Scientometrics (2026) 131:7025–7059
https://doi.org/10.1007/s11192-026-05774-z

Does scientific writing converge to U.S. english? evidence from generative AI-assisted publications

Dragan Filimonovic[1] · Christian Rutzer[1] · Jeffrey T. Macher[2] · Rolf Weder[1]

Received: 9 March 2026 / Accepted: 3 August 2026 / Published online: 22 August 2026
© The Author(s) 2026

Abstract
A growing literature documents that generative artificial intelligence (GenAI) is changing scientific writing. Much of this literature, however, focuses on absolute changes in vocabulary or readability. An important question remains unanswered: Does GenAI use lead to systematic convergence of writing styles toward U.S. scientific English? Unlike absolute changes, convergence signals whether language-related publication barriers are declining and suggests broader implications for participation and competition in global science. This study directly addresses this question using 5.65 million English-language scientific articles published from 2021 to 2024 and indexed in Scopus. We measure linguistic similarity to a U.S. benchmark corpus using SciBERT text embeddings, and estimate dynamic changes using an event-study difference-in-differences (DiD) design with repeated cross-sections centered on the late-2022 launch of ChatGPT. We find that GenAI-assisted publications from non-English-speaking countries exhibit statistically significant and increasing convergence toward U.S. scientific English, relative to non-GenAI-assisted publications from these countries. This effect is strongest for domestic author teams from countries more linguistically distant from English. The results suggest that GenAI tools are reducing language-related barriers in scientific publications, but whether this formal convergence alters underlying research quality, originality, or the trajectory of scientific discovery remains an open question.

Keywords Generative AI · Linguistic convergence · Scientific writing · SciBERT · Bibliometrics · Difference-in-differences

JEL Classification: O33 · D83 · C23 · C55 · I23

✉ Dragan Filimonovic
dragan.filimonovic@unibas.ch

[1] Center for Innovation Economics Basel, Faculty of Business and Economics, University of Basel, Peter Merian-Weg 6, Basel 4002, Switzerland

[2] Robert E. McDonough School of Business, Georgetown University, 37th and O Streets NW, Washington DC 20057, USA

Introduction

The dominance of English in the global scientific community has long conferred structural advantages on native English-speaking researchers and created persistent obstacles for non-native English-speaking researchers (Amano et al., 2016, 2023; Montgomery, 2013; Nature Human Behaviour, 2023). For instance, non-Anglophone researchers often face greater writing burdens and peer-review bias risks associated with language proficiency, with documented consequences for publication success and career progression (Amano et al., 2023; Castañeda, 2020; Clavero, 2010; Hanauer et al., 2019; Politzer-Ahles et al., 2020). These disadvantages have historically required substantial investments, including time spent in English-speaking research environments and payments made to professional language editing services (Balan, 2021; Stephan et al., 2016).

Yet the launch of ChatGPT in late 2022 introduced a widely accessible and low-cost alternative for text editing and language assistance, offering a potential means of reducing these longstanding linguistic barriers (Bubeck et al., 2023; Van Noorden & Perkel, 2023). Researchers have rapidly adopted these tools, and a growing body of bibliometric and empirical research has begun documenting shifts in scientific writing, including changes in vocabulary use and textual complexity (De Haan et al., 2025; Feyzollahi & Rafizadeh, 2025; Geng & Trotta, 2024; Gray, 2024; Kobak et al., 2025; Kousha & Thelwall, 2026; Liang et al., 2026; Thelwall & Kousha, 2026; Uribe & Maldupa, 2024). Scholars facing greater language frictions and those who are less established are among the most active adopters of GenAI writing tools (Lin et al., 2025; Liu et al., 2025), with the resulting productivity gains concentrated at institutions outside the Anglophone core (Kusumegi et al., 2025). Documented improvements in manuscript readability follow a similar pattern, yielding the most pronounced gains for authors from countries more linguistically distant from English (Prakash et al., 2025).

Despite these valuable insights, existing research has predominantly focused on absolute shifts in writing features, such as vocabulary use or readability scores, rather than directional convergence toward a well-defined linguistic standard. Furthermore, current evidence relies heavily on preprint repositories (Kusumegi et al., 2025) or narrowly-defined scientific fields (Lin et al., 2025; Liu et al., 2025; Prakash et al., 2025), leaving the broader peer-reviewed record unexamined at scale. The distinction between absolute change and convergence is not merely methodological but conceptually vital. Absolute adjustments indicate that writing is changing, whereas convergence reveals where it is moving and, by implication, which stylistic norms are being adopted. To illustrate this nuance, two papers may undergo equivalent absolute changes in their writing (e.g., both more concise or using a more diverse vocabulary), yet only one might move systematically toward a given reference register. Disentangling these distinct dynamics requires a comparative design that explicitly evaluates shifting writing styles against an established baseline standard.

To address this gap, we use U.S. scientific English as a global benchmark to provide large-scale evidence of linguistic convergence across multiple scientific fields. We track the stylistic evolution of 5.65 million English publications by authors from countries outside the Anglophone core, published from 2021 to 2024 and indexed in Scopus. For each publication, we measure linguistic similarity to a benchmark corpus of U.S. scientific writing using SciBERT text embeddings[1] (Beltagy et al., 2019). We then estimate dynamic

[1] SciBERT is a transformer architecture based on BERT (Devlin et al., 2019) and pretrained on large scientific corpora that effectively captures the semantic and stylistic properties of academic writing.

changes using an event study difference-in-differences (DiD) design centered on the late-2022 launch of ChatGPT. The results demonstrate that the diffusion of GenAI tools drives a pronounced convergence for post-2022 publications assisted by GenAI toward U.S. scientific writing, relative to publications without such assistance. This induced convergence is strongest where language barriers are greatest: namely, for domestic author teams from countries more linguistically distant from English.

This paper makes three contributions. First, we provide cross-field and transnational evidence that GenAI use induces measurable convergence in the linguistic presentation of published science, moving beyond the absolute changes documented in prior research. Against the broader backdrop of research on linguistic barriers and biases in global science (Amano et al., 2023; Castañeda, 2020; Clavero, 2010; Hanauer et al., 2019; Politzer-Ahles et al., 2020), our results provide evidence that GenAI-assisted publications from linguistically distant countries become more similar to dominant English-language scientific writing norms after 2022. This pattern is consistent with a partial reduction in language-related barriers, although our analysis concerns linguistic form rather than scientific substance. This potential reduction in linguistic frictions should be weighed against the risk that convergence toward a single dominant register may entail homogenization at the cost of linguistic diversity emphasized by some (Flowerdew, 2022) and questioned by others (House, 2003).

Second, we add to research that examines how digital technology adoption influences academic outcomes. For instance, internet access has been shown to increase research output by lowering discovery costs and facilitating collaboration (Barjak, 2006; da Fonseca Pachi et al., 2012), and higher internet penetration at the country level is positively correlated with national publication volumes (Xu & Reed, 2021). More directly relevant to our work, prior literature examines how earlier AI tools enhanced writing quality and efficiency for authors who are not native English speakers. Specifically, Ghufron and Rosyida (2018) and Fitria (2021) find that using Grammarly substantially improves English writing performance, while Mundt and Groves (2016) show that machine translation tools such as Google Translate enhance academic exchange and diversify research communities. We build on these research efforts by examining whether and how the latest generation of digital tools drives changes in linguistic patterns in the published scientific record at a much larger scale.

Third, we complement the expanding literature on the effects of GenAI adoption on research outcomes. Beyond the readability enhancements highlighted in the aforementioned studies, parallel research demonstrates that GenAI tools are reshaping knowledge production along other dimensions. Kusumegi et al. (2025) analyze three major preprint repositories and find that LLM usage substantially accelerates manuscript output and alters traditional quality signals, including a reversal of the historical correlation between language complexity and peer-review outcomes for LLM-assisted papers. Filimonovic et al. (2025) find that GenAI adoption in the social and behavioral sciences significantly boosts overall research productivity and modestly improves publication quality, particularly for scientists from non-English speaking countries. Tang et al. (2025) caution, however, that these benefits are not evenly distributed. In particular, they find using SSRN preprint data that male researchers' productivity gains from GenAI currently exceed those of female researchers, thereby widening existing gender gaps in science. By documenting linguistic convergence toward U.S. scientific English, our study adds a geographic and linguistic dimension to this literature, complementing existing evidence on productivity, quality, and inequality in research outcomes.

The rest of the paper is structured as follows. Section 2 presents the data and methods. Section 3 analyzes whether GenAI-assisted publications exhibit stronger convergence toward U.S. scientific writing. Section 4 assesses empirical robustness under alternative keyword thresholds, stricter GenAI identification rules, and different U.S. benchmark corpus definitions. Section 5 discusses the findings and their implications, and Section 6 concludes.

Data and methods

This section discusses the data and methods. Subsection 2.1 describes the dataset; Subsections 2.2 and 2.3 respectively detail how GenAI-assisted papers are identified and linguistic similarity is measured; and Subsection 2.4 explains the empirical strategy.

Data

Our starting dataset is all scientific publications written in English and indexed in the Scopus database from 2021 to 2024 (Elsevier, 2023; Singh et al., 2021). For each publication, we obtain information on the publication year, Scopus pre-defined scientific fields, journal impact factor, and author country (based upon institution affiliation). Scopus assigns scientific fields at the journal level using the All Science Journal Classification (ASJC), meaning that a journal can be classified into multiple fields and each of its publications inherits all field assignments of the journal. For authors with more than one affiliation, we use the first listed as it typically corresponds to the main institution. We drop observations for publications where all author affiliations are to English-speaking countries.[2] We focus on 5.65 million unique publications that include at least one author affiliation to a non-English-speaking country. Publications with author affiliations in multiple countries are informative about writing norms in each affiliated country, and publications whose journals are classified into multiple Scopus fields are relevant to each of those fields. To preserve this information, we represent each publication by all of its observed country and field assignments, constructing one observation for each unique publication, country and field triad.

Our unit of analysis is therefore the publication × country × field cell, with each cell weighted fractionally so that the total weight of any publication sums to one regardless of how many countries or fields it spans. This fractional cell structure corresponds to fractional counting in the bibliometric literature (Perianes-Rodriguez et al., 2016; Waltman, 2016) and yields an expanded sample of 14,315,948 observations. This approach preserves both country-affiliation and field-specific benchmark information without giving publications with multiple country or field assignments disproportionate influence in the estimation. Appendix Table 8 provides several empirical robustness tests that consider whether the counting method affects the results. First, we use full counting, where each country-field observation is weighted equally. Second, we assign each publication to the first author's country and to the first-listed scientific field in the Scopus classification. Third, we restrict the sample to publications assigned to a single country and a single Scopus field. The results are qualitatively identical to the baseline estimation.

[2] We retain observations for publications where at least one author affiliation is to a non-English-speaking country, as it allows for mixed-language author team analysis.

Table 1 Scopus scientific fields to aggregated group

Aggregated group	Scopus scientific fields
Life sciences	Biochemistry, genetics and molecular biology, medicine, public health, immunology, neuroscience, pharmacology, nursing, dentistry, veterinary science, agriculture
Physical sciences	Physics, chemistry, materials science, mathematics, environmental science, earth sciences
Engineering and technology	Engineering, Chemical Engineering, Computer Science, Energy Studies
Social sciences	Economics, business, psychology, sociology, decision sciences

This table maps Scopus scientific field classifications into aggregated groups

Table 2 Lexical patterns used to identify GenAI-assisted writing

Keywords and stems
delv*, groundbreak*, intric*, meticul*, realm*, revolution*, showcas*, underscore*, unveil*, while, elevat*, valuabl*, crucial*, empower, unleash, unlock, lever*, embarked, relentless, endeavour, enlightening, insight*, esteemed, resonate*, enhanc*, expertise*, offering*, tapestry, foster*, systemic*, inherent, synerg*, explor*, pivotal*, adhere, amplif*, embark*, invaluabl*, enlighten*, conceptual*, emphasiz*, complexit*, recogniz*, adapt*, promot*, critique, comprehensive, implication*, complementar*, perspective*, holistic, discern, multifacet*, nuanc*, underpinning*, cultivat*, integral, profound*, facilitat*, encompass*, elucidat*, unravel*, paramount, characteriz*, significant*

This table reports the full candidate keyword set. The subset actually used to identify GenAI-assisted publications is determined by a field-specific frequency filter (see Section 2.2 and Appendix Fig. 5). Asterisks (*) indicate stemmed keywords whereby all variants sharing the same root are grouped together: e.g., 'delv*' captures 'delve', 'delving', 'delved'

To improve interpretability in cross-scientific field comparisons, we aggregate the detailed Scopus journal-based scientific fields into four broad groups: Life Sciences, Physical Sciences, Engineering and Technology, and Social Sciences. Multidisciplinary and Arts and Humanities journals are excluded due to their small sample size and ambiguous classification. This dimensionality reduction simplifies cross-scientific field analysis while retaining the main structural differences among these scientific domains. Table 1 provides the mapping from Scopus scientific fields to these aggregated groups.

