publications

MDPI

Article

The Emergence of Generative AI in Scholarly Communication Through Lexical Classification, Disciplinary Diffusion, and Citation-Based Recognition

Carlos Hernán Suárez-Rodríguez *, Alba Mery Garzón-García * and Esteban Largo-Avila

Regionalization System, Universidad del Valle, Caicedonia 762540, Colombia; esteban.largo@correounivalle.edu.co

* Correspondence: carlos.hernan.suarez@correounivalle.edu.co (C.H.S.-R.); garzon.alba@correounivalle.edu.co (A.M.G.-G.)

Abstract

Generative artificial intelligence (GenAI) has gained prominence in scholarly communication, yet its lexical classification, disciplinary diffusion, semantic organization, and association with citation-based recognition remain understudied. This study analyzes 487,753 research and review articles from OpenAlex (August 2019–March 2026), combining rule-based lexical classification with interrupted time-series models, disciplinary mapping, keyword co-occurrence analysis, and citation-count models. The results show a discrete level change in the visibility of GenAI terminology after November 2022. This discontinuity remained robust under classification-error sensitivity analyses, whereas inference regarding the subsequent slope was more sensitive to classification assumptions. The proportion of GenAI-classified publications captured by high-specificity lexical rules increased from 86.65% before December 2022 to 93.64% afterward, consistent with greater lexical concentration. GenAI classification was broadly but unevenly distributed across fields, with the highest prevalence in Computer Science (49.42%) and Decision Sciences (38.57%). A curated keyword network showed overlapping thematic concentrations rather than sharply separated semantic subfields. Month-adjusted citation models showed a positive association between high-specificity classification and citation counts, with a smaller association with GenAI citations after December 2022. Overall, the study characterizes the temporal, lexical, disciplinary, thematic-network, and citation-related dimensions of GenAI's emergence in scholarly communication.

Keywords: disciplinary diffusion; generative artificial intelligence; interrupted time series; lexical classification; scholarly communication; semantic networks

Academic Editor: Andrew Kirby

Received: 15 August 2026
Revised: 10 September 2026
Accepted: 15 September 2026
Published: 18 September 2026

1. Introduction

GenAI has established a notable position in contemporary academic communication. Since the public deployment of conversational systems at the end of 2022, interest in GenAI has expanded across multiple scientific fields. This broader scholarly attention has led researchers to study it as a technology with the potential to reshape the processes of knowledge production, description, circulation, and evaluation. A steady stream of publications have addressed its methodological, ethical, educational, and epistemological implications (Quintans-Júnior et al., 2023; Ding et al., 2025; Khan et al., 2024; Alaqrabi & Alshagi, 2025).

However, the trajectory of GenAI in the scientific literature could go beyond increasing the volume of articles. Emerging research domains often show their evolution through

the gradual stabilization of their language, the transfer dynamics between disciplines, and recognition measured by citation practices. From this perspective, it has been suggested that academic communication functions as a space where scientific objects are semantically encoded, distributed within the disciplinary fabric, and progressively assimilated into the publishing system (Bawden, 2008; Mabe, 2010).

Under the framework of the science of science, scientific production is often modeled as a complex system where ideas, authors, publications, institutions, and citation networks interact (Fortunato et al., 2018; Krauss, 2024). Recent developments in this field also emphasize its transition toward large-scale, data-driven, and AI-assisted approaches, where scientometrics, citation analysis, complex networks, and computational methods are used to examine the internal dynamics of science and its societal role (Hou et al., 2025). Changes in this environment could not only alter the density of production but also the cognitive organization of knowledge and interdisciplinary boundaries (Fortunato et al., 2018). In this sense, the technical language reflected in titles, abstracts, and metadata can serve as an approximate empirical indicator to track the structure of research fields and their temporal evolution (Fortunato et al., 2018; Hou et al., 2025; Krauss, 2024).

Although the recent literature documents the rapid expansion of research on GenAI (Ding et al., 2025; Khan et al., 2024; Alaqrabi & Alshagi, 2025), and semantic change studies indicate that technological innovations often induce processes of terminological standardization (Hamilton et al., 2016; Sun et al., 2021; Rauch et al., 2024), there is still limited large-scale evidence regarding certain aspects of this phenomenon. Specifically, it remains to be explored in more detail whether the language associated with GenAI experienced structural variation after its public visibility, whether its expressions tended toward greater specificity, how it was distributed across disciplines, and in what ways citation patterns reflect its possible stabilization.

Addressing this gap is relevant because the emergence of a scientific domain does not depend solely on the accumulation of the literature, but also on processes of classification, thematic dispersion, temporal reconfiguration, and citation-based recognition (Bawden, 2008; Bornmann & Daniel, 2008; Larivière & Sugimoto, 2019). Examining these dimensions jointly provides a broader basis for understanding how an emerging research domain becomes visible, organized, and recognized within scholarly communication. In this study, citation patterns are interpreted as indicators of citation-based recognition rather than as direct measures of scientific quality.

