[2]

[1]School of Foreign Studies, Nanjing University, Nanjing, Jiangsu Province, People's Republic of China | [2]School of Business English, Sichuan International Studies University, Chongqing, People's Republic of China

**Correspondence:** Mingxin Yao (yao_mingx@163.com)

**Received:** 26 March 2026 | **Revised:** 15 September 2026 | **Accepted:** 7 November 2026

**Keywords:** abstracts | academic evaluation | AI | human | knowledge production | linguistic positivity

## ABSTRACT

Sentiment is widely examined in academic and non-academic contexts. However, with the advancement of generative artificial intelligence, no study has compared linguistic positivity in AI-generated and human-written academic texts. Adopting sentiment analysis, this study compares the sentiment scores of 420 research abstracts published in top journals across 14 hard and soft disciplines with 420 abstracts generated by DeepSeek. Our findings show that DeepSeek-generated abstracts and human-written journal abstracts show no significant difference in terms of their sentiment distribution. This pattern is observed across hard and the majority of the soft disciplines. These results indicate that AI-generated abstracts closely approximate human-written abstracts in linguistic positivity, as shown in lexicon-based sentiment scores. Given the findings, this study envisions human-AI collaboration in future knowledge production and calls for adjustments to current frameworks of academic evaluation.

## 1 | Introduction

Linguistic positivity bias refers to a general trend of adopting more positive than negative words in communication (Rozin et al. 2010; Augustine et al. 2011; Garcia et al. 2012; Dodds et al. 2015). Following these early discoveries in natural communication, linguistic positivity bias is identified and observed in academic writing (Vinkers et al. 2015; Holtz et al. 2017; Weidmann et al. 2018; Lerchenmueller et al. 2019; Cao et al. 2021; Wen and Lei 2022; Yuan and Yao 2022; Liu and Zhu 2023, 2025; Chen 2024; Yuan et al. 2025).

Originally, linguistic positivity bias can be traced back to Boucher and Osgood's (1969) work, which first put forth the Pollyanna hypothesis, pointing to a possible tendency to use positive words in human communication. The hypothesis was later adapted by Matlin and Stang (1978) as the Pollyanna principle. Following these preliminary findings, later research adopts this tendency as universal in language (Rozin et al. 2010; Augustine et al. 2011; Garcia et al. 2012; Dodds et al. 2015). For instance, in Rozin et al.'s (2010) research, researchers document a general tendency to adopt more positive words in 20 natural languages. Garcia et al. (2012) study linguistic positivity in three European languages and find a general trend to use more positive words than negative ones across three languages.

The above findings inspire researchers to further explore whether linguistic positivity bias exists in academic writing. Early studies such as Vinkers et al. (2015) start this line of research by exploring the distribution of positive and negative words in medical articles published between 1974 and 2014, retrieved from the PubMed database. Results of their study indicate a significant increase in the use of positive words and a relatively moderate increase in the use of negative ones in medical writing across time, which serve as the initial support for linguistic positivity bias in academic writing. Following Vinkers et al. (2015) and Weidmann et al. (2018) examine linguistic positivity bias in the abstracts of political science journals published

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made. All other rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies.

© 2026 The Author(s). *Learned Publishing* published by John Wiley & Sons Ltd on behalf of ALPSP.

**Highlights**

- This article compares the sentiment scores of research abstracts written by human writers and DeepSeek, across hard and soft disciplines.
- The findings show that DeepSeek-generated abstracts exhibit no significant difference from human-written abstracts in terms of sentiment scores, and this pattern is observed across disciplines.
- This finding sheds light on the contemporary landscape of research publishing and may attract attention to human-AI collaboration in knowledge production.

in the period of 1977–2014 and discover a similar increase in the use of positive words, based on Vinkers et al.'s (2015) list of 50 positive/negative words. However, they are limited in that much of their conclusion is derived from an analysis of Vinkers et al.'s (2015) small list, which consists of only 50 words.

To address the limitation of a small word-list approach, Wen and Lei (2022) first introduce large-lexicon-based sentiment analysis into the study of linguistic positivity bias in academic writing. Large-lexicon-based sentiment analysis is an approach to analyzing linguistic data by assigning sentiment scores to the sentences based on the calculation of sentiment words and valence shifters present in the texts. Following Wen and Lei (2022), more research applies sentiment analysis to analyze linguistic positivity bias in academic writing (Yuan and Yao 2022; Liu and Zhu 2023, 2025). For instance, Liu and Zhu (2023) apply sentiment analysis to the analysis of academic writing across disciplines. Specifically, they perform sentiment analysis on a 111-million-word corpus of research abstracts collected from the Web of Science, and compare the development of sentiment scores across eight disciplines including four disciplines from soft science and four from hard sciences. The results shed light on the disciplinary differences of linguistic positivity bias in academic writing by revealing "a higher and faster-growing degree of linguistic positivity in hard disciplines, compared to soft disciplines", adding new and more nuanced understanding to the field.

However, existing studies tend to follow a diachronic endeavour, with little attention directed to comparative research, especially that on AI models and human writers. With the advancement of artificial intelligence, comparison of AI-generated content with human-produced content has already received much scholarly attention (Goulart et al. 2024; Jiang and Hyland 2024, 2025a, 2025b, 2025c; Kong and Liu 2024; Lin and Crosthwaite 2024; Markowitz et al. 2024; Mo and Crosthwaite 2025; Nañola et al. 2025; Won et al. 2025; Yang et al. 2024; Zhang and Crosthwaite 2025). These studies compare the similarities and differences between AI and human writing by examining a wide range of linguistic features, such as lexical bundles (Jiang and Hyland 2024) or rhetorical moves (Kong and Liu 2024).

Some of these studies deepen our understanding of how AI models write differently from students (Jiang and Hyland 2024; Kong and Liu 2024; Nañola et al. 2025; Won et al. 2025; Yang et al. 2024; Zhang and Crosthwaite 2025). Findings of such research yield significant pedagogical implications on EAP teaching and research. For instance, Jiang and Hyland (2024) compare essays written by British students with ChatGPT-generated essays in terms of how they are different in the use of lexical bundles. They observe that ChatGPT tends to use more noun and preposition-based lexical bundles than L2 students. However, student essays exhibit a higher level of persuasion by the use of epistemic stance markers and authorial presence markers.