Identifying GenAI-assisted papers

We focus on publication titles and abstracts because they are especially well-suited for detecting stylistic shifts (Plavén-Sigray et al., 2017) and carry significant editorial weight (Ketcham et al., 2010), making them meaningful proxies for GenAI uptake in scientific writing. We identify GenAI use by detecting lexical pattern shifts in these texts pre- and post-launch of ChatGPT. Our approach builds on recent evidence that text produced with the assistance of large language models (LLMs) exhibits clear lexical and stylistic markers (Alafnan & Mohdzuki, 2023). We exploit a 65-term vocabulary previously linked to GenAI-related content (Feyzollahi & Rafizadeh, 2025; Filimonovic et al., 2025; Uribe & Maldupa, 2024). Table 2 lists the keywords that are applied in stemmed form to detect GenAI use in scientific publications.

Recent full-text studies provide support for this type of keyword-based identification strategy. Kousha and Thelwall (2026) track LLM-associated terms across major bibliometric databases and document sharp post-2022 increases in terms such as "delve", "underscore", and "intricate" in more than 2.4 million full-text articles from PubMed Central (PMC). Thelwall and Kousha (2026) extend this evidence to 1.25 million Multidisciplinary Digital Publishing Institute (MDPI) articles, and show that the prevalence of LLM-associated terms varies substantially across fields and journals. These studies support both the use of lexical markers to identify probable GenAI-assisted writing and our approach in determining relevant markers by scientific field rather than imposing a uniform vocabulary across disciplines. At the same time, our approach has two important limitations. First, lexical markers provide only an indirect proxy for GenAI use rather than definitive evidence. Second, a title and abstract focus limits the scope of the measure, as it may miss GenAI use elsewhere in the publication. Conversely, if authors use GenAI disproportionately to draft or polish abstracts, the measure may overstate the extent of GenAI assistance in the publication as a whole. Our classification should therefore be interpreted as a proxy for detectable GenAI-assisted writing in titles and abstracts, rather than as a comprehensive measure of GenAI use throughout the entire publication.

Recognizing that keyword occurrence may vary by scientific field (Thelwall & Kousha, 2026), we apply a field-specific filter that restricts the analysis to terms whose relative frequency increased by at least 300%: i.e., at least quadrupled in use from 2021 to 2024 and measured separately within each scientific field. We employ a high threshold as more modest shifts could reflect unrelated field-specific usage rather than any GenAI tool assistance. This approach helps to better isolate those terms whose diffusion is plausibly associated with GenAI adoption. Appendix Fig. 5 presents the final selection of terms by scientific field.

Since Scopus assigns scientific fields at the journal level, journals cross-listed in multiple fields cause their publications to inherit more than one scientific field assignment. Because linguistic baselines differ across fields, however, we identify GenAI markers and thresholds within field-year strata and use the following "any-field" rule: a publication is classified as GenAI-assisted if it is flagged—i.e., at least one of the filtered keywords in stemmed form appears in its title or abstract—in any of its assigned fields. All remaining publications are classified as non-GenAI-assisted. We adopt this "any-field" rule because GenAI assistance is inherently a property of the publication, which therefore requires a single, consistent treatment assignment regardless of how many fields the journal spans. Treating publication-field observations as independent so that the same publication may be classified both as GenAI-assisted and non-GenAI-assisted depending on the field is an inconsistent alternative. A stricter "all-field" rule, such as requiring all assigned fields to flag the publication, increases the risk that genuine GenAI-assisted publications are coded as non-GenAI-assisted whenever one assigned field fails to detect a marker. In our setting, such false negatives would place potentially treated publications into the comparison group. For this reason, and because our marker-based proxy is already conservative, we prefer an assignment rule that minimizes false negatives at the publication level. This approach does not imply that false positives are costless. More generally, when treatment misclassification is sufficiently severe, causal inference methods that compare treated and control groups may produce sign reversal in the estimated effect (Denteh & Kédagni, 2022). Appendix Table 9 therefore provides several empirical robustness tests that examine the sensitivity of our baseline results to the stricter "all-field" rule. The results change little, however, suggesting that misclassification bias is not driving our main findings. Importantly, we also examine the empirical robustness of this classification approach by applying

different frequency thresholds. More and less restrictive thresholds alter the final term composition used to identify GenAI-assisted publications, but our baseline findings remain qualitatively unchanged. Section Robustness provides a detailed discussion of these classification robustness tests.

Linguistic similarity measurement

We use text embeddings to measure the linguistic similarity between title-abstract texts by authors with institutional affiliations in non-English-speaking countries and by authors with institutional affiliations in the U.S. Modern text embeddings are numerical representations of text in a high-dimensional vector space, generated by neural language models trained on large corpora (Devlin et al., 2019; Mikolov et al., 2013). Texts with similar meaning and writing style are positioned closer together in this space, allowing similarity to be quantified through standard geometric measures such as cosine similarity. Because embeddings capture both style and content, all similarity comparisons are conditioned on field and publication year, which makes the resulting similarity measure more likely to reflect differences in writing style within comparable scientific contexts rather than broad field- or year-level differences in content. We use SciBERT rather than general BERT because it is pretrained from scratch on scientific text, employs a domain-specific vocabulary that represents technical terms more accurately, and better models the linguistic patterns typical of scientific writing (Beltagy et al., 2019).

For each publication, we construct an embedding by concatenating its title and abstract, tokenizing the resulting text with the pretrained SciBERT (`scivocab, uncased`) model from Hugging Face (Beltagy et al., 2019), and processing it with SciBERT using its maximum 512 token sequence length. Sequences shorter than 512 tokens are padded with the special [PAD] token, which is masked during mean pooling and therefore does not contribute to the resulting embedding. Sequences exceeding 512 tokens are truncated from the end, following standard BERT practice. In our analysis sample, between 0.5% and 9.0% of these concatenated texts require truncation depending on the field, with the median text length ranging from 209 to 350 tokens (Appendix Table 4 reports the full distribution by field). Truncation rates are highest in fields with longer abstracts (Veterinary Science with a median of 349 tokens and 9.0% truncated; Medicine with a median of 346 tokens and 8.3% truncated) and lowest in fields with shorter abstracts (Mathematics with a median of 221 tokens and 0.5% truncated; Economics with a median of 209 tokens and 0.8% truncated). Missing abstracts affect approximately 2% of publications overall, with somewhat higher rates in Medicine and Nursing (5–6%). These publications contribute their titles only. Missing titles affect fewer than 0.01% of publications. Overall, truncation removes only the end of long abstracts, and publications with missing abstracts still contribute their titles. Section Robustness nevertheless confirms that truncated and missing text do not materially affect our results in several empirical robustness tests.

From the model's final hidden states, we derive a 768-dimensional mean-pooled text vector by averaging token embeddings across the sequence. This approach is widely used for capturing overall text semantics with BERT models (Ortakci, 2024; Reimers & Gurevych, 2019). Embeddings are computed one publication at a time without gradient updates or fine-tuning, and serve as the basis for our analysis. We use L2-normalized embeddings and quantify linguistic proximity using cosine similarity, which we rescale to range from 0 to 1. Higher values indicate greater similarity in writing style and content between the compared texts. For each non-U.S. publication, we compute cosine similarities

Table 3 Selected variable descriptive statistics

Variable	Mean	SD	Min	Max
Linguistic similarity to U.S. benchmark	0.82	0.03	0.29	0.9
Number of authors	7.20	8.87	1	102
Year indicators (mean = share of sample)				
Year: 2021	0.23			
Year: 2022	0.24			
Year: 2023	0.26			
Year: 2024	0.27			
GenAI-assisted publication share by year				
Year: 2021	0.03			
Year: 2022	0.03			
Year: 2023	0.05			
Year: 2024	0.12			
Continent indicators				
Africa	0.04			
Americas	0.05			
Asia	0.59			
Europe	0.32			
Oceania	0.00			
Aggregated field indicators				
Engineering & technology	0.23			
Life sciences	0.32			
Physical sciences	0.38			
Social sciences	0.07			

This table reports descriptive statistics for the regression sample (N = 14,315,948 observations). The unit of observation is the publication-country-field combination. For binary variables, only the mean is reported. The nonzero GenAI-assisted publication shares in 2021–2022 reflect the presence of lexical markers later associated with GenAI-assisted writing that already existed before ChatGPT and should not be interpreted as evidence of GenAI use in those years

to U.S.–only publications: i.e., publications for which all authors have U.S. affiliations from the same scientific field and publication year. For example, a Swiss chemistry publication from 2022 is compared to U.S. chemistry publications from 2022. We then average the resulting pairwise similarities to obtain a similarity score for each publication-field pair relative to the field-year-specific U.S. benchmark corpus. These similarity scores form the basis of our empirical analysis, in which we test whether GenAI-assisted publications from non-English-speaking countries show greater convergence toward the U.S. benchmark post-launch of ChatGPT than non-GenAI-assisted publications from these same countries.

Section Robustness provides additional empirical robustness tests by preserving the same scientific-field match while varying the benchmark corpus of U.S.-only publications as follows: (1) only those in top journals (viz., top 10% by field-specific journal impact factor); (2) only those non-GenAI-assisted; and (3) only those from 2021 in the same scientific field (i.e., predating ChatGPT's launch).

Empirical strategy

We briefly present descriptive context of the final sample used in the analysis. Table 3 provides summary statistics for the main variables. Linguistic similarity to U.S. scientific writing averages 0.82 ($SD = 0.03$), indicating high overall similarity with modest transnational variation.

The sample is distributed relatively evenly across 2021–2024, with a modest increase from 23% in 2021 to 27% in 2024. The share of GenAI-assisted publications grows progressively over the sample window: viz., from 3% in 2021 and 2022, to 5% in 2023, and to 12% in 2024—reflecting the diffusion of GenAI writing tools following the late-2022 ChatGPT launch. Most observations originate from Asia (59 percent) and Europe (32 percent); the modest Americas share (five percent) reflects our publication restriction of at least one author affiliation in a non-English-speaking country. Life Sciences (32 percent) and Physical Sciences (38 percent) represent roughly two-thirds of the sample, with Engineering and Technology (23 percent) and Social Sciences (7 percent) accounting for the remainder.