In this context, the present work analyzes the emergence of GenAI through a corpus of 487,753 research and review articles retrieved from OpenAlex between 2 August 2019 and 30 March 2026. The study analyzes temporal, lexical, disciplinary, semantic-network, and citation-related dimensions through a pipeline that integrates rule-based lexical classification, interrupted time-series models, disciplinary mapping, term co-occurrence analysis, and citation-based recognition indicators. This integrated design is consistent with data science perspectives that emphasize combining statistical, computational, and domain-specific reasoning to support prediction, exploration, understanding, and interpretation in modern scientific research (Blei & Smyth, 2017). To guide the analysis, the following research questions are proposed:

(i) Do the data indicate a temporal discontinuity in GenAI terminology after the late-2022 visibility marker?
(ii) Did the relative prevalence of publications captured by the high-specificity lexical rules change after the late-2022 visibility marker?
(iii) How are GenAI-classified publications distributed across OpenAlex disciplinary fields, and what semantic structure characterizes the associated keyword network?

(iv) How are GenAI classification and lexical rule-set membership associated with citation counts, and does the relative GenAI citation association differ before and after the late-2022 visibility marker?

With this approach, the study aims to make contributions in three areas: first, to provide empirical evidence on the trajectory of GenAI in academic communication; second, to illustrate how lexical classification, disciplinary diffusion, semantic organization, and citation-based recognition indicators can be analyzed in an integrated manner to assess rapidly developing fields; third, to propose a reproducible methodological framework for monitoring emerging fields by combining text analytics, temporal modeling, and network metrics.

The results suggest that GenAI terminology exhibited a temporal discontinuity after the late-2022 visibility marker, characterized by a discrete level change rather than a sustained acceleration in the previous trend. Additionally, the data show a post-intervention increase in high-specificity lexical classification, a broad but heterogeneous distribution across OpenAlex disciplinary fields, and a keyword network characterized by a prominent computational component and several overlapping applied, educational, social, health-related, and software-oriented thematic concentrations. Overall, these observations suggest that the emergence of GenAI has involved a process of lexical organization and disciplinary circulation within the scientific publication ecosystem, providing a basis for examining how rapidly developing technological domains become increasingly visible and organized in scholarly communication.

2. Related Literature

2.1. Academic Communication and the Emergence of Scientific Domains

Academic communication constitutes a central infrastructure for the production, circulation, preservation, and evaluation of knowledge. Beyond serving as a channel for transmitting results, the publishing system organizes how research objects are named, classified, and recognized within disciplinary communities. In this sense, emerging scientific domains do not seem to consolidate solely through the quantitative increase in publications, but rather through the stabilization of vocabularies, categories, citation networks, and editorial spaces that facilitate their identification as legitimate areas of research (Bawden, 2008; Mabe, 2010). This perspective is also consistent with critical approaches to academic knowledge production, which emphasize that scholarly communication, evaluation, and recognition are embedded in institutional structures that shape how knowledge circulates and gains legitimacy within academia (Demeter et al., 2023). This view is also consistent with innovation science perspectives suggesting that innovation is not only a matter of novelty creation but also of diffusion, classification, and the organization of dispersed vocabularies and knowledge structures across social and scientific systems (Cao et al., 2022).

From the perspective of the science of science, scientific production is often modeled as a dynamic system in which ideas, authors, publications, institutions, and recognition structures interact (Fortunato et al., 2018; Krauss, 2024). This line of research suggests that the evolution of science goes beyond the cumulative growth of articles, involving changes in the cognitive organization of knowledge, the formation of scientific communities, and the reconfiguration of interdisciplinary relationships (Fortunato et al., 2018).

In this context, bibliographic metadata such as titles, abstracts, and keywords provide empirical evidence to observe the configuration and transformation of research fields. Here, the language of publications can serve as an approximation of the cognitive structure of the research, allowing the identification of naming patterns, thematic concentration, and

terminological stabilization. Therefore, analyzing the vocabulary associated with GenAI offers insights into how this subject becomes visible in academic communication.

The emergence of GenAI represents a significant case study, as recent works document its rapid expansion within scientific production and its growing presence across multiple fields of knowledge (Ding et al., 2025; Khan et al., 2024; Alaqrabi & Alshagi, 2025). However, this increase in volume must be interpreted alongside indicators that assess how the domain is codified and gains recognition. From this perspective, GenAI offers an empirical opportunity to analyze how a highly visible technological object is incorporated into the publishing system and tends to be shaped as an identifiable research domain.

2.2. Semantic Coding and Disciplinary Diffusion

Scientific language plays a fundamental role in the formation of research fields, where processes of semantic change allow us to observe how certain terms acquire new uses, become specialized, or stabilize as references within an emerging domain (Hamilton et al., 2016; Buhari & Kumala, 2024). In scientific systems, these linguistic transitions often reflect transformations in the conceptual, institutional, and disciplinary organization of knowledge.

The literature on the evolution of scientific language suggests that academic writing tends to become more specialized, technical, and information-dense as fields become established (Sun et al., 2021; Dong et al., 2023). Complementarily, computational approaches based on co-occurrence and semantic analysis have expanded the capacity to study these processes at scale, facilitating the identification of conceptual associations and shifts in the use of technical terms (Rauch et al., 2024).