Other studies focus on how AI models can be improved by comparing the differences of their writing with human writers' writing, especially how to make AI's writing more interactive and engaging (Jiang and Hyland 2025a, 2025b, 2025c; Mo and Crosthwaite 2025). For instance, Mo and Crosthwaite (2025) study student essays and compare the results with three AI models in terms of stance expression and engagement markers. They note that AI models tend to adopt more repetitive stance choices than students, suggesting that when it comes to expressing stance and engagement, AI models still have much to learn from human writers. Finally, there are studies that compare AI models' writing with humans in contexts other than student writing, including hotel reviews (Markowitz et al. 2024), teachers' feedback (Lin and Crosthwaite 2024) and journal writing (Kong and Liu 2024). For instance, Markowitz et al. (2024) study the differences and similarities between AI's output and human-written hotel reviews online. They find that ChatGPT's text is generally more analytical and descriptive but also less readable than reviews written by human users.

However, comparative studies between AI writing and human writing are limited in two ways. First, very few studies compare AI-generated academic writing with academic writing in journal articles. Second, to date, no study has compared the linguistic positivity in academic texts produced by AI and written by humans. The study responds to the knowledge gap about whether AI models generate academic texts more positively than human writers by examining the sentiment scores of 420 research abstracts collected from top journals across 14 disciplines and comparing them with 420 AI-generated abstracts. Findings of this research may generate implications for researchers, peer reviewers, editors and AI developers.

Finally, three related but distinct concepts that recur throughout this paper should be further clarified. First, the term sentiment is adopted in this study in its computational sense, referring to the evaluative polarity of a text as operationalized by a lexicon-based sentiment score. It is a technical measure of the positive or negative valence of words and phrases (Rinker 2019; Jockers 2017). Second, linguistic positivity denotes the broader communicative tendency, documented since Boucher and Osgood (1969), for language users to employ more positive than negative words. In this sense, a sentiment score could be understood as one operationalization of this tendency. Third, affective expression (or emotion) refers to the psychological reality of human feeling, evaluative intent and disciplinary rhetorical stance (Martin and White 2005). Therefore, it should be acknowledged that a lexicon-based sentiment score captures only one particular dimension of evaluative language and therefore does not directly measure human emotion, judgment or disciplinary affect. Accordingly, throughout this paper, similarity



17414857, 2027, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/leap.2126 by Université Paul-Valéry-Montpellier, Wiley Online Library on [22/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

in sentiment scores between AI and human-written abstracts is interpreted as one manifestation of linguistic positivity, rather than as evidence of affective or emotional mastery.

This comparative study compares the linguistic positivity between AI-generated abstracts and journal abstracts and intends to answer the questions below:

**RQ1**. Does AI compose abstracts more positively than journal abstracts in general?

**RQ2**. Does AI compose abstracts more positively than journal abstracts in hard disciplines?

**RQ3**. Does AI compose abstracts more positively than journal abstracts in soft disciplines?

## 2 | Methodology

In this section, we present the methodology of the study by demonstrating how the datasets are compiled in the data collection section and how the data is handled in the data analysis section.

### 2.1 | Data Collection

To address the research questions, this study built two corpora or datasets for comparison, namely the journal abstract dataset representing human writing and the AI-generated abstract dataset. The journal abstract dataset includes 420 research abstracts collected from 14 disciplines, seven from hard disciplines and seven from soft disciplines. Selection of disciplines followed the classification based on Becher (1994) and Shapin (2022), with reference to Zhu et al.'s (2024) list of disciplines. As a result, the study examined seven hard disciplines, including biology (BIO), computer science (CMS), civil engineering (CVE), geology (GEO), mechanics (MEC), physics (PHY) and telecommunication (TEL) as well as seven soft disciplines, including communication (COM), education (EDU), history (HSY), linguistics (LIN), political science (POL), public administration (PBM) and sociology (SOC). Note that medicine (MED) and law (LAW) are not included in our examined disciplines because initial screening showed that in these two disciplines, research abstracts tend to follow a different format from the rest of the disciplines.

Next, with reference to JCR—impact factors and rankings 2023, we selected the Top 1 journal from the ranking of each discipline. The first 30 research articles published in these journals in 2025 are collected for later analysis. It should also be noted that, since our examination is largely based on the article type of 'research articles,' journals that solely published review articles are excluded, such as *Physics of Life Reviews*. In addition, journals with a low publication number are also excluded because they have not published 10 articles by the time of data collection. In the above cases, lower ranked journals are selected as replacements. In sum, the study only selected journals that publish 'research articles' and have already published more than 10 articles in 2025 as data for analysis. Finally, 30 research articles from the following 14 journals are collected, as shown in Table 1.

**TABLE 1** | Disciplines and journals examined by this study.

| Disciplines | Journal name | No. of articles |
|---|---|---|
| Hard disciplines | | 210 |
| Biology | *Current Biology* | 30 |
| Computer Science | *IEEE Transactions on Pattern Analysis and Machine Intelligence* | 30 |
| Civil Engineering | *Automation in Construction* | 30 |
| Geology | *Geology* | 30 |
| Mechanics | *Energy Conversion and Management* | 30 |
| Physics | *Nature Materials* | 30 |
| Telecommunications | *IEEE Communications Surveys and Tutorials* | 30 |
| Soft disciplines | | 210 |
| Communication | *Journal of Communication* | 30 |
| Education | *International Journal of Educational Technology in Higher Education* | 30 |
| History | *Contemporary European History* | 30 |
| Linguistics | *Computer Assisted Language Learning* | 30 |
| Political Science | *Policy and Society* | 30 |
| Public Administration | *Public Administration Review* | 30 |
| Sociology | *Annals of Tourism Research* | 30 |

Third, to compile the dataset for AI-generated abstracts, we manually removed the abstracts in the journal abstract dataset before feeding them to the AI model—DeepSeek, following Kong and Liu's (2024) study. The following prompt is given to DeepSeek to generate abstracts for each article:

1. *You are an expert academic writer specializing in [discipline]. I will provide a research paper text in the txt. files. Please analyze the full content to understand its research objectives, methodology, findings and contributions.*
2. *Note that in these txt. files, article contents do not include a summary or abstract*



17414857, 2027, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/leap.2126 by Université Paul-Valéry-Montpellier, Wiley Online Library on [22/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3. *Based on the provided research paper, generate a structured academic abstract that includes*: (1) *research purpose and gap*, (2) *methodology*, (3) *key findings and* (4) *implications. The abstract must be written in formal academic English and strictly adhere to word count between* [@] *and* [@] *words.*
4. *During all generation of the abstracts, set the temperature to the value of 0 to ensure that the outputs are consistent and replicable as well as that random variability across generations is eliminated.*

All abstracts were generated via the official DeepSeek API (DeepSeek-V3.1) on 16 September 2025. Following the prompt, the temperature was set to 0, and all other generation parameters were kept at their default values to ensure consistent and replicable outputs. The following python function was coded to call access DeepSeek API.

```
    def AI_promot(prompt):
        try:
            response = client.chat.completions.
create(
                model="deepseek-chat",
                messages=[
                    {"role": "user", "content":
prompt},
                ],
                stream=False
            )
            return response.choices[0].message.
content
        except Exception as e:
            print(f"API error: {e}")
            return "Error"
```

Given that the word count range may differ between hard disciplines and soft disciplines, a python script was coded to calculate the range of word counts for each discipline and the results show that in hard disciplines, abstracts generally range from 112 to 550 words while the number is lower in soft disciplines, ranging between 90 and 260 words. Such information is also given to DeepSeek as reference, as shown in the above prompts. Finally, DeepSeek generated 420 abstracts correspondingly across 14 disciplines.