Figure 1 shows linguistic similarity trends for publications from non-English-speaking countries relative to U.S. publications at aggregated scientific field levels. Similarity levels remain largely stable from 2021 to 2022. Between 2022 and 2024, similarity rises in Engineering and Technology, Physical Sciences, and Social Sciences, with the largest

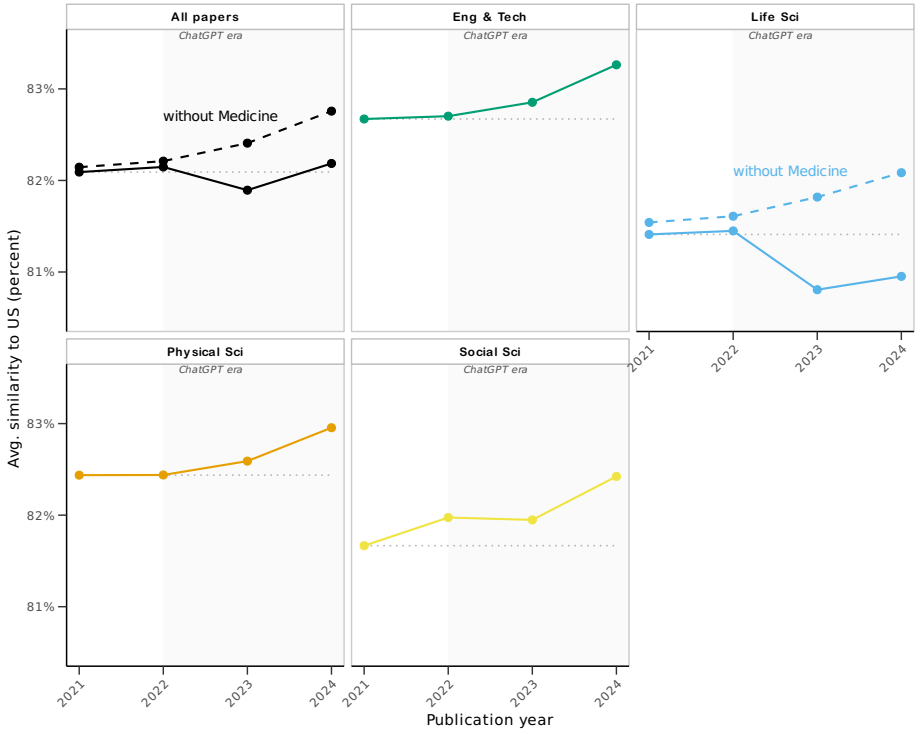


Fig. 1 Trends in linguistic similarity to U.S. scientific writing. This figure shows average linguistic similarity of publications from non-English-speaking countries to U.S. scientific writing, measured using SciBERT embeddings, and shown by publication year and aggregated scientific domain. Dashed lines denote values when publications in medicine are excluded

increase in Engineering and Technology. Life Sciences instead drop in 2023 and remain below their 2022 level in 2024. This pattern suggests that the SciBERT-based measure captures both linguistic and topical dimensions. A plausible explanation is that COVID-19-related research temporarily raised linguistic similarity over 2021–2022, but as such research receded in 2023, Life Sciences publications returned to more typical topics and subsequently lowered linguistic similarity. Calculations that exclude medicine publications from Life Sciences support this explanation, revealing a notable upward trend.

This descriptive context is informative, but does not establish whether the observed linguistic convergence reflects GenAI use or broader research topic, field composition, or editorial practice shifts. We therefore turn to a difference-in-differences (DiD) framework that compares marker-positive and marker-negative publications within the same journal-year cells and accounts for field- and country-level heterogeneity. DiD is a widely used econometric method that estimates the causal effect of an intervention by comparing changes in outcomes over time between a treated group (here, GenAI-assisted publications) and a control group (here, non-GenAI-assisted publications) (Cunningham, 2021). Under the identifying assumptions discussed below, this design helps distinguish changes driven by the diffusion of GenAI tools from broader compositional and stylistic shifts that may have occurred concurrently. To clarify the underlying identification logic and the remaining potential bias sources, we rely on a Directed Acyclic Graph (DAG) framework. DAGs provide a transparent way to represent causal pathways and statistical dependencies among variables (Brumback, 2021).

Appendix Fig. 4 uses the canonical 2×2 DiD DAG proposed by Knaus and Pfleiderer (2026), adapted to our empirical setting. *GenAI* denotes whether a publication is GenAI-assisted, proxied by our marker-based indicator. Y_{pre} and Y_{post} denote linguistic similarity to the U.S. scientific benchmark in the pre- and post-launch ChatGPT periods, respectively. Because our data consist of repeated cross-sections rather than the same publications observed over time, these variables should be understood as the relevant pre- and post-period outcomes in the stylized DiD setup, rather than repeated measurements of the same publication. By construction, each publication appears only in its publication year. X captures observed publication characteristics that may influence both GenAI use and linguistic similarity, such as coauthorship structure or affiliation with English-speaking countries. U represents unobserved factors, including latent writing ability, institutional support, funding, and other background conditions that may jointly affect observed publication characteristics, the probability of GenAI use, and linguistic similarity. The arrow from *GenAI* to Y_{post} represents the causal relationship of interest—whether GenAI-assisted publications exhibit greater convergence toward U.S. scientific writing after the launch of ChatGPT.

The DAG also clarifies the logic of our DiD identification strategy. In our repeated cross-section setting, it is not possible to difference out publication-specific unobserved heterogeneity as in a traditional panel data setup, because the same publication is not observed before and after treatment.[3] Instead, the estimator compares average changes in linguistic similarity between marker-positive and marker-negative publications over time. The design therefore accounts for stable differences in average outcomes between these two groups, but remains vulnerable to differential compositional changes that are not captured by observed covariates and fixed effects. Identification accordingly relies on a conditional

[3] In a traditional panel setting, the difference-in-differences estimator is based on changes within the same observational units over time. This differencing removes unit-specific unobserved characteristics that remain constant across periods.

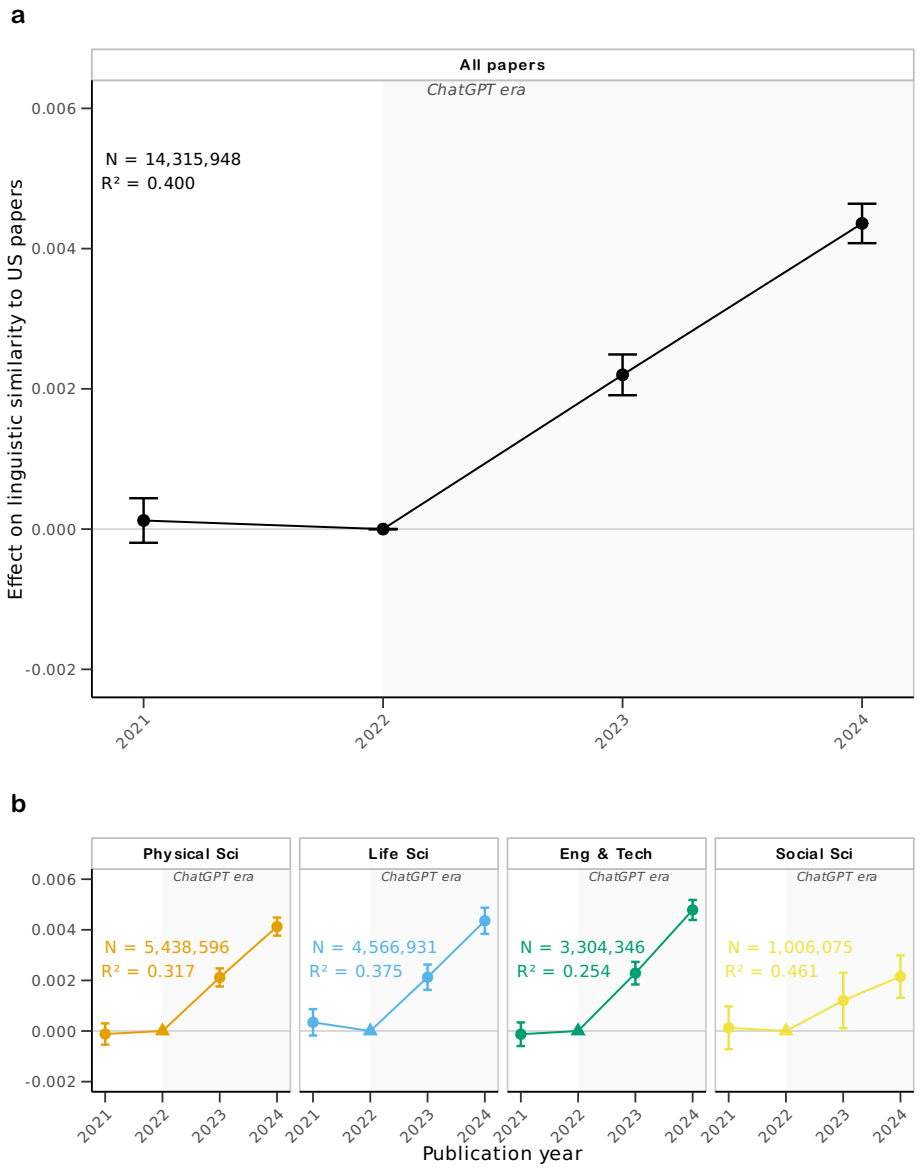


Fig. 2 GenAI-assisted writing and linguistic convergence to U.S. publications. This figure shows estimated effects based on the event-study DiD regression (Eq. 1), comparing GenAI-assisted and non-GenAI-assisted publications from non-English-speaking countries. Coefficients represent year-specific differences in linguistic convergence relative to the 2022 baseline. Panel **a** shows convergence overall; Panel **b** shows convergence by scientific field. All models include country, field, journal, year and journal-year fixed effects. Appendix Table 5 provides full regression results

parallel-trends assumption: i.e., absent the broader diffusion of GenAI tools, the average linguistic similarity of marker-positive and marker-negative publications would have evolved similarly over time, conditional on observables and fixed effects. Our event-study

specification helps assess the plausibility of this assumption by testing whether the two groups already exhibited divergent trajectories prior to 2023, when GenAI tools were not yet widely accessible to the broader public.

Figures 2 and 3 show no evidence of differential pre-period patterns—a result that holds across all analyses. In the context of our empirical design, the pre-2023 coefficients also function as a placebo test of our marker-based classification. If our keywords merely captured pre-existing stylistic preferences correlated with similarity to U.S. writing, marker-positive publications would be expected to differ systematically from marker-negative publications already before the widespread adoption of GenAI writing tools. The null

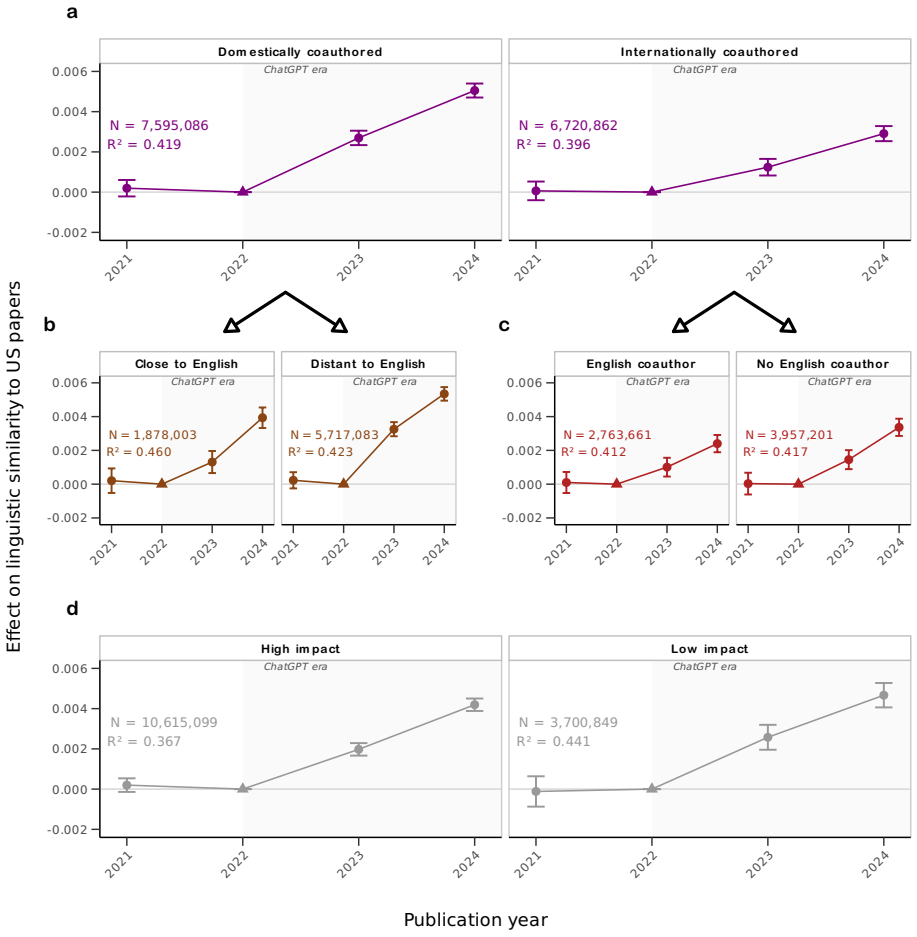


Fig. 3 Linguistic similarity convergence: subsample estimations. This figure shows estimated effects based on the event-study regression (Eq. 1), comparing GenAI-assisted to non-GenAI-assisted publications. Panel **a** contrasts domestically vs. internationally coauthored publications; Panel **b** contrasts countries linguistically close vs. linguistically distant from English within domestically coauthored publications; Panel **c** contrasts coauthor teams with vs. without an English-speaking coauthor within internationally coauthored publications; Panel **d** contrasts publications in high-impact vs. low-impact journals. All models include country, field, journal, year and journal-year fixed effects. Appendix Table 6 provides full regression results

pre-period coefficients therefore reduce the concern that our markers capture purely pre-existing stylistic variation, supporting the interpretation that observed post-2022 convergence reflects GenAI adoption rather than persistent stylistic differences between groups. At the same time, our fixed-effects structure absorbs all shocks common to publications appearing in the same journal-year, thereby accounting for time-varying journal-specific factors such as editorial practices, outlet-specific stylistic norms, and other contemporaneous changes at the journal level. The robustness of our findings to alternative GenAI classification rules and frequency thresholds further supports the interpretation that observed convergence reflects GenAI use rather than measurement artifacts.