In the case of GenAI, semantic coding can be traced in the shift from broad expressions to more explicit terms such as generative artificial intelligence, large language models, ChatGPT, foundation models, and prompt engineering. The recurring presence of these terms may be compatible with increasing terminological concentration and possible stabilization, potentially reducing ambiguity in scholarly communication and contributing to the delimitation of the domain.

Disciplinary diffusion constitutes another key dimension in the formation of emerging fields. According to innovation diffusion models, new ideas or technologies spread through social systems in a heterogeneous and non-linear manner (Rogers, 2003). In science, this dispersion depends on the ability of concepts to integrate into pre-existing knowledge structures (Cheng et al., 2023), so the expansion of scientific language often shows differentiated circulation depending on the conceptual proximity of each discipline.

GenAI illustrates this pattern, since although its conceptual origin is linked to computer science, machine learning, and natural language processing, its presence in medicine, education, social sciences, business, and engineering suggests broad dissemination. However, this expansion may remain anchored in a computational core, indicating that broad disciplinary diffusion does not necessarily eliminate conceptual hierarchies within the field. Thus, the analysis of terms and co-occurrences allows for the simultaneous evaluation of both the expansion and the internal organization of the discourse surrounding GenAI.

2.3. Citation-Based Recognition and Legitimization in Scientific Publishing

Citations are among the most visible mechanisms of recognition in scientific publishing, although they do not fully capture the quality or validity of a contribution. They operate as indicators of circulation, use, and academic attention (Bornmann & Daniel, 2008). In bibliometric studies, citations are regularly used to analyze impact, influence, and the consolidation processes of scientific fields.

However, the literature points out limitations in the use of citation indicators, as citation distributions are asymmetric, tend to concentrate in a minority of publications, and are influenced by factors such as the prestige of the journals, the prior visibility of the authors, open access, metric optimization, and the cumulative dynamics of recognition (Bornmann & Daniel, 2008; D'Ippoliti, 2021; Fire & Guestrin, 2019; Larivière & Sugimoto, 2019). Therefore, citations are interpreted here as indicators of visibility within the system, rather than as absolute measures of quality. This caution is particularly relevant because citation-based and publication-based indicators may become strategic targets within academic evaluation systems, reducing their validity as proxies for scientific quality or contribution (Fire & Guestrin, 2019).

In emerging fields, citation visibility can exhibit distinct dynamics, with areas associated with high-profile technologies often experiencing an initial phase of concentrated attention, in which a subset of publications receive early recognition due to their novelty or topical relevance. This behavior is consistent with the Matthew effect, according to which initially visible contributions tend to accumulate advantages in attracting subsequent attention (Merton, 1968; Fortunato et al., 2018).

However, as the domain expands and becomes integrated into the literature, that initial advantage could diminish. The increase in the volume of articles tends to diversify attention, leading to a normalization of the average impact. In this sense, the legitimization of an emerging field seems to be expressed not only through high citation levels but also through a shift from concentrated visibility to a more routine integration into academic communication.

Against this background, analyzing citations in the GenAI literature makes it possible to examine how citation-based recognition varies across lexical classification categories and publication periods. Comparing publications captured by the high-specificity rules with those captured only by the expanded rules, as well as comparing GenAI-classified and classifier-negative retrieved records before and after the late-2022 visibility marker, helps assess whether citation associations differ according to semantic classification and temporal context. Citation metrics are therefore integrated with lexical, disciplinary, and semantic-network analyses to provide a broader account of the evolution of GenAI-related scholarly communication.

3. Materials and Methods

3.1. Corpus Construction and Data Collection

The corpus was built via programmatic queries to the OpenAlex Works API, covering publications from 2 August 2019 to 30 March 2026. This window was chosen deliberately to capture observations both before and after the public release of ChatGPT on 30 November 2022, which we treat as the study's principal temporal inflection point. OpenAlex was selected for this task given its broad disciplinary coverage, open accessibility, and demonstrated suitability for reproducible, large-scale bibliographic research (Culbert et al., 2025; Arumugam & Gunavathi, 2025).

The retrieval strategy combined the following search expressions using the Boolean operator OR: ChatGPT, generative AI, generative artificial intelligence, large language model, large language models, LLM, foundation model, and foundation models. The analytical sample was restricted to records classified by OpenAlex as research articles or review articles. Paratextual and retracted records were excluded during retrieval through the OpenAlex filters described in Appendix A, including is_paratext:false and is_retracted:false. Because the final analytical file did not retain a record-level retraction indicator, the retraction exclusion could not be independently re-audited at the analysis stage.

Records were retrieved through cursor-based pagination, capped at 100 records per request, with retry logic, rate-limit handling, and periodic checkpoints built into the collection script to accommodate the scale of the query. For each record, we extracted OpenAlex identifiers, DOIs, bibliographic metadata, citation counts, authorship and affiliation data, publication venue, topics, disciplinary fields, concepts, keywords, and abstracts, with the latter reconstructed from OpenAlex's inverted-index format. Duplicate records were removed first by OpenAlex identifier and, where available, cross-checked by DOI.

After application of the publication-type and date criteria, the final analytical corpus comprised 487,753 records: 465,135 research articles and 22,618 review articles. The corpus contained no duplicate OpenAlex identifiers, no duplicated non-empty DOIs, and no missing publication dates.