## 2.2 | Data Analysis

Data analysis generally follows the procedures proposed in Figure 1. The first step is corpus preparation, which was complete during data collection. Before data analysis, we coded a python script to delete any non-English symbols in the datasets, using a regular expression. Once data cleaning is complete, we ran a sentiment analysis to calculate the sentiment scores of each step. Sentiment analysis generally follows previous studies, such as Yuan and Yao (2022), Wen and Lei (2022) and Liu and Zhu (2023). To be specific, a lexicon-based sentiment analysis is performed in R on the datasets, with the help of the R package—Sentimentr (Rinker 2019). In Sentimentr, the Jockers sentiment lexicon (Jockers 2017) is adopted to measure the sentiment scores. The sentiment scoring is performed on sentence levels. Scores are initially assigned to each sentence before being added up as raw scores for all sentences in one abstract. Since the aim of the research is to compare the results of two datasets, standardization of the sentiment scores is not needed, as the same procedure is administered to both the journal abstract dataset and the AI-generated abstract dataset.

Finally, for statistical analysis, a Mann–Whitney *U* test (guided by Zhu et al. 2024) is administered to the results of the two datasets to detect any statistical difference in the sentiment scores between journal abstracts and DeepSeek-generated abstracts. Note that a non-parametric test was adopted because the sentiment scores are not normally distributed but positively skewed with a long right tail, as reflected in the wide and asymmetric ranges of the scores (journal abstracts: −2.48 to 5.57; DeepSeek-generated abstracts: −1.2 to 3.95). Consequently, the median was

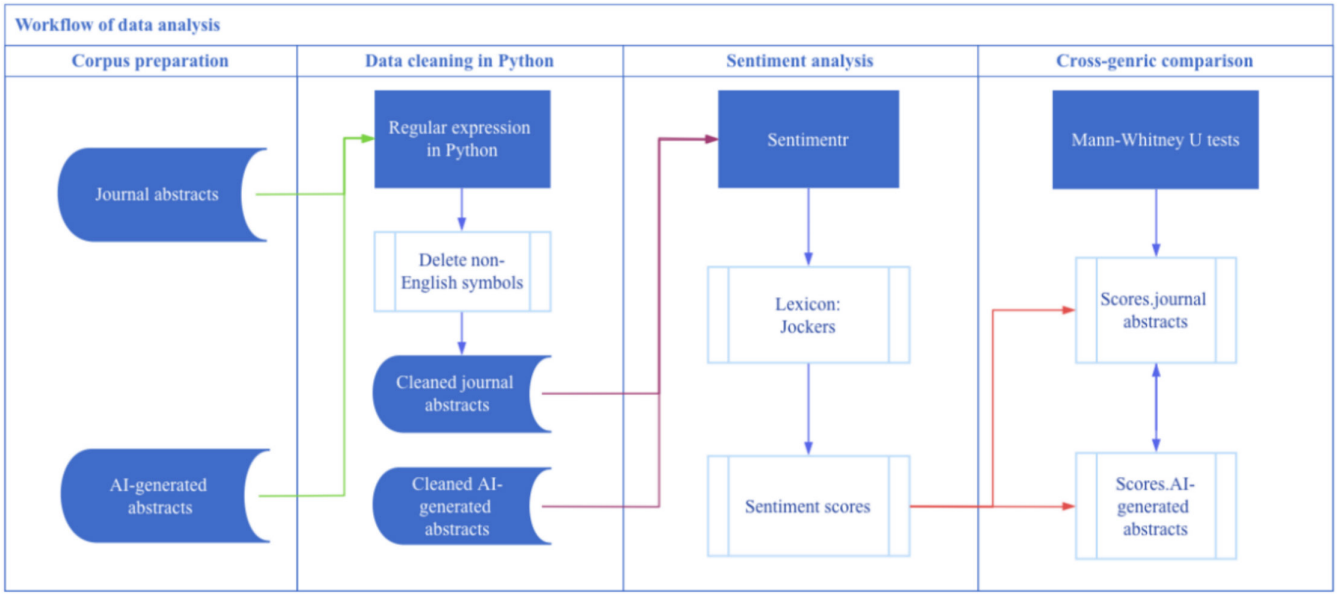



**FIGURE 1** | Workflow of data analysis.



17414857, 2027, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/leap.2126 by Université Paul-Valéry-Montpellier, Wiley Online Library on [22/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

used as the measure of central tendency rather than the mean. Since the distributional assumptions of parametric tests are not fulfilled (i.e., normality and homogeneity of variance), the Mann–Whitney *U* test, which makes no such assumptions, was appropriate for comparing the two independent samples.

## 3 | Results

In this section, we first report the differences in sentiment scores between journal abstracts and AI-generated abstracts before further presenting the distribution of sentiment scores across hard and soft disciplines.

### 3.1 | Overall Sentiment for Journal Abstracts and AI-Generated Abstracts

First, we present the overall sentiment distribution of journal abstracts and DeepSeek-generated abstracts. Generally, statistics show that journal abstracts and DeepSeek-generated abstracts do not differ in terms of sentiment scores ($U$_Statistic = 48,763, $p$_value = 0.06708 > 0.05). Figure 2 displays and compares the distribution of overall sentiment scores.

Although DeepSeek-generated abstracts tend to exhibit a higher median sentiment score (sentiment median = 1.10) than that of journal abstracts (sentiment median = 1.04), the two datasets do not differ significantly. In terms of the range of sentiment scores, DeepSeek-generated abstracts range from −1.200259597 to 3.947094089 whereas that of journal abstracts exhibits a broader range from −2.479963704 to 5.567073907, indicating that DeepSeek-generated abstracts exhibit more controlled and consistent sentiment expression, while human-authored journal abstracts demonstrate greater emotional variability and range. Finally, while DeepSeek and journal writing show no significant difference in sentiment distribution at the overall level, this trend may conceal important disciplinary variations. Therefore, a closer investigation of sentiment distribution across hard and soft disciplines is reported in the following sections.