Two caveats nevertheless remain worthy of mention. First, our treatment measure is marker-based and therefore captures observable traces of GenAI assistance rather than its full latent use. Section Identifying GenAI-Assisted Papers notes that this limitation is common to marker-based approaches in the literature. We address it by imposing a marker frequency threshold for classifying GenAI-assisted publications and by showing that the results remain stable across alternative thresholds. Alternatives that rely on AI-based detectors have been shown to be unreliable (Liang et al., 2023). Second, because the analysis is based on repeated cross-sections, the identifying assumption may still be violated if the composition of marker-positive and marker-negative publications changes differentially over time in ways that are not fully absorbed by observables and fixed effects (Ye et al., 2025). This concern is partly mitigated by our broad coverage of Scopus-indexed publications and by the inclusion of journal-by-year fixed effects, which ensure our comparisons are robust to shifting editorial preferences or changes in topical composition over time. The estimates should therefore be interpreted as credible evidence under standard DiD assumptions, with the limitations of marker-based treatment assignment and potential compositional shifts borne in mind.

To formalize this identification strategy, we implement an event-study DiD design with repeated cross-sections, comparing marker-positive and marker-negative publications relative to the U.S. benchmark corpus across calendar years, with 2022 as the reference year. The specification is given by:

$$\begin{aligned} y_{ifcjt} = \delta_0 \cdot \mathrm{GenAI}_i + \sum_{\tau \in \mathcal{T} \setminus \{2022\}} \beta_\tau \cdot \mathbf{1}\{t = \tau\} \times \mathrm{GenAI}_i \\ + \gamma_1 \cdot \mathrm{Authors}_i + \gamma_2 \cdot \mathrm{Eng}_i \\ + \alpha_c + \alpha_f + \alpha_j + \alpha_t + \alpha_{j \times t} + \varepsilon_{ifcjt}, \end{aligned} \tag{1}$$

The dependent variable y_{ifcjt} is our previously defined continuous similarity score for publication i in field f, country c, journal j, and year t. GenAI_i is an indicator equal to one if GenAI-related linguistic markers are detected in publication i and zero otherwise. Because baseline temporal changes are absorbed by our fixed effects, our main coefficients of interest, β_τ, are estimated from the interaction between GenAI_i and year indicators τ. These coefficients capture how the similarity gap between GenAI-assisted and non-GenAI-assisted publications in year τ differs from the corresponding gap in the 2022 reference year. For post-2022 years, positive values indicate that GenAI-assisted publications exhibit greater similarity to the U.S. benchmark relative to non-GenAI-assisted publications, compared with 2022. The 2021 interaction coefficient serves as a pre-trend check. We control at the publication level for the number of authors ($\mathrm{Authors}_i$) and whether at least one coauthor is affiliated with an English-speaking country (Eng_i). We also include a rich set of fixed effects to control for unobserved heterogeneity: α_c are country fixed effects that

absorb time-invariant country-level writing differences; α_f are Scopus field fixed effects (distinct from our aggregated groups) that absorb time-invariant field-specific writing styles; α_j are journal fixed effects that absorb outlet- and editorial-specific time-invariant factors; and α_t and $\alpha_{j\times t}$ are respective year and journal-by-year fixed effects that absorb common and journal-specific temporal shocks, such as shifts in the topics covered. Finally, ϵ_{ifcjt} represents an idiosyncratic error term. Standard errors are clustered at the journal level. We use an alpha level of 0.05 in all empirical analyses to determine statistical significance. Estimates with p-values below 0.05 are considered statistically significant, whereas those with p-values of 0.05 or above are not.

We estimate Equation (1) by weighted least squares using fractional publication weights ω_{ifc}. These weights are defined such that the weights of all country–field observations belonging to the same publication sum to one:

$$\sum_{(f,c)\in\mathcal{R}_i} \omega_{ifc} = 1,$$

where $\mathcal{R}_i$ denotes the set of country–field records associated with publication i. Equivalently, the estimator minimizes:

$$\sum_{i,f,c,j,t} \omega_{ifc}\left(y_{ifcjt} - \hat{y}_{ifcjt}\right)^2.$$

This weighting scheme prevents publications assigned to multiple fields or countries from receiving disproportionately large influence in the estimation.

GenAI use and convergence to U.S. english

This section presents the empirical results. Section 3.1 provides baseline analyses of linguistic convergence; Section 3.2 extends this analysis by taking into account heterogeneous factors that may affect linguistic convergence.

Baseline analyses

We first estimate Equation (1) in a baseline analysis that includes all controls and fixed effects. Because the U.S. benchmark is constructed separately by scientific field and publication year, it evolves over time. The journal-by-year fixed effects additionally absorb contemporaneous changes common to publications appearing in the same journal and year. Our estimates therefore isolate differential shifts in how GenAI-assisted and non-GenAI-assisted publications converge toward or diverge from the evolving U.S. benchmark, net of such journal-year-specific shocks.

Figure 2 shows that the launch of ChatGPT coincided with an increasing convergence in scientific writing styles—overall and by scientific field. Appendix Table 5 provides full regression results. Panel (a) indicates that GenAI-assisted non-U.S. publications became more linguistically similar to the U.S. benchmark post-2022 than comparable non-GenAI-assisted publications. This overall effect is both statistically significant and increasing in magnitude: i.e., by 0.2 percentage points in 2023 and by 0.4 percentage points in 2024, relative to the 2022 baseline. The absence of any pre-trend suggests that this shift emerges post-launch of ChatGPT. Linguistic similarity scale magnitudes are small but as expected:

viz., Fig. 1 indicates that the average exceeds 0.8. Perhaps more informative than the absolute magnitudes is the clear directional shift that emerges post-launch of ChatGPT. Because the embedding measure can reflect both language and topical content shifts, we estimate models with country and journal-year fixed effects. Comparisons are thus made within the same year and journal—an empirical design that ensures the observed convergence among GenAI-assisted publications is unlikely to be driven by topical composition changes. The patterns remain highly consistent across robustness tests that vary both the identification of GenAI-assisted publications and the linguistic similarity specifications—including stricter keyword thresholds, alternative U.S. benchmark corpora, and the exclusion of GenAI-assisted U.S. texts (see Section Robustness for full details and estimates).

Panel (b) indicates that linguistic convergence occurred across the scientific landscape. In particular, consistent and relatively large upward trends are observed in Engineering and Technology (0.5 percentage points), Physical Sciences (0.4 percentage points) and Life Sciences (0.4 percentage points), with a more modest increase in the Social Sciences (0.2 percentage points).

Heterogeneity analyses

We re-estimate Equation (1) in several heterogeneity analyses that consider other factors that might affect linguistic convergence. If GenAI acts as a converging force in scientific writing, its effects should be most pronounced for those researchers who have historically faced the greatest language-related publication barriers. Figure 3 provides results that are consistent with this prediction. Appendix Table 6 provides full regression estimates.

Panel (a) compares publications by domestic author teams (all same-country authors) with those by international author teams (at least one different-country author). International author teams typically bring broader complementary capabilities (that plausibly include language support) and achieve higher impact (Adams, 2012, 2013; Wuchty et al., 2007), in comparison to domestic author teams. Consistent with this argument, we find that the convergence effect of GenAI is significantly larger and rises more sharply for domestic author teams. By 2024, the estimated convergence is 0.5 percentage points for domestic teams and 0.3 percentage points for international teams. The difference is statistically significant (Appendix Table 6).

Panel (b) compares author team publications from countries linguistically close to English with those from countries linguistically distant from English. Although English dominates international scientific communication, its distance to other languages varies considerably and likely shapes both the initial need for GenAI support and the potential for linguistic convergence. We capture this dimension using the Common Language Index (CLI) that aggregates three linguistic components (Melitz & Toubal, 2014): (i) common native language (CNL) – measuring the probability that two randomly chosen individuals from two countries share the same mother tongue; (ii) common official language (COL) – capturing institutionalized support for communication via official status; and (iii) linguistic proximity (LP) – quantifying similarities between basic vocabularies based on the Automated Similarity Judgment Program (ASJP) (Wichmann et al., 2025). The resulting normalized score ranges from 0 to 1, where higher values indicate greater closeness to U.S. English. We split the sample at the median CLI score into proximate countries (above median) and distant countries (below median), and assess whether GenAI-driven linguistic convergence is stronger where researchers face larger inherent language barriers. Consistent with the interpretation that GenAI reduces language-related frictions, the convergence

effect is significantly larger for linguistically distant countries post-launch of ChatGPT. The estimated difference in convergence in 2024 between distant countries and proximate countries is statistically significant (Appendix Table 6).

Panel (c) contrasts author team publications that include at least one English-speaking coauthor with those that include none. Because such teams likely face fewer language frictions (Amano et al., 2023), we expect smaller gains where English-speaking expertise is present. Consistent with this argument, we find stronger convergence for author teams without any English-speaking coauthor. The difference between the two subsamples is statistically significant only in 2024 (Appendix Table 6), providing some evidence that this substitution pattern emerges with a lag. However, the smaller convergence effect among teams with an English-speaking coauthor could be specific to teams with a U.S.-based coauthor. Collaboration with a U.S.-based coauthor may already align a team's writing more closely with the U.S. English benchmark used in our analysis. We therefore further split the English-speaking-coauthor subsample into teams with a U.S. coauthor and teams without any U.S. coauthor. The estimated convergence effects are similar across the two groups, and their differences are not statistically significant (Appendix Table 7). If anything, the point estimates are slightly larger for teams with non-U.S. English-speaking coauthors. These results suggest that the weaker convergence among teams with an English-speaking coauthor is not driven specifically by teams with a U.S.-based coauthor.

Panel (d) considers this substitution mechanism across the scientific publication landscape by comparing low- and high-impact journals. The point estimates indicate stronger linguistic convergence in lower-impact journals (where authors may have less access to professional editing services) and weaker convergence in top-tier journals (where manuscripts may already undergo more extensive language polishing and thus face a potential ceiling effect). This pattern is consistent with previous evidence that shows peer review in high-impact journals is substantially more thorough—with longer reviews and greater attention to substantive aspects—than in lower-impact outlets (Severin et al., 2023). It also aligns with recent evidence from biomedicine indicating that less established researchers and lower-ranked institutions are among the primary GenAI tool adopters (Liu et al., 2025). However, the difference between high- and low-impact journals is not statistically significant (Appendix Table 6). We thus interpret these estimates as suggestive rather than definitive evidence that GenAI writing assistance may be particularly useful in publication contexts where language support is less readily available.

Taken together, the heterogeneity analyses indicate that GenAI-related convergence is concentrated in settings with greater language-related publication frictions. This pattern is statistically strongest for domestic author teams and for countries linguistically distant from English. The evidence is more nuanced in other dimensions, where differences by English-speaking-coauthor presence become statistically significant only in 2024.

Robustness

This section conducts a series of empirical robustness tests to assess the stability of the baseline and heterogeneity estimation results. These tests include alternative keyword thresholds, keyword threshold restrictions, U.S. publication corpus definitions, and sample restrictions based on text completeness.

Alternative Keyword Thresholds The baseline estimation identifies GenAI-assisted publications using a field-specific frequency threshold of 300% (i.e., occurrences increase by

a factor of four from 2021 to 2024) for GenAI-related terms. We test whether this result is sensitive to this threshold level: viz., thresholds too low may produce excessive false-positives; thresholds too high may produce excessive false-negatives. We re-estimate the baseline specification using looser (i.e., 200% or a three-fold occurrence increase) and stricter (i.e., 400% or a five-fold occurrence increase) thresholds.