3.2. Lexical Classification of the Corpus

3.2.1. Text Representation and Normalization

For each record, we constructed a unified textual representation by integrating the available textual and semantic metadata, including the title, display name, abstract, keywords, concepts, topics, and disciplinary fields, following document-representation approaches that combine multiple textual features of scientific publications (Kozlowski et al., 2021). Missing fields were replaced with empty strings prior to concatenation. The resulting text was normalized by converting it to lowercase, removing diacritics and URLs, filtering unsupported special characters, and standardizing whitespace (Siino et al., 2024).

3.2.2. Rule-Based Classification

Lexical classification relied on predefined regular-expression dictionaries (Manning et al., 2008; Sebastiani, 2002) applied at two levels of specificity. The strict criterion prioritized explicit GenAI terminology, including ChatGPT, large language models, foundation models, and generative artificial intelligence. Records matching at least one strict pattern and no applicable exclusion rule were assigned to treatment_strict.

The expanded criterion incorporated broader technical terminology, including generative AI, LLMs, GPT model variants, transformers, retrieval-augmented generation (RAG), fine-tuning, instruction tuning, predictive prompting, and prompt engineering. To reduce overclassification, records not already assigned to treatment_strict were classified as treatment_broad only if they contained either two or more qualifying lexical matches or a selected unambiguous GenAI expression. Thus, treatment_broad represents expanded-only matches. Records meeting neither criterion were assigned to other, and the final GenAI-classified subset was defined as the union of treatment_strict and treatment_broad.

Legal-context exclusions were limited to ambiguous uses of "LLM" or "LLMs" when accompanied by legal terminology and when no unambiguous GenAI expression was present. This procedure excluded 26 records from positive assignment. The expanded-only category thus corresponds to (1):

$$\text{ExclusiveBroad}_i = \text{Broad}_i(1 - \text{Strict}_i) \quad (1)$$

The complete dictionaries, regular expressions, exclusion patterns, and assignment hierarchy are provided in Appendix A and in the project repository.

3.2.3. Validation of the Lexical Classifier

The lexical classification procedure was initially evaluated using a random, manually labeled subsample of 299 publications, each assigned a reference label indicating whether it substantively addressed generative artificial intelligence.

The high-specificity specification was used for the primary monthly count analysis because it applied the most restrictive lexical rules and reduced exposure to ambiguous matches. The expanded-only category was retained to evaluate whether records captured solely by the broader lexical rules exhibited different temporal and citation patterns.

Manual validation assessed substantive relevance to GenAI; it did not independently validate a latent construct of semantic explicitness. Accordingly, comparisons between the two classifications are interpreted as differences between lexical rule sets rather than as direct measurements of linguistic explicitness. Detailed sampling procedures, coding criteria, reference and predicted labels, and standard classification-performance metrics (Sokolova & Lapalme, 2009) are reported in Appendix A.

To strengthen the reference validation, an independent second coder evaluated a reproducible random subsample of 100 records from the original 299-record validation set, which had been manually labeled by a first coder. The second coder was blinded to the original manual labels and classifier predictions and applied the same binary criterion of substantive GenAI relevance. Inter-rater reliability was assessed using observed agreement and Cohen's κ, with a bootstrap 95% confidence interval based on 10,000 resamples. Disagreements were adjudicated only after the independent agreement analysis using the predefined operational criteria. Full procedures and adjudicated results are reported in Appendix A.9.

3.3. Analytical Preparation and Temporal Coding

Publication dates were standardized and aggregated by calendar month. Since ChatGPT was released on 30 November 2022, we treated November 2022 as the final pre-intervention month and December 2022 as the first complete post-intervention month, defining the monthly intervention indicator as (2):

$$\text{Post}_t = \begin{cases} 0, & t < \text{December 2022} \\ 1, & t \geq \text{December 2022} \end{cases} \quad (2)$$

$$\text{TimeAfter}_t = \begin{cases} 0, & t < \text{December 2022} \\ t - T_0, & t \geq \text{December 2022} \end{cases}$$

where T_0 denotes the temporal index corresponding to December 2022. Under this coding, $\text{TimeAfter}_t = 0$ in December 2022.

We constructed a continuous monthly time index beginning in August 2019. To obtain a directly interpretable level-change coefficient, we also defined a centered post-intervention time variable. This variable was set to zero throughout the pre-intervention period and in December 2022, the first complete post-intervention month, and then increased sequentially by one for each subsequent month. The resulting specification separates the underlying pre-intervention time trend, the discrete level difference at the start of the post-intervention period, and the change in slope thereafter. Before model estimation, we verified the absence of missing publication dates, duplicate records, and gaps in the 80-month sequence, which extended from August 2019 to March 2026.

3.3.1. Empirical Strategy

An interrupted time-series design was used to examine whether the frequency of GenAI-related publications exhibited changes in level or trend after the intervention (Bernal et al., 2021; Linden, 2015). The segmented model was specified as (3):

$$\log[E(Y_t)] = \beta_0 + \beta_1 \text{Time}_t + \beta_2 \text{Post}_t + \beta_3 \text{TimeAfter}_t \quad (3)$$

In this parameterization, β_0 represents the expected log count at the beginning of the observation period, β_1 represents the pre-intervention monthly trend, β_2 represents the discrete level difference in December 2022 relative to the counterfactual continuation of the pre-intervention trend, and β_3 represents the change in slope after the intervention. The post-intervention slope is therefore given by $\beta_1 + \beta_3$. Because the model uses a log link, exponentiated coefficients are interpreted as incidence rate ratios.