### 3.2 | Sentiment for Journal Abstracts and AI-Generated Abstracts in Hard Disciplines

Unlike the general trends, sentiment scores in journal abstracts and AI-generated abstracts do not exhibit statistically significant differences ($U$_Statistic = 11,096, $p$_value = 0.9837 > 0.05).

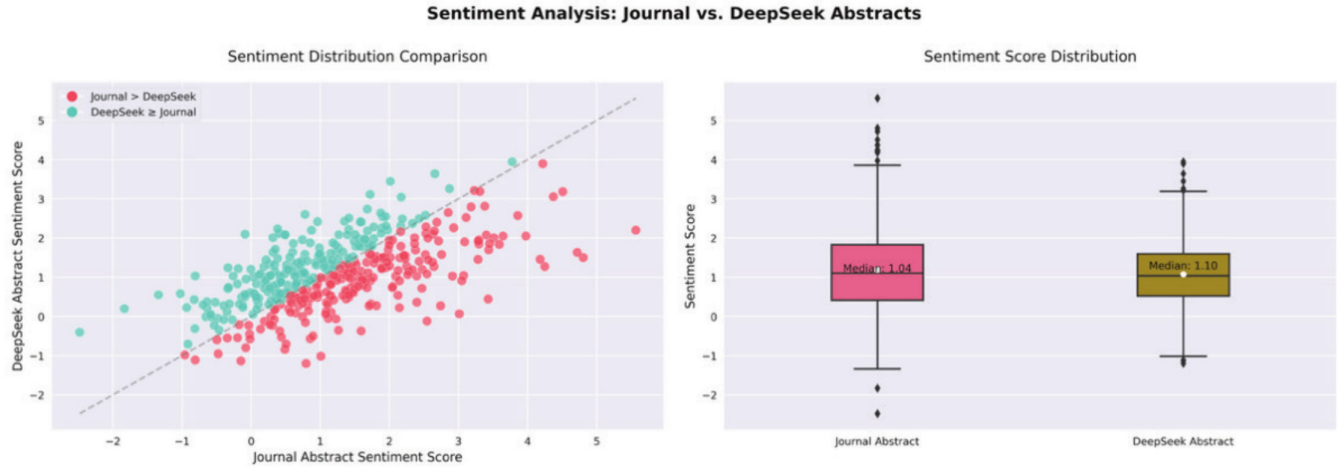



**FIGURE 2** | Sentiment distribution comparison for overall disciplines.

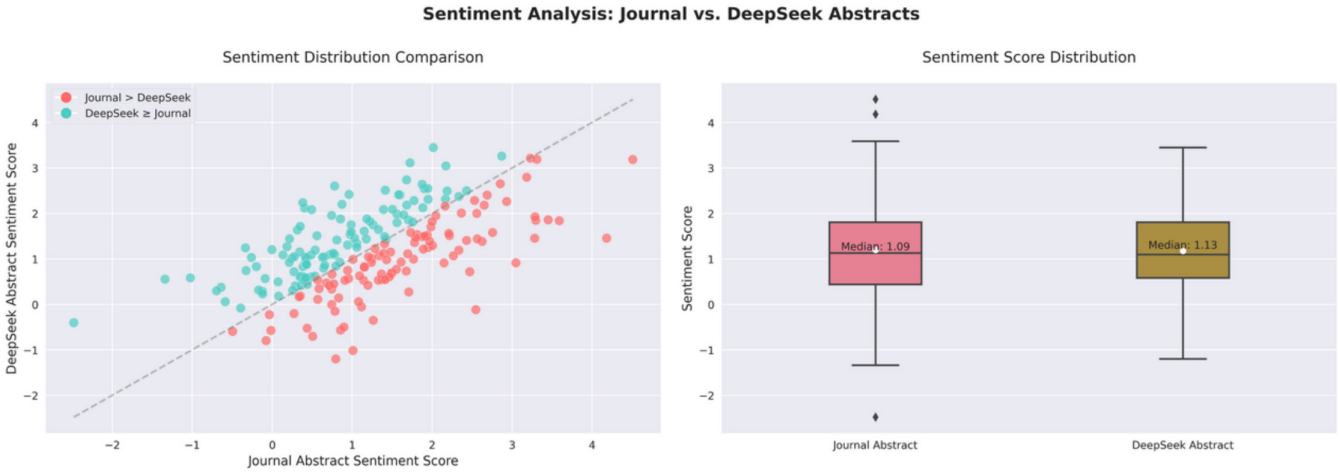



**FIGURE 3** | Sentiment distribution comparison for hard disciplines.



17414857, 2027, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/leap.2126 by Université Paul-Valéry-Montpellier, Wiley Online Library on [22/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Figure 3 displays the distribution of sentiment scores across the two groups, with distribution comparison by higher value on the left and sentiment range on the right.

Specifically, the results in hard disciplines show that DeepSeek-generated abstracts and journal abstracts exhibit similar sentiment medians. The former exhibits a median sentiment score of 1.09 and the latter a similar 1.13. In terms of the range of sentiment scores, DeepSeek-generated abstracts range from −1.200259597 to 3.448486561 and journal abstracts from −2.479963704 to 4.508756784. In addition, the boxplot in Figure 3 shows that the sentiment scores of both datasets tend to cluster within the range below 2 and above 0.

Next, a closer examination across seven hard disciplines is performed. The results, as displayed in Figure 4 and Table 2, show that among seven examined disciplines, no disciplines show significant differences in terms of their sentiment distribution between DeepSeek-generated and journal abstracts ($p > 0.05$ across seven disciplines). Given the results above, it may be concluded that DeepSeek-generated abstracts align with journal abstracts in terms of their sentiment expression across all hard disciplines, with no significant difference observed within specific disciplines. Taken together, these results indicate that, at the level of sentiment scores, DeepSeek-generated abstracts closely reproduce the linguistic positivity patterns of scientific discourse across each discipline.

### 3.3 | Sentiment for Journal Abstracts and AI-Generated Abstracts in Soft Disciplines

Results from soft disciplines show that sentiment scores between DeepSeek-generated abstracts and journal abstracts exhibit a significant difference ($U$_Statistic = 13,312, $p$_value = 0.0111 < 0.05). Figure 5 presents the distribution of sentiment scores across soft disciplines.

In detail, the results in soft disciplines show that journal abstracts and DeepSeek-generated abstracts tend to slightly differ in terms of their sentiment medians (Journal sentiment median = 0.99; DeepSeek sentiment median = 1.10). The two datasets also exhibit a different range of sentiment distribution. Journal sentiment distribution ranges from −1.831957561 to 5.567073907, whereas that of DeepSeek-generated abstracts ranges between −1.133947293 and 3.947094089.