Appendix Fig. 5 displays GenAI-linked words whose frequency increased relative to their pre-ChatGPT baseline disaggregated by scientific field. The figure includes only those keywords from our initial list that meet our lowest threshold of a three-fold increase. The shading indicates the magnitude of the increase: i.e., more words are included under the looser (3x) threshold; fewer words are included under the stricter (5x) threshold. Identification is again conducted at the scientific field level, meaning words are selected according to their abnormal post-2022 frequency growth within each field. The looser criterion captures a broader set of linguistic markers, ensuring that our findings are not driven only by highly-distinctive cases. The stricter criterion focuses exclusively on the most pronounced GenAI-related terms.

The left panels of Appendix Fig. 6 illustrate how the share of GenAI-assisted publications varies across these thresholds: Panel (a) uses the looser (200%) threshold; Panel (b) uses the baseline (300%) threshold; and Panel (c) uses the stricter (400%) threshold. Across all three thresholds, the share of GenAI-assisted publications follows a similar temporal pattern. The increase is roughly the same under the 300% and 400% thresholds and markedly larger under the 200% threshold.

The right panels of Appendix Fig. 6 present estimation results of linguistic similarity convergence across these alternative keyword thresholds. For each threshold, the core result of greater convergence in linguistic similarity between non-U.S. and GenAI-assisted publications and the U.S. corpus holds, relative to non-U.S. and non-GenAI-assisted publications. The variation in magnitudes across thresholds likely reflects differences in field composition and writing conventions among publications with stronger versus weaker GenAI signals. Importantly, these differences do not alter our main conclusion that convergence is robust to alternative keyword thresholds.

Alternative Keyword Threshold Restrictions The baseline estimation classifies a publication as GenAI-assisted if at least one GenAI-linked keyword appears in its title or abstract. This definition maximizes sensitivity to early stylistic shifts, but may also capture publications where such language reflects authors' unique writing style rather than genuine GenAI assistance. We examine this directly by implementing a stricter criterion that requires at least two distinct GenAI-related terms to be present. Publications flagged only by a single keyword are excluded from the sample. This robustness test ensures that our findings are not driven by marginal cases where GenAI-linked keywords appear without reflecting systematic adoption of GenAI writing support. Convergence that persists under this stricter definition provides additional support that the observed effects capture meaningful GenAI-related shifts in linguistic style rather than noise from spurious single-word matches.

Appendix Fig. 7 presents estimation results of linguistic similarity convergence, directly comparing the different identification approaches. The left panel reports results for publications containing exactly one GenAI-related term, whereas the right panel reports results for publications containing at least two such terms. Under the stricter identification definition, the point estimates are larger in both 2023 and 2024 than among publications containing exactly one GenAI-related term. The estimates are statistically significant in both years relative to the 2022 reference year. This pattern provides additional evidence that the baseline results are not driven by isolated single-term matches.

Alternative U.S. Publication Corpus Definitions A critical assumption underlying our baseline estimation is that our set of U.S.-authored benchmark publications serves as an accurate and stable reference for "standard" scientific English. To ensure that our findings are not mechanically driven by how this benchmark is defined, we consider three alternatives.

First, we re-compute similarity scores relative to a U.S. benchmark corpus from which publications flagged as GenAI-assisted have been excluded. This test is used to verify that convergence patterns are not an artifact of GenAI adoption in the benchmark itself, which could otherwise compress measured linguistic distances. Panel (a) in Appendix Fig. 8 presents estimation results that closely mirror those from our baseline estimation, indicating that the inclusion of GenAI-assisted publications in the benchmark corpus does not affect our findings.

Second, we restrict the benchmark to U.S. publications from 2021 (i.e., pre-launch of ChatGPT), thereby providing a temporally stable reference point that is unaffected by potential shifts in U.S. writing post-launch of ChatGPT. Panel (b) in Appendix Fig. 8 presents the estimation results and shows a similar pattern and pace of convergence over time compared with the baseline estimation.

Third, we construct a benchmark consisting of all U.S. publications appearing in the top 10% of journals within each Scopus field, ranked by impact factor. This allows for an assessment of whether the convergence pattern also holds relative to the linguistic register of leading publication outlets rather than being specific to the broader U.S. benchmark. Panel (c) in Appendix Fig. 8 presents the estimation results and shows a similar pattern of convergence over time to the baseline specification, but with larger point estimates. The larger point estimate, particularly, in 2024 may partly reflect the greater linguistic homogeneity of top-journal publications, which provides a more concentrated benchmark against which convergence is easier to detect.

Restricting to Complete Title-Abstract Sample To verify that truncation and missing-text issues do not drive our results, we re-estimate Equation (1) on a restricted sample that excludes both publications with truncated abstracts (those exceeding 512 tokens) and publications with missing titles or abstracts. This restriction reduces the sample by 3.2% (from 14,315,948 to 13,851,454 observations). Appendix Table 10 reports the results, which are qualitatively identical to our baseline estimates and confirm that data limitations do not drive the observed convergence.

These robustness tests help strengthen the credibility of our empirical strategy by demonstrating that the observed convergence of GenAI-assisted texts does not hinge on specific choices regarding keyword thresholds, identification rules, U.S. benchmark corpus definitions, or text completeness restrictions.

Discussion

The convergence patterns documented above are strongest among domestic author teams from linguistically distant countries. Point estimates are also larger in lower-impact journals, although the difference is not statistically significant. These results suggest that GenAI may reduce some language-related barriers in global science by narrowing the gap toward U.S. scientific writing norms, with disproportionately larger associations where linguistic and institutional resources are most constrained. This interpretation is consistent with complementary evidence showing asymmetric productivity gains from LLM adoption

(Filimonovic et al., 2025; Kusumegi et al., 2025) and readability improvements among non-native English-speaking authors (Prakash et al., 2025). Our analysis also extends prior studies focused on single fields or preprint data (Lin et al., 2025; Liu et al., 2025) to published journal articles across four broad scientific domains.

We interpret these findings using insights from international trade theory. This literature illustrates how lowering structural barriers can simultaneously expand access, diversify participation, and challenge established structural advantages (Amiti & Khandelwal, 2013; Medina, 2024; Melitz, 2003). Through this lens, GenAI may act as a technology that reduces language frictions. By doing so, it may reduce one barrier to participation in international journals for authors from outside the Anglophone core. This shift could gradually alter the historical advantages long held by dominant groups (Amano et al., 2021; Nuñez & Amano, 2021). As a result, broader inclusion could make highly competitive publication environments even more crowded. This dynamic may dilute the historical linguistic premium enjoyed not only by native English speakers but also by highly proficient authors who have invested heavily in mastering the language. Linguistic convergence may therefore produce both potential inclusionary shifts and associated adjustments within global science.

The same trade dynamics that drive global market integration also frequently lead to market homogenization and the crowding out of local markets (Melitz, 2003). Two channels may be especially important in the context of scientific writing. First, just as global systems often favor standardized formats, GenAI-assisted writing may reduce stylistic variety within English scholarship, producing language that is more uniform but potentially less expressive. This concern may be particularly relevant in fields where argumentative voice and rhetorical variation play a central role. Second, the reallocation of research toward English as the dominant global lingua franca may have implications for local venues. By lowering the implicit submission barriers to English journals, GenAI could weaken incentives to publish in regional and domestic outlets, with potential consequences for the linguistic plurality of the scholarly ecosystem.

Our study has some limitations. First, our GenAI identification strategy relies on field-specific frequency shifts of lexical markers in titles and abstracts. Although we validate empirical robustness using stricter thresholds and alternative benchmarks (e.g., excluding GenAI-assisted U.S. publications, using a 2021 U.S. benchmark, and restricting to U.S. publications in high-impact journals), full-text stylometry and validated benchmark datasets would further sharpen inference (Hao et al., 2026). A related concern is that our candidate keyword set is adopted from a specific subset of the literature (Feyzollahi & Rafizadeh, 2025; Filimonovic et al., 2025; Uribe & Maldupa, 2024) and may not capture all GenAI markers identified by alternative approaches such as excess-vocabulary methods (Kobak et al., 2025). This incomplete identification of GenAI-assisted publications likely attenuates our estimates toward zero under classical measurement error in a binary treatment, suggesting that our findings are conservative. A promising direction for future work is therefore to construct a set of GenAI-associated terms that have been identified across multiple studies and detection approaches, and to validate whether these terms reliably identify GenAI-assisted writing using benchmark corpora in which GenAI involvement

is known with high confidence. Second, we infer language background from institutional affiliation, treating affiliations in non-English-speaking countries as non-English-speaking and U.S. affiliations as the English-language benchmark. This approach does not capture individual language proficiency and therefore introduces measurement noise. Third, our design is observational. Linking author-level adoption to outcomes (e.g., acceptance rates, time to publication, citations, and reviewer text) and exploiting natural experiments (e.g., institutional rollouts and disclosure mandates) would strengthen causal claims. Finally, deeper understanding of field-specific norms, disclosure behavior, and interactions between human editing and GenAI assistance will be important for assessing how language technologies reshape scientific practice.

Conclusion

This study examines whether the diffusion of GenAI in scientific writing drives measurable linguistic convergence toward dominant benchmarks of scientific English. Specifically, it examines whether non-U.S. publications identified as GenAI-assisted exhibit stronger similarity to U.S. scientific writing than comparable non-U.S. and non-GenAI-assisted publications, and whether such patterns vary systematically across linguistic and team-composition contexts.

Using a large-scale bibliometric dataset of 5.65 million scientific publications representing 14 million unique publication-country-field observations over a pre- and post-launch ChatGPT window, we document two main findings. First, non-U.S. and GenAI-assisted publications show statistically significant and increasing convergence toward U.S. scientific writing post-launch of ChatGPT, relative to non-U.S. and non-GenAI-assisted publications. Second, this convergence is strongest among domestic author teams from countries linguistically distant from English. These patterns are robust to alternative keyword thresholds, stricter identification criteria, different U.S. benchmark corpus definitions, and truncation and missing text sub-sample restrictions.

The study makes several contributions. Methodologically, we combine lexical markers unique to specific fields with embedding similarity measures to provide a scalable framework for tracking stylistic shifts in published science. Conceptually, we reframe the empirical study of how GenAI affects science by shifting the focus from absolute changes in vocabulary or readability toward directional convergence. As the empirical results show, this alternative lens reveals dynamics that remain invisible in prior approaches.

The patterns documented are robust across a range of specifications and contexts, but they speak to the linguistic surface of scientific writing rather than to changes in the quality, originality, or impact of the research. Whether this linguistic convergence translates into deeper changes in the trajectory of scientific discovery is a separate question that we do not address but raise as an important agenda item for future research.

Appendix A

See Tables 4, 5, 6, 7, 8, 9, 10 and Figs. 4, 5, 6, 7 and 8.