Because the outcome was a monthly count, a Poisson model with a log link was initially estimated. The Pearson dispersion statistic was approximately 649.76, indicating an extreme violation of the Poisson equidispersion assumption. Substantive inference was therefore based on a negative binomial NB2 model estimated by maximum likelihood, with its dispersion parameter estimated from the data. The Poisson specification was retained only as a diagnostic comparison. For the primary interrupted time-series specification, heteroskedasticity- and autocorrelation-consistent standard errors were calculated using a Newey–West covariance estimator with a maximum lag of three months. Model convergence, finite likelihood values, and dispersion estimates were examined before substantive interpretation. Additional model parameters and diagnostic procedures are reported in Appendix A.

3.3.2. Robustness and Model Design

Robustness was assessed through alternative distributional, temporal, and diagnostic specifications. First, the negative binomial model was compared with a Poisson specification, although the latter was not used for substantive inference because of extreme overdispersion. Second, the interrupted time-series model was re-estimated using September 2022 as an anticipatory cutoff and February 2023 as a delayed cutoff. In each specification, the post-intervention time variable was centered at the first post-cutoff month so that the level-change coefficient retained the same interpretation. Model convergence, finite parameter estimates, likelihood values, and dispersion estimates were examined for all alternative specifications. The February 2023 model did not yield a finite converged negative binomial solution and was therefore not interpreted substantively.

Because no credible unexposed bibliographic comparison series with a sufficiently similar pre-intervention trajectory could be identified, the study retained a single-series interrupted time-series design. Accordingly, the estimates are interpreted as temporal discontinuities associated with the selected visibility marker rather than as causal treatment effects.

Classification uncertainty was propagated through the ITS estimates using the two validation scenarios described in Section 3.2.3: the original 299-record validation sample and the adjudicated 100-record reference. Deterministic reweighting reconstructed expected monthly GenAI counts from the estimated probabilities of substantive GenAI relevance conditional on classifier status. In addition, a probabilistic sensitivity analysis generated 1000 Monte Carlo reconstructions of the monthly series, allowing these probabilities to vary according to beta distributions derived from the validation counts. The segmented NB2 model was re-estimated for each reconstruction. These analyses were interpreted as sensitivity scenarios under approximately nondifferential classification error rather than as estimates of latent true prevalence.

Finally, placebo interrupted time-series models were estimated using only the pre-intervention period (August 2019–November 2022). Four artificial cutoffs were evaluated: August 2020, February 2021, August 2021, and February 2022, using the same centered NB2 specification and Newey–West covariance estimator as in the primary model. This diagnostic assessed whether comparable level discontinuities emerged at arbitrary points within the pre-intervention series.

3.3.3. Analysis of Classification Composition

Classification composition was examined within the subset of publications assigned to either treatment_strict or treatment_broad. A binary indicator distinguished records captured by the high-specificity lexical rules from those captured exclusively by the expanded rules shown in (4):

$$\text{Strict}_i = \begin{cases} 1, & \text{if publication i was classified as treatment\_strict} \\ 0, & \text{if publication i was classified as treatment\_broad} \end{cases} \quad (4)$$

The overall probability of assignment to the high-specificity category was first estimated using an intercept-only logistic model as (5):

$$\text{logit}[P(\text{Strict}_i = 1)] = \gamma_0 \quad (5)$$

Temporal differences in classification composition were evaluated using a 2 × 2 contingency table, Pearson's chi-square test, Cramér's V, and a binary logistic regression as shown in (6):

$$\text{logit}[P(\text{Strict}_i = 1)] = \gamma_0 + \gamma_1 \text{Post}_i \quad (6)$$

where $\text{Post}_i = 1$ for publications dated from December 2022 onward and $\text{Post}_i = 0$ otherwise. The exponentiated coefficient e^{γ_1} was interpreted as the post- versus pre-intervention odds ratio for assignment to the high-specificity lexical category. Because manual validation assessed substantive relevance to GenAI rather than an independent construct of semantic explicitness, these analyses were interpreted as changes in classifier composition rather than direct measurements of linguistic specificity (Agresti, 2013).

3.3.4. Disciplinary Diffusion

Publications were mapped to the OpenAlex disciplinary fields associated with each record. Because OpenAlex permits multiple field assignments, the data were transformed into a publication–field structure containing one row for each observed assignment. Of the 487,753 publications in the analytical corpus, 487,438 had at least one disciplinary field assignment and 315 had no field information. The resulting dataset contained 865,123 publication–field assignments. Publications had between one and three assigned fields, with a mean of 1.77 and a median of two assignments. For each field f, the relative prevalence of GenAI classification was calculated as (7):

$$\text{GenAIShare}_f = \frac{N_{\text{GenAI,f}}}{N_{\text{Total,f}}} \quad (7)$$

where $N_{\text{GenAI,f}}$ denotes the number of publication–field assignments corresponding to records classified as either treatment_strict or treatment_broad, and $N_{\text{Total,f}}$ denotes all corpus assignments to field f. These proportions describe the relative presence of GenAI-classified records within each field and should not be interpreted as rates of technology adoption by researchers or institutions.