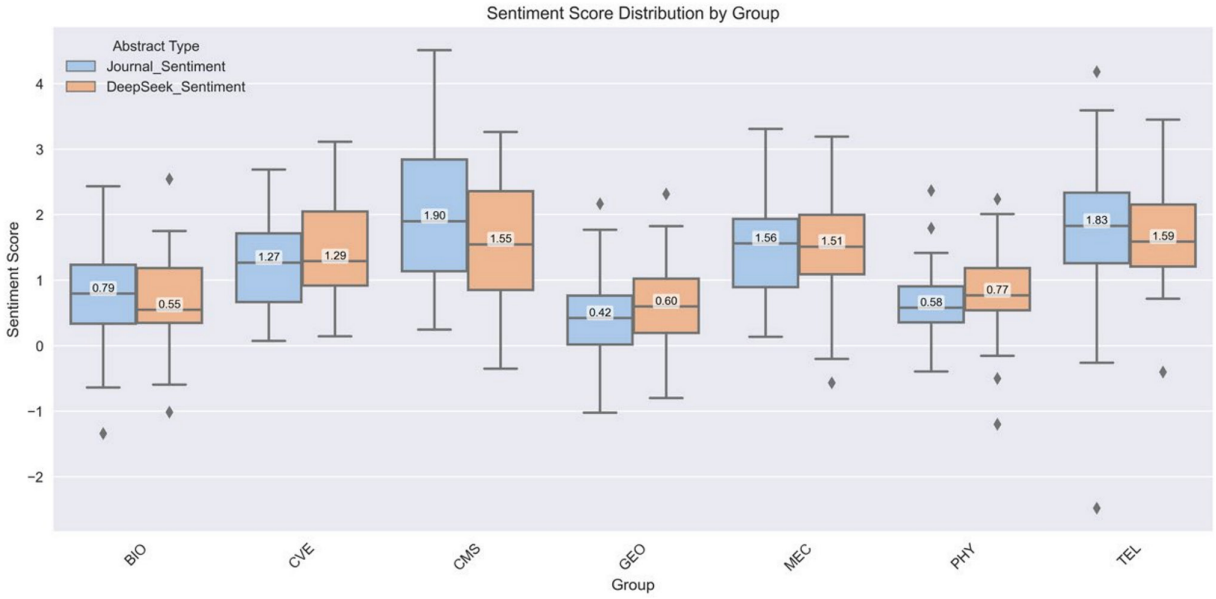



**FIGURE 4** | Sentiment score distribution across hard disciplines.

**TABLE 2** | Difference in sentiment distribution across seven hard disciplines.

| Discipline | Journal mean | DeepSeek mean | Mean difference | $U$_Statistic | $p$ |
|---|---|---|---|---|---|
| BIO | 0.755 | 0.649 | −0.106 | 498.0 | 0.4825 |
| CMS | 1.958 | 1.610 | −0.348 | 529.0 | 0.2458 |
| CVE | 1.219 | 1.438 | 0.219 | 383.0 | 0.3255 |
| GEO | 0.455 | 0.614 | 0.159 | 386.0 | 0.3478 |
| MEC | 1.528 | 1.481 | −0.047 | 443.0 | 0.9234 |
| PHY | 0.677 | 0.775 | 0.098 | 372.0 | 0.2519 |
| TEL | 1.773 | 1.657 | −0.115 | 511.0 | 0.3711 |



17414857, 2027, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/leap.2126 by Université Paul-Valéry-Montpellier, Wiley Online Library on [22/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