Table 4 Token length and missing data statistics by field

Field	Median tokens	Truncated (%)	Missing abstract (%)	Missing title (%)
Agriculture	329	6.1	1.1	0.000
Biochemistry, genetics and molecular biology	307	4.2	1.1	0.000
Business	245	1.3	2.9	0.000
Chemical engineering	285	1.4	0.3	0.000
Chemistry	277	1.7	0.3	0.000
Computer science	262	1.0	0.3	0.001
Decision sciences	249	1.2	0.9	0.000
Dentistry	350	6.3	2.9	0.002
Earth sciences	326	7.2	0.8	0.000
Economics	209	0.8	1.7	0.000
Energy studies	282	1.7	0.4	0.000
Engineering	268	1.5	0.6	0.000
Environmental science	318	3.4	0.6	0.000
Public health	335	6.6	2.9	0.006
Immunology	325	6.0	1.3	0.000
Materials science	274	1.5	0.3	0.000
Mathematics	221	0.5	0.4	0.000
Medicine	346	8.3	5.5	0.001
Neuroscience	320	4.2	1.8	0.000
Nursing	328	5.2	5.7	0.002
Pharmacology	323	5.1	1.4	0.003
Physics	255	1.7	0.3	0.000
Psychology	274	1.6	1.8	0.000
Sociology	251	1.5	5.3	0.000
Veterinary science	349	9.0	2.7	0.000

This table reports statistics computed across all publications in our analysis sample (2021–2024) by Scopus field. Truncation occurs when the title-abstract concatenation exceeds SciBERT's 512-token limit

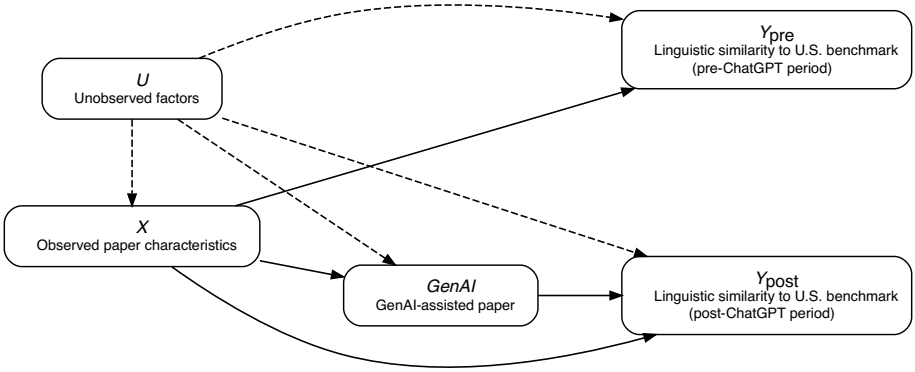


Fig. 4 Directed acyclic graph. This graph summarizes the main relationships between treatment, outcomes, observed covariates, and potential unobserved factors in our empirical setting. The primary channel of interest is the effect of GenAI assistance on post-ChatGPT linguistic similarity, $GenAI \rightarrow Y_{\text{post}}$. Dashed arrows denote potential unobserved factors (U)

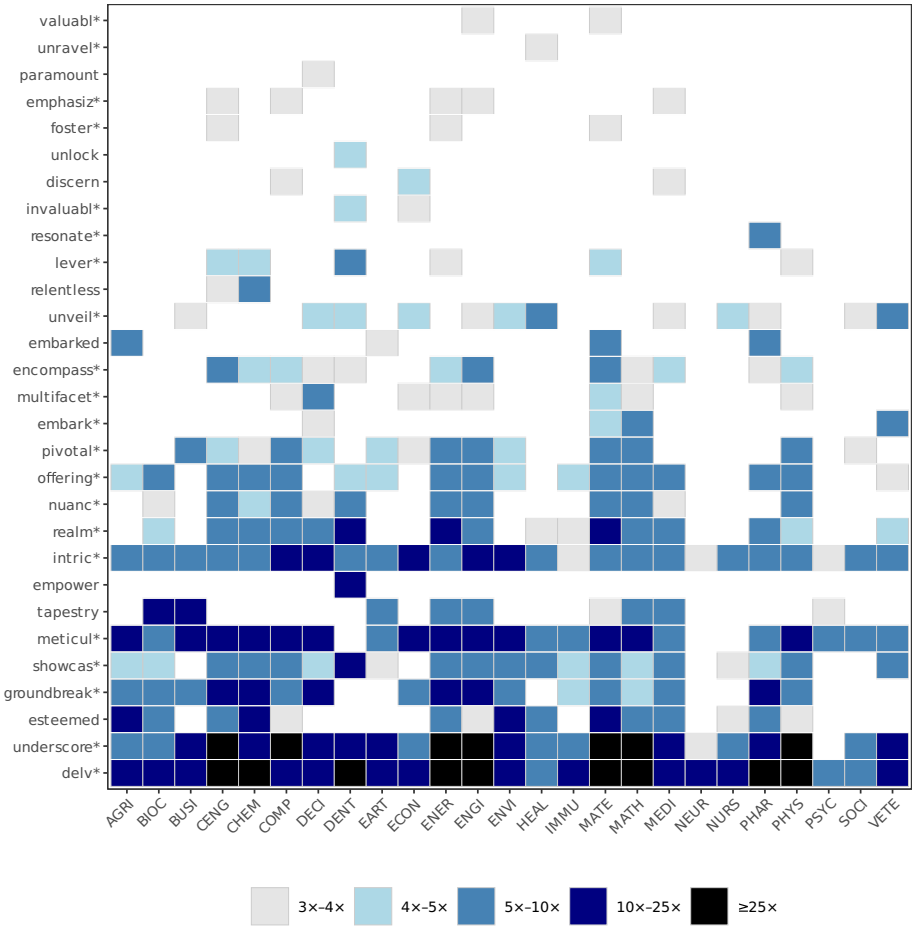


Fig. 5 Change in GenAI-related term frequency over time and by field. This heatmap displays the change in the frequency of GenAI-related terms across scientific fields between 2021 and 2024. Each row corresponds to a linguistic pattern (GenAI-related term) and each column to a field. The shading indicates the fold-change in usage, with darker colors representing larger increases. Values are grouped into several change categories based on their magnitude (3–4×, 4–5×, 5–10×, 10–25×, and ≥ 25×). GenAI-related terms in Table 2 that did not reach a three-fold increase in any field are excluded

Table 5 GenAI-assisted writing and linguistic convergence to U.S. publications

Sample	(1)	(2)	(3)	(4)	(5)
Variables					
Year 2021 × GenAI	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Year 2023 × GenAI	0.002**	0.002**	0.002**	0.002**	0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Year 2024 × GenAI	0.004**	0.005**	0.004**	0.004**	0.002**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GenAI paper	0.004**	0.004**	0.005**	0.003**	0.005**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of authors	0.000**	0.000	0.000**	0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
English-speaking coauthor	0.004**	0.004**	0.004**	0.004**	0.003**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Fixed-effects					
Country	Yes	Yes	Yes	Yes	Yes
Main field	Yes	Yes	Yes	Yes	Yes
Journal	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes
Year-journal	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	14,315,948	3,304,346	4,566,931	5,438,596	1,006,075
R^2	0.400	0.254	0.375	0.317	0.461
Within R^2	0.007	0.012	0.011	0.009	0.006

The dependent variable is the linguistic similarity of a paper to U.S. scientific writing, measured using SciBERT embeddings. Column (1) reports results for the full sample of papers. Columns (2)–(5) report estimates for the respective scientific field groups: Engineering & Technology, Life Sciences, Physical Sciences, and Social Sciences. *GenAI* indicates papers classified as GenAI-assisted using the marker-based approach. Coefficients from the event-study DiD specification capture differences between GenAI and non-GenAI papers by publication year relative to 2022 (reference year). All models include country, main field, journal, year, and year–journal fixed effects. Standard errors are clustered at the journal level and reported in parentheses. Statistical significance: * $p < 0.05$, ** $p < 0.01$

Table 6 Linguistic convergence: subsample estimations

Sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables								
2021 × GenAI	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
2023 × GenAI	0.003**	0.001**	0.001**	0.003**	0.001**	0.001**	0.002**	0.003**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
2024 × GenAI	0.005**	0.003**	0.004**	0.005**	0.002**	0.003**	0.004**	0.005**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GenAI paper	0.004**	0.004**	0.004**	0.004**	0.004**	0.004**	0.004**	0.005**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of authors	0.000**	0.000	0.001**	0.000**	0.000	0.000	0.000*	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
English-speaking coauthor		0.002**					0.004**	0.003**
		(0.000)					(0.000)	(0.000)
Fixed-effects								
country	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
main_field	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
journal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
year_journal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics								
Observations	7,595,086	6,720,862	1,878,003	5,717,083	2,763,661	3,957,201	10,615,099	3,700,849
R^2	0.419	0.396	0.460	0.423	0.412	0.417	0.367	0.441
Within R^2	0.007	0.005	0.006	0.007	0.002	0.004	0.008	0.007
Differences across paired samples								
Difference in 2023 × GenAI	0.0015**		−0.0019**		−0.0004		−0.0006	
	(0.0003)		(0.0004)		(0.0004)		(0.0004)	

Table 6 (continued)

Sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Difference in 2024 × GenAI	0.0021**		−0.0014**		−0.0010**		−0.0005	
	(0.0003)		(0.0004)		(0.0004)		(0.0003)	

The dependent variable is the linguistic similarity of a paper to U.S. scientific writing, measured using SciBERT embeddings. Columns (1)–(8) report estimates for the following groups: domestically coauthored papers (Column (1)), internationally coauthored papers (Column (2)), papers from countries linguistically close to English (Column (3)), papers from countries linguistically distant from English (Column (4)), papers with at least one English-speaking coauthor (Column (5)), papers without an English-speaking coauthor (Column (6)), papers published in high-impact journals (Column (7)), and papers published in low-impact journals (Column (8)). *GenAI* indicates papers classified as GenAI-assisted using the marker-based approach. Coefficients from the event-study DiD specification capture differences between GenAI and non-GenAI papers by publication year relative to 2022, the reference year. All models include country, main field, journal, year, and year–journal fixed effects. Standard errors are clustered at the journal level and reported in parentheses. The English-speaking coauthor indicator is omitted in specifications where it is collinear with the subsample definition. The bottom panel reports difference-in-coefficients z-tests across paired subsamples. The reported differences are Column (1) minus Column (2), Column (3) minus Column (4), Column (5) minus Column (6), and Column (7) minus Column (8), respectively. Statistical significance: * $p < 0.05$, ** $p < 0.01$

Table 7 Linguistic convergence: U.S. versus other english-speaking coauthors

Sample	(1) U.S. coauthor	(2) Other english-speaking coauthor
Variables		
2021 × GenAI	0.000	0.000
	(0.000)	(0.000)
2023 × GenAI	0.001*	0.001**
	(0.000)	(0.000)
2024 × GenAI	0.002**	0.003**
	(0.000)	(0.000)
GenAI paper	0.004**	0.003**
	(0.000)	(0.000)
Number of authors	0.000	0.000**
	(0.000)	(0.000)
Fixed effects		
Country	Yes	Yes
Main field	Yes	Yes
Journal	Yes	Yes
Year	Yes	Yes
Year × journal	Yes	Yes
Fit statistics		
Observations	1,604,539	1,159,122
R^2	0.435	0.442
Within R^2	0.002	0.002
Differences across paired samples		
Difference in 2023 × GenAI	−0.0004	
	(0.0006)	
Difference in 2024 × GenAI	−0.0009	
	(0.0005)	

The dependent variable is the linguistic similarity of a paper to U.S. scientific writing, measured using SciBERT embeddings. The sample is restricted to internationally coauthored papers with at least one coauthor from an English-speaking country. Column (1) includes papers with at least one U.S. coauthor, including papers that also have a coauthor from another English-speaking country. Column (2) includes papers with at least one coauthor from the United Kingdom, Canada, Australia, or Ireland, but no U.S. coauthor. *GenAI* indicates papers classified as GenAI-assisted using the marker-based approach. Coefficients from the event-study DiD specification capture differences between GenAI and non-GenAI papers by publication year relative to 2022, the reference year. All models include country, main-field, journal, year, and year–journal fixed effects. Standard errors are clustered at the journal level and reported in parentheses. The bottom panel reports difference-in-coefficients z-tests across the two subsamples. The reported differences are Column (1) minus Column (2). The corresponding p-values are 0.536 for 2023 and 0.073 for 2024. Statistical significance: * $p < 0.05$, ** $p < 0.01$

Table 8 Linguistic convergence: alternative publication counting

Sample	Fractional count (1)	Full count (2)	First-first (3)	Same-same (4)
Variables				
2021 × GenAI	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
2023 × GenAI	0.002**	0.002**	0.002**	0.004**
	(0.000)	(0.000)	(0.000)	(0.000)
2024 × GenAI	0.004**	0.004**	0.005**	0.006**
	(0.000)	(0.000)	(0.000)	(0.000)
GenAI paper	0.004**	0.004**	0.004**	0.004**
	(0.000)	(0.000)	(0.000)	(0.000)
Number of authors	0.000**	0.000	0.000**	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
English-speaking coauthor	0.004**	0.003**	0.003**	
	(0.000)	(0.000)	(0.000)	
Fixed-effects				
Country	Yes	Yes	Yes	Yes
Main field	Yes	Yes	Yes	Yes
Journal	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Year-journal	Yes	Yes	Yes	Yes
Fit statistics				
Observations	14,315,948	14,315,948	5,379,278	1,508,034
R^2	0.400	0.366	0.410	0.463
Within R^2	0.007	0.007	0.007	0.009