Because individual publications could contribute assignments to more than one field, publication–field observations were not treated as statistically independent. Field differences were therefore evaluated using a binomial regression in which GenAI classification was the binary outcome and disciplinary field was represented by a full set of field indicators. Cluster-robust standard errors were calculated at the publication level to account for dependence among assignments from the same publication. An omnibus cluster-robust Wald test assessed whether the predicted prevalence of GenAI classification differed across

fields. Raw field-specific proportions and cluster-robust confidence intervals were used for interpretation and visualization.

A conventional Pearson chi-square test was additionally calculated for descriptive comparison. Because its independence assumption is not satisfied by the overlapping publication–field structure, it was not used for primary inference.

3.3.5. Semantic Structure Analysis

The semantic structure analysis was conducted on OpenAlex keywords associated with the GenAI-classified subset. OpenAlex concepts were initially evaluated as a complementary source; however, audit procedures identified substantial overlap with keywords, broad disciplinary labels, ontology artifacts, and context-dependent expressions that could distort thematic interpretation. Keywords were therefore retained as the primary representation for the network analysis, while concepts were excluded from the principal analysis.

For the network analysis, keywords were separately normalized through lowercasing, whitespace standardization, and consolidation of equivalent term forms. Broad disciplinary labels, malformed expressions, confirmed ontology artifacts, and parenthetical disambiguation labels were excluded. Terms appearing in fewer than 100 GenAI-classified publications were also removed to limit the influence of highly infrequent expressions. After filtering, 915 eligible keywords remained, covering 137,476 of the 148,797 GenAI-classified publications; 11,321 publications had no keyword meeting the final eligibility criteria.

A binary publication–term representation was constructed so that each keyword contributed at most once per publication. Pairwise similarity between terms was calculated using cosine similarity based on document-occurrence profiles. The 50 highest-frequency eligible keywords were retained for the principal visualization. One isolated term in the initial top-50 set was replaced by the next-highest-frequency eligible keyword to obtain a connected network; this replacement was documented in the reproducibility files.

An undirected edge was included when cosine similarity was at least 0.05, with edge weights defined by cosine similarity rather than raw co-occurrence counts. Sensitivity analyses evaluated thresholds of 0.04, 0.05, 0.06, 0.075, and 0.10. The 0.05 threshold was retained because it yielded a connected and interpretable network without the substantially greater density observed at 0.04 or the fragmentation observed at higher thresholds.

Network structure was summarized using degree, weighted degree, betweenness centrality, density, connected components, and community structure. Louvain community detection was performed with resolution = 1 and random seed = 42. Community labels were assigned post hoc for descriptive interpretation and were not treated as externally validated thematic categories.

As a complementary analysis, hierarchical clustering was performed on the same 50-keyword set using cosine distance and average linkage with optimal leaf ordering. Average linkage was selected after comparison with complete and single linkage using cophenetic correlation. Agreement between the four-community Louvain partition and a four-cluster hierarchical solution was assessed using the adjusted Rand index and normalized mutual information. Hierarchical clustering was used to examine local proximity among terms rather than to validate or reproduce the Louvain communities.

3.4. Citation Indicators and Citation-Based Recognition

Citation counts were obtained from the OpenAlex cited_by_count field at the time of data extraction. Because citation distributions were highly right-skewed and included a substantial proportion of zero values, they were summarized using the mean, median, standard deviation, maximum, and proportion of uncited publications. Citation exposure was measured as elapsed months between publication and the data-extraction cutoff. This

continuous age measure was used in sensitivity specifications rather than publication year because records published within the same calendar year could differ substantially in citation opportunity.

Two negative binomial NB2 models with a log link were estimated, with the dispersion parameter estimated from the data. Primary specifications included publication-month fixed effects to control flexibly for publication timing and common month-specific citation exposure, with standard errors clustered by publication month. Sensitivity specifications modeled publication age using splines and heteroskedasticity-robust HC1 standard errors.

The first model was restricted to GenAI-classified publications and compared records assigned to treatment_strict with those assigned exclusively to treatment_broad. The first model was specified as Equation (8):

$$\log[\mathrm{E}(\mathrm{Citations}_i)] = \beta_0 + \beta_1 \mathrm{Strict}_i + \delta_{m(i)} \quad (8)$$

where $\mathrm{Strict}_i = 1$ for treatment_strict and $\mathrm{Strict}_i = 0$ for expanded-only treatment_broad, while $\delta_{m(i)}$ represents publication-month fixed effects. The exponentiated coefficient e^{β_1} was interpreted as the citation incidence rate ratio associated with high-specificity classification relative to expanded-only classification.