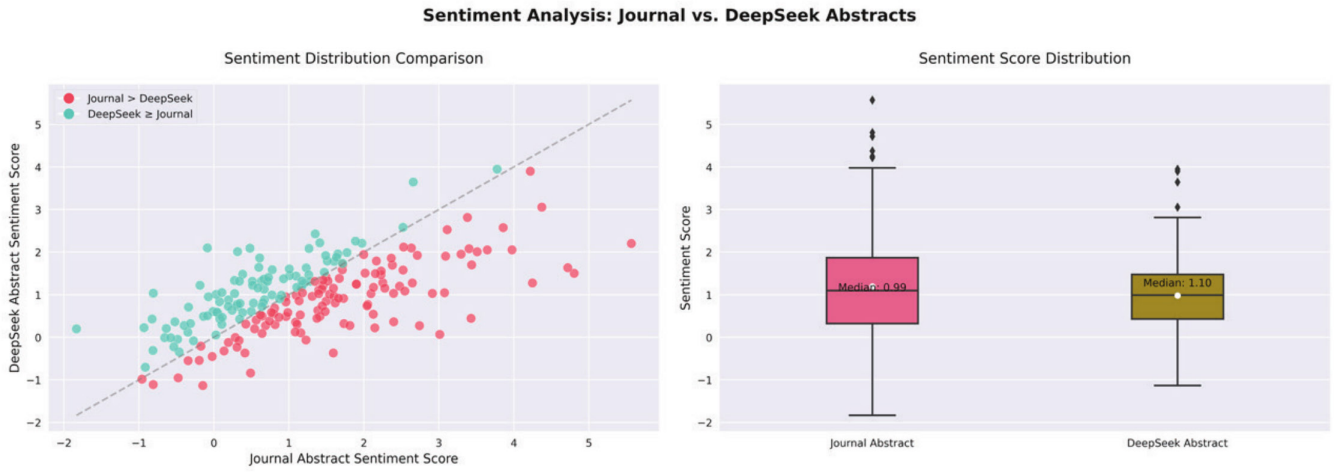



**FIGURE 5** | Sentiment score distribution across soft disciplines.

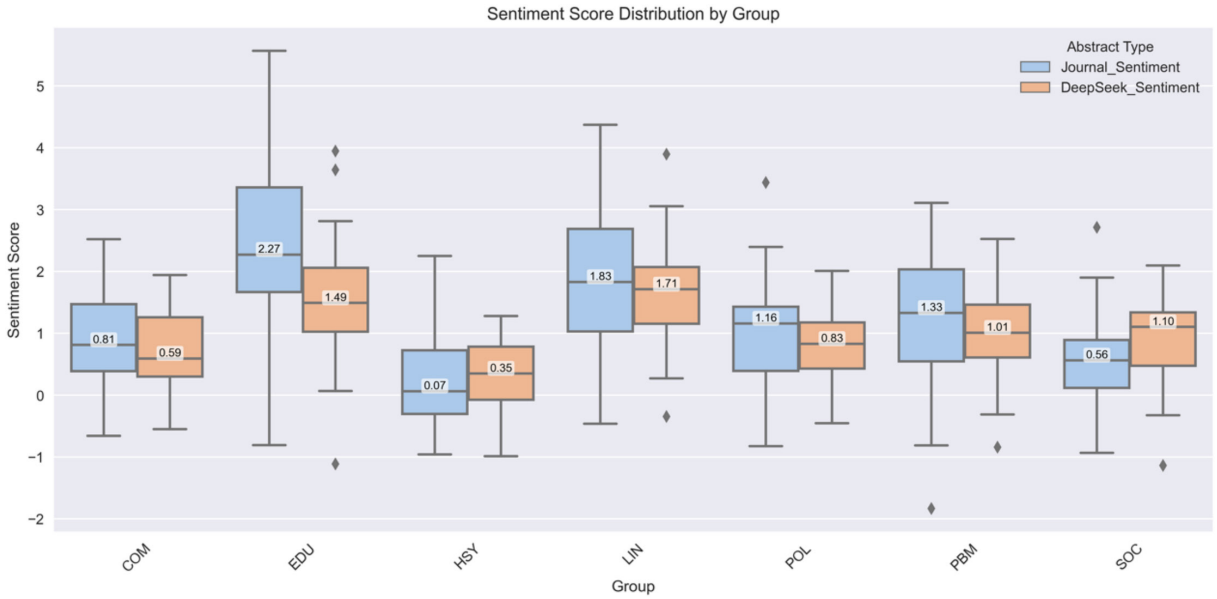



**FIGURE 6** | Sentiment score distribution across hard disciplines.

**TABLE 3** | Difference in sentiment distribution across seven soft disciplines.

| Discipline | Journal mean | DeepSeek mean | Mean difference | *U*_Statistic | *p* |
|---|---|---|---|---|---|
| COM | 0.904 | 0.68 | −0.224 | 533 | 0.2226 |
| EDU | 2.394 | 1.53 | −0.864 | 638 | 0.0056*** |
| HSY | 0.332 | 0.287 | −0.045 | 427 | 0.7394 |
| LIN | 1.961 | 1.637 | −0.325 | 515 | 0.3403 |
| PBM | 1.203 | 0.977 | −0.226 | 526 | 0.2643 |
| POL | 0.974 | 0.797 | −0.176 | 528 | 0.2519 |
| SOC | 0.595 | 0.948 | 0.354 | 291 | 0.0191* |

*Note:* $*p < 0.05$, $***p < 0.01$.

However, a closer investigation of the sentiment distribution across seven soft disciplines, as shown in Figure 6 and Table 3, suggests that the difference between journal and DeepSeek-generated abstracts in sentiment distribution across soft disciplines can be mainly attributed to two specific disciplines, namely Education (EDU) and Sociology (SOC).



17414857, 2027, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/leap.2126 by Université Paul-Valéry-Montpellier, Wiley Online Library on [22/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Sentiment distribution between the two datasets in EDU exhibits a statistically significant difference ($p_value = 0.0056 < 0.01$), with a −0.864 mean difference and $U$ statistic at 638. Similarly, sentiment scores of SOC show that the two datasets differ significantly ($p_value = 0.0191 < 0.05$), in terms of a 0.354 mean difference and $U$ statistic at 291.

Given the results above, there exists a significant difference between journal and DeepSeek-generated writing in sentiment distribution. However, internal data analysis reveals that this difference is not universal but is limited to two specific sub-disciplines, suggesting that AI-human differences are confined to specific soft disciplines rather than the entire genre. This may help us conclude that journal abstracts and DeepSeek-generated abstracts do not differ in sentiment expression. However, it may fail to align with particular forms of disciplinary discourse.

## 4 | Discussion

The current study examined linguistic positivity in abstracts written by human authors and those generated by AI. To do so, we compared the sentiment scores of 420 abstracts published in top journals across disciplines, both hard and soft ones, with 420 abstracts generated by DeepSeek. The results show that DeepSeek-generated abstracts show no significant difference from human-written journal abstracts in terms of their expression of sentiment in academic writing across hard and soft disciplines. Specifically, across all seven hard disciplines, journal abstracts and DeepSeek-generated abstracts show no significant difference in their sentiment distribution, indicating that in terms of lexicon-based sentiment scores, DeepSeek-generated abstracts align closely with those written by human authors across the hard sciences. In soft disciplines, although the overall sentiment distribution between DeepSeek-generated and journal abstracts differs significantly, the difference is largely attributed to two specific disciplines, that is, in EDU and SOC. In the rest five soft disciplines, no significant difference in sentiment distribution is observed, suggesting that other than certain disciplines, AI may be capable of producing academic texts with a closer-to-human sentiment level.

It is particularly interesting to observe the two fields diverging from the rest. Prior research has shown that the evaluative language and rhetorical structure of abstracts vary markedly across disciplines. For instance, Stotesbury (2003) found that abstracts in social sciences and humanities contain substantially more explicit evaluative lexis than those in natural sciences. Liu and Li (2024) showed that abstract writers systematically adjust their linguistic choices to communicative purposes and audience. These disciplinary conventions may account for the two exceptions observed here. Education abstracts written by human authors score comparatively high in terms of sentiment levels, as they may draw on a narrative rich evaluatively rich tradition, whereas the DeekSeek's more uniform style fails to reproduce such feature and thus score significantly lower. Sociology, by contrast, is a field whose research typically addresses social problems, and human authors tend to encode their evaluations in a cautious manner. Therefore, their abstracts may remain relatively restrained in sentiment. DeekSeek may have default to an optimistic and template-like style, yielding texts that are more positive than their human counterparts.

Findings of this study diverge from existing discussion on the comparison between human and AI-generated academic writing. While much of the existing literature remains focused on the differences between AI-generated and human academic writing (i.e., Jiang and Hyland 2025a, 2025b, 2025c; Kong and Liu 2024; Markowitz et al. 2024; Zhang and Crosthwaite 2025), this study reveals that in terms of sentiment expression, such differences have largely disappeared. This suggests that, at the level of sentiment scores, AI-generated abstracts may have converged with human-written abstracts in their linguistic positivity, approximating the positive evaluative norms of scholarly discourse. This convergence raises profound questions about the future of academic creativity, at a time when AI is capable of replicating not only the structural and stylistic patterns of academic writing, but also the sentiment dimension of its evaluative language. Ultimately, the production of knowledge, the modes of human–AI collaboration and the entire framework of scholarly evaluation may be altered.