The dependent variable is the linguistic similarity of a paper to U.S. scientific writing, measured using SciBERT embeddings. The table reports robustness checks using alternative ways of assigning publications to countries and scientific fields. Column (1) uses fractional weights for publication–country–field observations (our baseline). Column (2) treats each publication–country–field cell as an independent observation. Column (3) assigns each publication to the first author's country and to the first-listed scientific field in the Scopus classification. Column (4) restricts the sample to publications whose authors are all affiliated with the same country and that are assigned to only one Scopus field. *GenAI paper* indicates publications that are identified as GenAI assisted. The event-study coefficients report differences between GenAI and non-GenAI papers by publication year relative to 2022, the reference year. All models include country, main-field, journal, year, and year-journal fixed effects. Standard errors are clustered at the journal level and reported in parentheses. The English-speaking coauthor indicator is omitted in Column (4) because of lack of identifying variation in the restricted sample. Statistical significance: * $p < 0.05$, ** $p < 0.01$

Table 9 Linguistic convergence: alternative GenAI flag

Sample	(1)	(2)	(3)	(4)	(5)
Variables					
Year 2021 × GenAI	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Year 2023 × GenAI	0.003**	0.002**	0.002**	0.002**	0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Year 2024 × GenAI	0.005**	0.005**	0.004**	0.004**	0.002**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GenAI paper	0.004**	0.004**	0.005**	0.003**	0.005**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of authors	0.000**	0.000	0.000**	0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
English-speaking coauthor	0.004**	0.004**	0.004**	0.004**	0.003**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Fixed-effects					
Country	Yes	Yes	Yes	Yes	Yes
Main field	Yes	Yes	Yes	Yes	Yes
Journal	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes
Year-journal	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	14,315,948	3,304,346	4,566,931	5,438,596	1,006,075
R^2	0.399	0.254	0.375	0.317	0.461
Within R^2	0.007	0.012	0.011	0.009	0.006

Note: The dependent variable is the linguistic similarity of a paper to U.S. scientific writing, measured using SciBERT embeddings. Column (1) reports results for the full sample of papers. Columns (2)–(5) report estimates for the respective scientific field groups: Engineering & Technology, Life Sciences, Physical Sciences, and Social Sciences. *GenAI* indicates papers classified as GenAI-assisted using the marker-based approach. In this robustness check, the GenAI indicator is assigned using an "all-field" rule, i.e. for papers assigned to multiple Scopus fields, a paper is classified as GenAI only if the number of detected GenAI-related keywords is positive in all fields assigned to that paper. Coefficients from the event-study DiD specification capture differences between GenAI and non-GenAI papers by publication year relative to 2022, the reference year. The coefficient on *GenAI paper* captures the baseline GenAI–non-GenAI difference in 2022. All models include country, main-field, journal, year, and year–journal fixed effects. Standard errors are clustered at the journal level and reported in parentheses. Statistical significance: * $p < 0.05$, ** $p < 0.01$

Table 10 Linguistic convergence: complete title-abstract sample

Sample	(1)	(2)	(3)	(4)	(5)
Variables					
Year 2021 × GenAI	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Year 2023 × GenAI	0.002**	0.002**	0.002**	0.002**	0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Year 2024 × GenAI	0.004**	0.005**	0.004**	0.004**	0.002**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GenAI paper	0.004**	0.004**	0.004**	0.003**	0.005**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of authors	0.000**	0.000	0.000**	0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
English-speaking coauthor	0.004**	0.004**	0.004**	0.004**	0.003**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Fixed-effects					
Country	Yes	Yes	Yes	Yes	Yes
Main field	Yes	Yes	Yes	Yes	Yes
Journal	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes
Year-journal	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	13,851,454	3,266,422	4,273,379	5,320,608	991,045
R^2	0.404	0.257	0.387	0.321	0.462
Within R^2	0.008	0.012	0.012	0.009	0.006

The sample is restricted to papers for which both title and abstract are available and the abstract is not truncated by the 512-token processing limit. The dependent variable is the linguistic similarity of a paper to U.S. scientific writing, measured using SciBERT embeddings. Column (1) reports results for the full sample of papers. Columns (2)–(5) report estimates for the respective scientific field groups: Engineering & Technology, Life Sciences, Physical Sciences, and Social Sciences. *GenAI* indicates papers classified as GenAI-assisted using the marker-based approach, assigned using the baseline "any-field" rule. Coefficients from the event-study DiD specification capture differences between GenAI and non-GenAI papers by publication year relative to 2022, the reference year. The coefficient on *GenAI paper* captures the baseline GenAI–non-GenAI difference in 2022. All models include country, main-field, journal, year, and year–journal fixed effects. Standard errors are clustered at the journal level and reported in parentheses. Statistical significance: * $p < 0.05$, ** $p < 0.01$

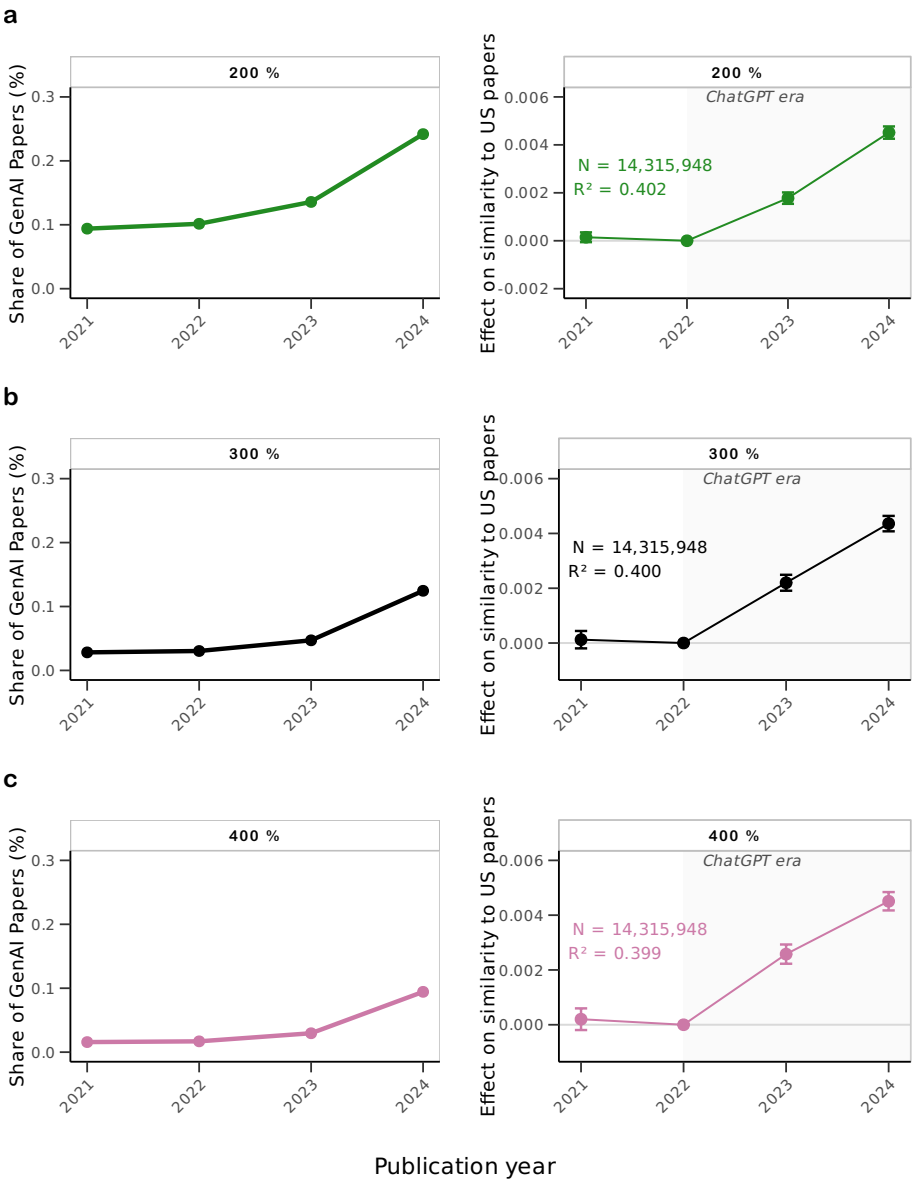


Fig. 6 Linguistic similarity convergence: alternative keyword thresholds. Panels **a**–**c** apply three alternative definitions for identifying GenAI-assisted publications based on changes in keyword frequency between 2021 and 2024: 200% **a**, 300% (b; baseline), and 400% **c** growth. In each panel, the left subfigure shows the evolution of the share of publications flagged as GenAI over time. The right subfigure reports coefficients from the event-study specification described in Data and Methods. Estimated effects compare GenAI-assisted and non-GenAI-assisted publications originating from non-English speaking countries. The dependent variable is the SciBERT-based linguistic similarity to U.S. scientific writing. Coefficients represent period-specific differences relative to the 2022 reference year

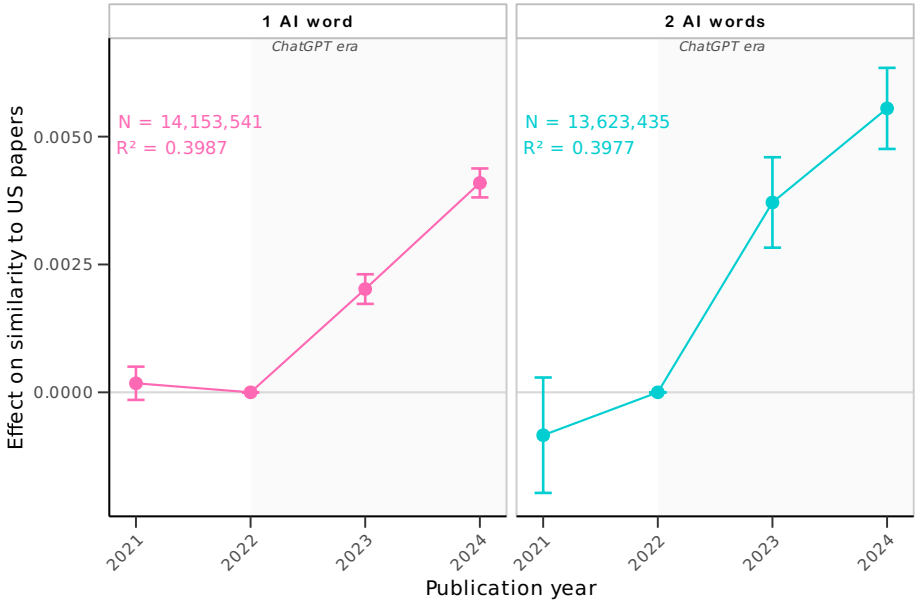


Fig. 7 Linguistic similarity convergence: stricter identification. Both panels report coefficients from the specification described in Data and Methods, i.e., our baseline model. The dependent variable is the SciBERT-based linguistic similarity to U.S. scientific writing, and estimates compare GenAI-assisted and non-GenAI-assisted publications from non-English-speaking countries relative to the 2022 reference year. The left panel defines GenAI uptake based on the presence of exactly one GenAI-related term in either the title or the abstract. The right panel uses a stricter definition, flagging only papers that contain at least two such terms; papers with exactly one occurrence are excluded from both treatment and control groups