The second model used the full retrieved corpus and evaluated whether the citation difference between GenAI-classified and classifier-negative records changed after the late-2022 visibility marker, as shown in Equation (9):

$$\log(\mathbb{E}[\mathrm{Citations}_i]) = \beta_0 + \beta_1 \mathrm{GenAI}_i + \beta_2(\mathrm{GenAI}_i \times \mathrm{Post}_i) + \delta_{m(i)} \quad (9)$$

where $\mathrm{GenAI}_i = 1$ for records assigned to either treatment_strict or treatment_broad, $\mathrm{GenAI}_i = 0$ for classifier-negative retrieved records, and $\mathrm{Post}_i = 1$ for publications dated from December 2022 onward. Because publication-month fixed effects absorb the main effect of the post-intervention period, the standalone post-period coefficient was omitted from the primary specification. The interaction coefficient represents the relative change in the GenAI citation rate ratio after the visibility marker.

All citation models were interpreted as associational analyses of citation-based recognition and do not establish causal effects of lexical classification, GenAI relevance, or publication timing on citation counts.

Classification-error sensitivity was evaluated using the validation scenarios described in Section 3.2.3. Strict, expanded-only, and classifier-negative records were reweighted using category-specific probabilities of substantive GenAI relevance under the original-validation and adjudicated/hybrid scenarios. Because the independently coded 100-record subsample contained no expanded-only records, the corresponding probability in the hybrid scenario was retained from the original validation sample. These analyses were used to assess the stability of citation associations under plausible classification error.

3.5. Ethics Statement

This study used publicly available bibliographic metadata and did not involve human participants, animals, experimental interventions, private records, or the direct collection of sensitive personal information. The analyses were conducted at the publication and aggregated disciplinary levels. Institutional ethical approval and informed consent were therefore not required.

4. Results

4.1. Lexical Classification and Growth of Scholarly Communication on GenAI

Lexical classification was applied to 487,753 research and review articles retrieved from OpenAlex between 2 August 2019 and 30 March 2026. Of these, 338,956 records were assigned to the classifier-negative other category (69.49%), 139,272 to treatment_strict (28.55%), and 9525 exclusively to treatment_broad (1.95%). The final GenAI-classified subset therefore comprised 148,797 publications, representing 30.51% of the analytical corpus.

In the original manual validation, the strict specification yielded a precision of 0.689, a recall of 0.750, and an F1-score of 0.718, while the expanded specification yielded a precision of 0.671, a recall of 0.779, and an F1-score of 0.721.

Independent double coding of the 100-record subsample produced 79.0% observed agreement and Cohen's $\kappa = 0.529$ (bootstrap 95% CI: 0.367–0.681). Against the adjudicated reference, both specifications showed higher precision and lower recall than in the original validation. Detailed adjudicated results are reported in Appendix A.9.

Monthly counts of publications captured by the high-specificity GenAI rules showed an overall increase over the 80-month observation period. In the centered negative binomial specification, the pre-intervention trend was positive and statistically significant ($\beta = 0.0444$, $SE = 0.0141$, $p = 0.002$). Exponentiation of this coefficient yields an incidence rate ratio of 1.045, corresponding to an estimated 4.5% monthly increase before December 2022.

At the beginning of the post-intervention period, the model estimated a positive discrete level difference of $\beta = 2.8927$ ($SE = 0.4760$, $p < 0.001$). The corresponding incidence rate ratio was 18.04 (95% CI: 7.10–45.86), indicating that the fitted count for December 2022 was approximately 18 times the count projected under continuation of the pre-intervention trend. This estimate represents a modeled level discontinuity; its causal limitations are discussed in Section 5.1.

The estimated change in slope after the intervention was positive but not statistically significant ($\beta = 0.0252$, $SE = 0.0167$, $p = 0.131$; IRR = 1.026, 95% CI: 0.993–1.060). Thus, although the fitted post-intervention trajectory remained upward, the data did not provide sufficient evidence that its monthly growth rate differed from the pre-intervention rate. The estimated negative binomial dispersion parameter was $\alpha = 0.4441$, supporting the use of the negative binomial model over the Poisson alternative.

Table 1 summarizes the centered negative binomial interrupted time-series estimates.

Table 1. Centered negative binomial interrupted time-series model for monthly high-specificity GenAI publication counts.

Variable	β	SE	95% CI	*p*	IRR
Intercept	1.8383	0.2431	[1.3618, 2.3148]	<0.001	6.286
Time	0.0444	0.0141	[0.0168, 0.0719]	0.002	1.045
Post	2.8927	0.4760	[1.9598, 3.8256]	<0.001	18.041
Time after intervention	0.0252	0.0167	[−0.0075, 0.0579]	0.131	1.026
Dispersion (α)	0.4441	0.1276	[0.1940, 0.6942]	<0.001	—

The centered parameterization distinguishes the immediate modeled discontinuity from the subsequent change in trend. The results indicate a large positive level difference at the beginning of the post-intervention period, whereas the estimated slope change was not statistically distinguishable from zero. The observed pattern is therefore consistent with a discrete increase in the visibility of high-specificity GenAI terminology rather than evidence of an additional sustained acceleration in the monthly growth rate.

Figure 1 illustrates this pattern. Before November 2022, the series remains at relatively low levels; after the intervention, the series displays a pronounced upward level shift. The fitted trajectories do not indicate a statistically significant additional change in slope.

Figure 1. Observed and fitted monthly counts of publications captured by the high-specificity GenAI rules, August 2019–March 2026.