First, it is necessary for us to envision a future where AI may have the potential to produce human-level academic writing. In traditional academic writing, researchers bear the dual responsibility of both knowledge production and its textual expression. Our findings suggest that DeepSeek is capable of producing text with the same level of sentiment scores as human authors, demonstrating that AI models may have been developed towards research automation (Sætra 2025), where AI is capable of 'most if not all steps in scientific research' (Sætra 2025, 1). At the very least, our evidence indicates that AI may now effectively manage sentiment expression in academic writing, potentially freeing human researchers to focus more intensively on knowledge innovation itself. However, this does not necessarily direct us to outright replacement of human writing since 'outsourcing of cognitive effort to machines risks the atrophy of essential research skills' (Sætra 2025, 9), particularly if researchers become passive supervisors rather than active intellectual partners.

Second, we must acknowledge the possibility of human–AI collaboration in scholarly practice. Academic practice in the future may operate through what Hayles (2017) terms as 'cognitive assemblage', where AI handles technical writing tasks while researchers focus on core innovation. This is also discussed in Watermeyer et al.'s (2024) work, where the 'automating of academia' may serve as a 'positive disruption' (p. 446) to academia and relieve academics from the hard labour of academic production and focus more on the 'scholarly craftsmanship' (p. 446). The key challenge then would be to distinguish between 'automating drudgery' and 'automating thought' (Sætra 2025, 10).

Finally, and perhaps most urgently, our findings highlight the challenge of the current framework of academic evaluation. AI's alignment with human writing in sentiment expression in academic writing makes another traditional quality metric less useful, meaning that in terms of sentiment scores, one will not be able to distinguish AI-generated or human-written academic writing. To the academic and publishing community, it may be urgent to develop new assessment benchmarks that shift focus from writing quality to research quality, possibly with a shift of



17414857, 2027, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/leap.2126 by Université Paul-Valéry-Montpellier, Wiley Online Library on [22/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

focus towards (1) the formulation and framing of research, what Battleday and Gershman (2026) described as the 'hard problems' that still challenge current AI systems; (2) process transparency over output, which marks a shift towards strengthened documentation and disclosure of intellectual workflows in ways that distinguish human from AI contributions, as advocated by Dwivedi et al. (2023).

### Author Contributions

**Ying Wei:** methodology, writing – original draft, data curation. **Mingxin Yao:** conceptualization, methodology, supervision, writing – review and editing.


### Funding

This research was supported by Research Start-up Fund of Sichuan International Studies University (Grant No. 5001/222125035).


### Conflicts of Interest

The authors declare no conflicts of interest.

### Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

### References


Augustine, A. A., M. R. Mehl, and R. J. Larsen. 2011. "A Positivity Bias in Written and Spoken English and Its Moderation by Personality and Gender." *Social Psychological and Personality Science* 2, no. 5: 508–515. https://doi.org/10.1177/1948550611399154.

Battleday, R., and S. Gershman. 2026. "Artificial Intelligence for Science: The Easy and Hard Problems." *Philosophical Transactions of the Royal Society a Mathematical Physical and Engineering Sciences* 384, no. 2320. https://doi.org/10.1098/rsta.2024.0530.

Becher, T. 1994. "The Significance of Disciplinary Differences." *Studies in Higher Education* 19, no. 2: 151–161. https://doi.org/10.1080/03075079412331382007.

Boucher, J., and C. E. Osgood. 1969. "The Pollyanna Hypothesis." *Journal of Verbal Learning and Verbal Behavior* 8, no. 1: 1–8. https://doi.org/10.1016/S0022-5371(69)80002-2.

Cao, X., L. Lei, and J. Wen. 2021. "Promoting Science With Linguistic Devices: A Large-Scale Study of Positive and Negative Words in Academic Writing." *Learned Publishing* 34, no. 2: 82–88. https://doi.org/10.1002/leap.1322.

Chen, S. 2024. "Mining Emotions in Academic Writing: A Subdisciplinary Probe Into Medical Research Articles." *English for Specific Purposes* 74: 55–67. https://doi.org/10.1016/j.esp.2024.01.002.

Dodds, P. S., E. M. Clark, S. Desu, et al. 2015. "Human Language Reveals a Universal Positivity Bias." *Proceedings of the National Academy of Sciences of the United States of America* 112, no. 8: 2389–2394. https://doi.org/10.1073/pnas.1411678112.

Dwivedi, Y. K., N. Kshetri, L. Hughes, et al. 2023. "Opinion Paper:"So What if ChatGPT Wrote It?" Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy." *International Journal of Information Management* 71: 102642.

Garcia, D., A. Garas, and F. Schweitzer. 2012. "Positive Words Carry Less Information Than Negative Words." *EPJ Data Science* 1, no. 1: 1–12. https://doi.org/10.1140/epjds3.

Goulart, L., M. L. Matte, A. Mendoza, L. Alvarado, and I. Veloso. 2024. "AI or Student Writing? Analyzing the Situational and Linguistic Characteristics of Undergraduate Student Writing and AI-Generated Assignments." *Journal of Second Language Writing* 66: 101160. https://doi.org/10.1016/j.jslw.2024.101160.

Hayles, N. K. 2017. *Unthought: The Power of the Cognitive Nonconscious*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226447919.001.0001.

Holtz, P., E. Deutschmann, and H. Dobewall. 2017. "Cross-Cultural Psychology and the Rise of Academic Capitalism: Linguistic Changes in CCR and JCCP Articles, 1970–2014." *Journal of Cross-Cultural Psychology* 48, no. 9: 1410–1431. https://doi.org/10.1177/0022022117724902.

Jiang, F., and K. Hyland. 2024. "Does ChatGPT Argue Like Students? Bundles in Argumentative Essays." *Applied Linguistics* 46, no. 3: 375–391. https://doi.org/10.1093/applin/amae052.

Jiang, F., and K. Hyland. 2025a. "Rhetorical Distinctions: Comparing Metadiscourse in Essays by ChatGPT and Students." *English for Specific Purposes* 79: 17–29. https://doi.org/10.1016/j.esp.2025.03.001.

Jiang, F. K., and K. Hyland. 2025b. "Metadiscursive Nouns in Academic Argument: ChatGPT vs Student Practices." *Journal of English for Academic Purposes* 75: 101514. https://doi.org/10.1016/j.jeap.2025.101514.

Jiang, F., and K. Hyland. 2025c. "Does ChatGPT Write Like a Student? Engagement Markers in Argumentative Essays." *Written Communication* 42, no. 3: 463–492. https://doi.org/10.1177/07410883251328311.

Jockers, M. L. 2017. "Syuzhet: Extracts Sentiment and Sentiment-Derived Plot Arcs From Text." https://cran.r-project.org/web/packages/syuzhet.

Kong, X., and C. Liu. 2024. "A Comparative Genre Analysis of AI-Generated and Scholar-Written Abstracts for English Review Articles in International Journals." *Journal of English for Academic Purposes* 71: 101432. https://doi.org/10.1016/j.jeap.2024.101432.