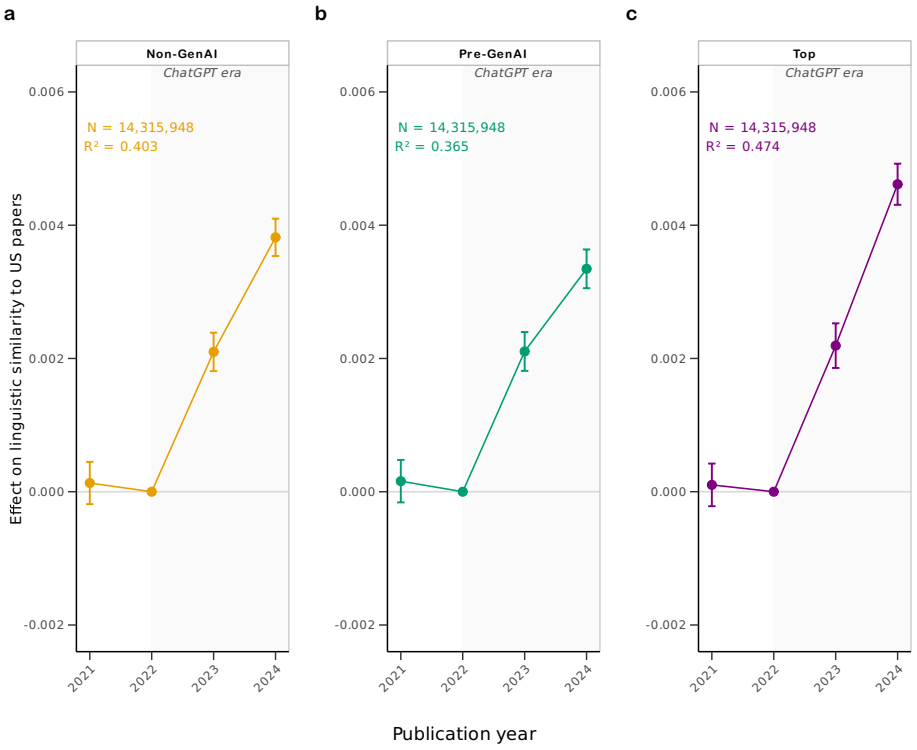


Fig. 8 Linguistic similarity convergence: various U.S. corpus benchmarks. Each panel reports estimates using the same specification described in Data and Methods, i.e., our baseline model. The dependent variable is the SciBERT-based linguistic similarity to U.S. scientific writing, and coefficients compare GenAI-assisted and non-GenAI-assisted publications from non-English-speaking countries relative to the 2022 reference year. The panels differ in how the U.S. benchmark corpus is defined: Panel **a** (*Non-GenAI benchmark*) computes similarity scores relative to a U.S. corpus from which GenAI-assisted publications have been removed. Panel **b** (*Pre-GenAI 2021 benchmark*) uses only U.S. publications from 2021, i.e., before the introduction of ChatGPT. Panel **c** (*Top-US benchmark*) restricts the benchmark to U.S. publications appearing in journals that belong to the top 10 percent within each Scopus field, ranked by impact factor

Acknowledgements We thank the editor, the guest editor and two anonymous referees for their thorough and very helpful feedback that helped us to improve the paper. We also thank participants at the Institutions and Innovation Conference (IIC) for helpful comments on earlier versions. We are also grateful for the general support by the Zaeslin-Bustany-Foundation.

Author contributions D.F. contributed primarily to the conceptualization, writing, and empirical analysis, and assisted with data creation. C.R. contributed primarily to the conceptualization, writing, and data creation, and assisted with the empirical analysis. J.M. contributed to the conceptualization and writing. R.W. contributed to the conceptualization and writing.

Funding Open access funding provided by University of Basel.

Data availability Raw Scopus data are subject to Elsevier licensing restrictions and cannot be shared publicly. Anonymized data will be made available upon request to enable replication.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Adams, J. (2012). The rise of research networks. *Nature, 490*(7420), 335–336.
Adams, J. (2013). The fourth age of research. *Nature, 497*(7451), 557–560.
Alafnan, M. A., & Mohdzuki, S. F. (2023). Do artificial intelligence chatbots have a writing style? An investigation into the stylistic features of ChatGPT-4. *Journal of Artificial Intelligence and Technology, 3*(3), 85–94.
Amano, T., Berdejo-Espinola, V., Christie, A. P., Willott, K., Akasaka, M., Báldi, A., & Sutherland, W. J. (2021). Tapping into non-english-language science for the conservation of global biodiversity. *PLoS Biology,19*(10), Article e3001296.
Amano, T., González-Varo, J. P., & Sutherland, W. J. (2016). Languages are still a major barrier to global science. *PLoS Biology, 14*(12), Article e2000933.
Amano, T., Ramírez-Castañeda, V., Berdejo-Espinola, V., Borokini, I., Chowdhury, S., Golivets, M., & Veríssimo, D. (2023). The manifold costs of being a non-native English speaker in science. *PLoS Biology*. https://doi.org/10.1371/journal.pbio.3002184
Amiti, M., & Khandelwal, A. K. (2013). Import competition and quality upgrading. *Review of Economics and Statistics, 95*(2), 476–490.
Balan, S. (2021). English as the language of research: But are we missing the mark? *Exploratory Research in Clinical and Social Pharmacy, 3*, Article 100043.
Barjak, F. (2006). Research productivity in the internet era. *Scientometrics, 68*(3), 343–360.
Beltagy, I., Lo, K., & Cohan, A. (2019). Scibert: A pretrained language model for scientific text arXiv:1903.10676.
Brumback, B. A. (2021). *Fundamentals of Causal Inference: With R* (1st ed.). Chapman and Hall/CRC.
Bubeck, S., Chandrasekaran, V., & Eldan, R. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4 arXiv:2303.12712.
Castañeda, V. R. (2020). Disadvantages in preparing and publishing scientific papers caused by the dominance of the english language in science: The case of colombian researchers in biological sciences. *PLoS ONE, 15*(9), 1–15.
Clavero, M. (2010). Awkward wording. Rephrase": Linguistic injustice in ecological journals. *Trends in Ecology & Evolution, 25*(10), 552–553.
Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.
da Fonseca Pachi, C. G., Yamamoto, J. F., da Costa, A. P. A., & Lopez, L. F. (2012). Relationship between connectivity and academic productivity. *Scientometrics, 93*(2), 265–278.
De Haan, S., Liu, Y., & Bollen, J. (2025). Gpt editors, not authors: The stylistic footprint of llms in academic preprints arXiv:2505.17327.
Denteh, A., & Kédagni, D. (2022). Misclassification in difference-in-differences models arXiv:2207.11890.
Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers) 4171–4186
Elsevier. (2023). The scopus content coverage guide: A complete overview of the content coverage in scopus and corresponding policies.
Feyzollahi, M., & Rafizadeh, N. (2025). The adoption of large language models in economics research. *Economics Letters*. https://doi.org/10.1016/j.econlet.2025.112265
Filimonovic, D., Rutzer, C., & Wunsch, C. (2025). Can genai improve academic performance? evidence from the social and behavioral sciences arXiv:2510.02408.
Fitria, T. N. (2021). Grammarly as ai-powered english writing assistant: Students' alternative for writing english. *The Journal of English Language and Literature, 5*, 65–78.

Flowerdew, J. (2022). Models of english for research publication purposes. *World Englishes, 41*(4), 571–583.
Geng, M., & Trotta, R. (2024). Is ChatGPT Transforming Academics' Writing Style? arXiv:2404.08627.
Ghufron, M. A., & Rosyida, F. (2018). The role of Grammarly in assessing English as a foreign language (EFL) writing. *Lingua Cultura, 12*(4), 395–403.
Gray, A. (2024). Chatgpt "contamination": estimating the prevalence of llms in the scholarly literature arXiv:2403.16887.
Hanauer, D. I., Sheridan, C. L., & Englander, K. (2019). Linguistic injustice in the writing of research articles in english as a second language: Data from taiwanese and mexican researchers. *Written Communication, 36*(1), 136–154.
Hao, Q., Xu, F., Li, Y., & Evans, J. (2026). Artificial intelligence tools expand scientists' impact but contract science's focus. *Nature, 649*(8099), 1237–1243.
House, J. (2003). English as a lingua franca: A threat to multilingualism? *Journal of Sociolinguistics, 7*(4), 556–578.
Ketcham, C. M., Hardy, R. W., Rubin, B., & Siegal, G. P. (2010). What editors want in an abstract. *Laboratory Investigation, 90*, 4–5.
Knaus, M. C., & Pfleiderer, H. (2026). Causal graphs for conditional parallel trends arXiv:2604.12818.
Kobak, D., González-Márquez, R., Horvát, EÁ., & Lause, J. (2025). Delving into llm-assisted writing in biomedical publications through excess vocabulary. Science Advances *11*(27), eadt3813.
Kousha, K., & Thelwall, M. (2026). How much are LLMs changing the language of academic papers after ChatGPT? A multi-database and full-text analysis. *Scientometrics*. https://doi.org/10.1007/s11192-026-05601-5
Kusumegi, K., Yang, X., Ginsparg, P., De Vaan, M., Stuart, T., & Yin, Y. (2025). Scientific production in the era of large language models. *Science, 390*(6779), 1240–1243.
Liang, W., Izzo, Z., & Zhang, Y. (2026). Monitoring ai-modified content at scale: A case study on the impact of chatgpt on ai conference peer reviews arXiv:2403.07183.
Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). Gpt detectors are biased against non-native english writers. *Patterns4*(7), Article 100779.
Lin, D., Zhao, N., & Tian, D. (2025). Chatgpt as linguistic equalizer? quantifying llm-driven lexical shifts in academic writing arXiv:2504.12317.
Liu, J., He, Y., & Zheng, Z. (2025). Ai-assisted writing is growing fastest among non-english-speaking and less established scientists arXiv:2511.15872.
Medina, P. (2024). Import competition, quality upgrading, and exporting: Evidence from the peruvian apparel industry. *Review of Economics and Statistics, 106*(5), 1285–1300.
Melitz, J., & Toubal, F. (2014). Native language, spoken language, translation and trade. *Journal of International Economics, 93*(2), 351–363.
Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica, 71*(6), 1695–1725.
Mikolov, T., Chen, K., & Corrado, G. (2013). Efficient estimation of word representations in vector space arXiv:1301.3781.
Montgomery, S. L. (2013). *Does science need a global language?: English and the future of research*. University of Chicago Press.
Mundt, K., & Groves, M. (2016). A double-edged sword: The merits and the policy implications of google translate in higher education. *European Journal of Higher Education, 6*(4), 387–401.
Nature Human Behaviour. (2023). Scientific publishing has a language problem. *Nature Human Behaviour,7*, 1019–1020.
Nuñez, M. A., & Amano, T. (2021). Monolingual searches can limit and bias results in global literature reviews. *Nature Ecology & Evolution, 5*(3), 264–264.
Ortakci, Y. (2024). Revolutionary text clustering: Investigating transfer learning capacity of sbert models through pooling techniques. *Engineering Science and Technology, an International Journal, 55*, Article 101730.
Perianes-Rodriguez, A., Waltman, L., & Van Eck, N. J. (2016). Constructing bibliometric networks: A comparison between full and fractional counting. *Journal of Informetrics, 10*(4), 1178–1195.
Plavén-Sigray, P., Matheson, G. J., Schiffler, B. C., Thompson, W. H. (2017). The readability of scientific texts is decreasing over time. eLife 6, e27725.
Politzer-Ahles, S., Girolamo, T., & Ghali, S. (2020). Preliminary evidence of linguistic bias in academic reviewing. *Journal of English for Academic Purposes, 47*, Article 100895.
Prakash, A., Aggarwal, S., Varghese, J. J., & Varghese, J. J. (2025). Writing without borders: Ai and cross-cultural convergence in academic writing quality. *Humanities and Social Sciences Communications, 12*(1), 1058.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (pp. 3982–3992). Association for Computational Linguistics.

Severin, A., Strinzel, M., Egger, M., Barros, T., Sokolov, A., Mouatt, J. V., & Müller, S. (2023). Relationship between journal impact factor and the thoroughness and helpfulness of peer reviews. *PLOS Biology,21*(8), Article e3002238.

Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of web of science, scopus and dimensions: A comparative analysis. *Scientometrics, 126*, 5113–5142.

Stephan, P., Franzoni, C., & Scellato, G. (2016). Global competition for scientific talent: Evidence from location decisions of phds and postdocs in 16 countries. *Industrial and Corporate Change, 25*(3), 457–485.

Tang, C., Li, S. K., Hu, S., Zeng, F., & Du, Q. (2025). Gender disparities in the impact of generative artificial intelligence: Evidence from academia. PNAS Nexus 4(2):pgae591.

Thelwall, M., & Kousha, K. (2026). Have llm-associated terms increased in article full texts in all fields? arXiv:2604.07565

Uribe, S. E., & Maldupa, I. (2024). Estimating the use of ChatGPT in dental research publications. *Journal of Dentistry, 149*, Article 105275.

Van Noorden, R., & Perkel, J. M. (2023). AI and science: What 1,600 researchers think. *Nature, 621*(7980), 672–675.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of informetrics, 10*(2), 365–391.

Wichmann, S., Holman, E. W., & Brown, C. H. (2025). *The ASJP Database*. version 21.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*(5827), 1036–1039.

Xu, X., & Reed, M. (2021). The impact of internet access on research output-a cross-country study. *Information Economics and Policy, 56*, Article 100914.

Ye, K., Bilinski, A., & Lee, Y. (2025). Difference-in-differences analysis with repeated cross-sectional survey data. Health Services and Outcomes Research Methodology.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.