Lerchenmueller, M. J., O. Sorenson, and A. B. Jena. 2019. "Gender Differences in How Scientists Present the Importance of Their Research: Observational Study." *BMJ* 367: l6573. https://doi.org/10.1136/bmj.l6573.

Lin, S., and P. Crosthwaite. 2024. "The Grass Is Not Always Greener: Teacher vs. GPT-Assisted Written Corrective Feedback." *System* 127: 103529. https://doi.org/10.1016/j.system.2024.103529.

Liu, X., and H. Zhu. 2023. "Linguistic Positivity in Soft and Hard Disciplines: Temporal Dynamics, Disciplinary Variation, and the Relationship With Research Impact." *Scientometrics* 128, no. 5: 3107–3127. https://doi.org/10.1007/s11192-023-04679-5.

Liu, X., and H. Zhu. 2025. "The Diachronic Change in Linguistic Positivity in the Academic Book Reviewing of Language Studies: A Text-Mining Analysis." *Scientometrics* 130, no. 1: 133–157. https://doi.org/10.1007/s11192-024-05208-8.

Liu, Y., and T. Li. 2024. "Comparing the Syntactic Complexity of Plain Language Summaries and Abstracts: A Case Study of Marine Science Academic Writing." *Journal of English for Academic Purposes* 68: 101350. https://doi.org/10.1016/j.jeap.2024.101350.

Markowitz, D. M., J. T. Hancock, and J. N. Bailenson. 2024. "Linguistic Markers of Inherently False AI Communication and Intentionally False Human Communication: Evidence From Hotel Reviews." *Journal of Language and Social Psychology* 43, no. 1: 63–82. https://doi.org/10.1177/0261927X231200201.





17414857, 2027, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/leap.2126 by Université Paul-Valéry-Montpellier, Wiley Online Library on [22/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Martin, J. R., and P. R. R. White. 2005. *The Language of Evaluation: Appraisal in English*. Palgrave Macmillan.

Matlin, M. W., and D. J. Stang. 1978. *The Pollyanna Principle: Selectivity in Language, Memory, and Thought*. Schenkman.

Mo, Z., and P. Crosthwaite. 2025. "Exploring the Affordances of Generative AI Large Language Models for Stance and Engagement in Academic Writing." *Journal of English for Academic Purposes* 75: 101499. https://doi.org/10.1016/j.jeap.2025.101499.

Nañola, E. L., R. L. Arroyo, N. J. T. Hermosura, M. Ragil, J. N. U. Sabanal, and H. B. Mendoza. 2025. "Recognizing the Artificial: A Comparative Voice Analysis of AI-Generated and L2 Undergraduate Student-Authored Academic Essays." *System* 130: 103611. https://doi.org/10.1016/j.system.2025.103611.

Rinker, T. 2019. "Calculate Text Polarity Sentiment." https://cran.r-project.org/web/packages/sentimentr/sentimentr.pdf.

Rozin, P., L. Berman, and E. Royzman. 2010. "Biases in Use of Positive and Negative Words Across Twenty Natural Languages." *Cognition and Emotion* 24, no. 3: 536–548. https://doi.org/10.1080/02699930902793462.

Sætra, H. S. 2025. "The Rise of the Research Automaton: Science as Process or Product in the Era of Generative AI?" *AI & Society* 41, no. 3: 1865–1879. https://doi.org/10.1007/s00146-025-02557-7.

Shapin, S. 2022. "Hard Science, Soft Science: A Political History of a Disciplinary Array." *History of Science* 60, no. 3: 287–328. https://doi.org/10.1177/00732753221094739.

Stotesbury, H. 2003. "Evaluation in Research Article Abstracts in the Narrative and Hard Sciences." *Journal of English for Academic Purposes* 2, no. 4: 327–341. https://doi.org/10.1016/S1475-1585(03)00049-3.

Vinkers, C. H., J. K. Tijdink, and W. M. Otte. 2015. "Use of Positive and Negative Words in Scientific PubMed Abstracts Between 1974 and 2014: Retrospective Analysis." *BMJ* 351: h6467. https://doi.org/10.1136/bmj.h6467.

Watermeyer, R., L. Phipps, D. Lanclos, and C. Knight. 2024. "Generative AI and the Automating of Academia." *Postdigital Science and Education* 6, no. 2: 446–466. https://doi.org/10.1007/s42438-023-00440-6.

Weidmann, N. B., S. Otto, and L. Kawerau. 2018. "The Use of Positive Words in Political Science Language." *PS, Political Science & Politics* 51, no. 3: 625–628.

Wen, J. U., and L. Lei. 2022. "Linguistic Positivity Bias in Academic Writing: A Large-Scale Diachronic Study in Life Sciences Across 50 Years." *Applied Linguistics* 43, no. 2: 340–364. https://doi.org/10.1093/applin/amab037.

Won, D. O., Y. K. Shin, H. J. Kim, and I. W. Yoo. 2025. "Advancing Language Assessment With GPT: Is It Nonnative-Language Friendly?" *Language Assessment Quarterly* 22: 1–28. https://doi.org/10.1080/15434303.2024.2444349.

Yang, S., S. Chen, H. Zhu, J. Lin, and X. Wang. 2024. "A Comparative Study of Thematic Choices and Thematic Progression Patterns in Human-Written and AI-Generated Texts." *System* 126: 103494. https://doi.org/10.1016/j.system.2024.103494.

Yuan, Z. M., C. Jiang, and M. Yao. 2025. "The Art of Positive Reporting in Media Discourse: Time Magazine's Linguistic Positivity Bias Over the 20th Century." *Digital Scholarship in the Humanities* 40, no. 3: 1041–1050. https://doi.org/10.1093/llc/fqaf039.

Yuan, Z. M., and M. Yao. 2022. "Is Scientific Writing Becoming More Positive? A Large-Scale Diachronic Case Study of Science Research Articles Across 25 Years." *Scientometrics* 127, no. 11: 6191–6207.

Zhang, M., and P. Crosthwaite. 2025. "More Human Than Human? Differences in Lexis and Collocation Within Academic Essays Produced by ChatGPT-3.5 and Human L2 Writers." *International Review of Applied Linguistics in Language Teaching* 64: 733–760. https://doi.org/10.1515/iral-2024-0196.

Zhu, H., T. Wang, and N. Pang. 2024. "Diachronic Changes in Lexical Density of Research Article Abstracts: A Corpus-Based Study." *Lingua* 312: 103837. https://doi.org/10.1016/j.lingua.2024.103837.



17414857, 2027, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/leap.2126 by Université Paul-Valéry-Montpellier, Wiley Online Library on [22/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License