4.2. Temporal Change in GenAI Classification Composition

Among the 148,797 publications classified as GenAI-related, 139,272 (93.60%) were assigned to the high-specificity treatment_strict category, while 9525 (6.40%) were classified exclusively as expanded-only treatment_broad.

The proportion of strict classifications increased from 86.65% before December 2022 (675 of 779 publications) to 93.64% in the post-intervention period (138,597 of 148,018 publications). The association between period and classification category was statistically significant ($\chi^2(1) = 61.95$, $p < 0.001$), although its magnitude was small (Cramér's $V = 0.020$).

Consistent with this result, the logistic regression yielded a positive post-intervention coefficient, $\beta = 0.8183$ ($SE = 0.1059$, $z = 7.73$, $p < 0.001$). The corresponding odds ratio was 2.27 (95% CI: 1.84–2.79), indicating higher odds of assignment to the high-specificity category after the intervention. Table 2 summarizes the distribution of both classification categories across periods.

Table 2. Distribution of high-specificity and expanded-only GenAI classifications before and after the intervention.

Period	Expanded-Only, N	High-Specificity Strict, N	Total GenAI, N	Strict Classification, %
Pre-intervention	104	675	779	86.65
Post-intervention	9421	138,597	148,018	93.64
Total	9525	139,272	148,797	93.60

Overall, the findings suggest a post-intervention shift toward the high-specificity lexical category. This pattern is compatible with greater concentration around terms captured by the restrictive dictionary, including ChatGPT, large language models, generative artificial intelligence, and foundation models. However, because the classifier was not

designed to measure semantic stabilization directly, the result should be interpreted as a compositional lexical shift rather than as definitive evidence of field stabilization.

4.3. Disciplinary Distribution and Semantic Structure

Disciplinary field information was available for 487,438 of the 487,753 publications in the corpus. Because publications could receive up to three OpenAlex field assignments, these records generated 865,123 publication–field observations. A total of 193,934 publications had one field assignment, 209,323 had two, and 84,181 had three.

The relative prevalence of GenAI classification differed substantially across disciplinary fields. Computer Science showed the highest proportion at 49.42%, followed by Decision Sciences at 38.57%, Materials Science at 30.65%, Engineering at 27.65%, Social Sciences at 27.27%, Medicine at 27.11%, Psychology at 26.36%, and Physics and Astronomy at 25.76%.

The cluster-robust binomial model indicated that GenAI classification varied significantly across disciplinary fields, with an omnibus Wald statistic of $\chi^2(25) = 69,147.52$, $p < 0.001$. Standard errors were clustered by publication to account for the fact that the same record could contribute to multiple field assignments. The conventional assignment-level Pearson statistic was also large, $\chi^2(25) = 70,194.28$, $p < 0.001$, but it was treated as descriptive because repeated assignments violate the assumption of independent observations.

Figure 2 presents the observed percentage of publication–field assignments associated with GenAI-classified records across the 26 OpenAlex fields. These percentages describe the relative prevalence of GenAI classification within each field. Because disciplinary assignments overlap, they should not be interpreted as mutually exclusive percentages of publications or as rates of GenAI technology adoption by individual researchers or institutions.

These results suggest broad but heterogeneous cross-field distribution. GenAI-classified records appeared across all 26 OpenAlex fields, but their relative prevalence remained highest in Computer Science and Decision Sciences. The substantial shares observed in Materials Science, Engineering, Social Sciences, Medicine, Psychology, and Physics and Astronomy show that the literature extends beyond its computational origins. Nevertheless, because field assignments overlap and the analysis does not measure collaboration, cognitive integration, or reference diversity within individual publications, the findings should be interpreted as disciplinary diffusion rather than a direct measure of interdisciplinarity.

The semantic audit identified substantial differences between OpenAlex concepts and keywords. Concepts frequently included broad disciplinary categories, ontology artifacts, and context-dependent labels that were not sufficiently specific for thematic interpretation. The principal semantic analysis was therefore restricted to a curated keyword representation. Of the 148,797 GenAI-classified publications, 137,476 contained at least one eligible keyword after filtering, corresponding to 92.39% of the GenAI subset. The final vocabulary contained 915 eligible keywords and 488,561 publication–keyword pairs.

The final network comprised 50 keywords and 193 edges at a cosine-similarity threshold of 0.05. The network formed a single connected component, had a density of 0.158, and contained no isolated nodes. Louvain community detection identified four communities with a modularity of 0.353. Artificial intelligence was the most structurally central keyword, with a degree of 32, a weighted degree of 4.69, and a betweenness centrality of 0.441.

The first community contained 18 terms associated primarily with educational, cognitive, ethical, health-related, and social dimensions, including mathematics education, knowledge management, engineering ethics, cognitive science, social psychology, medical education, higher education, health care, qualitative research, and transformative learning. The second community contained 15 terms related to language-model architectures,

inference, model deployment, and computational workflows, including language model, transformer, deep learning, inference, workflow, automation, software deployment, data modeling, generalization, and interpretability.

The third community contained 14 terms centered on artificial intelligence, natural language processing, machine learning, data science, information retrieval, human–computer interaction, computer vision, data mining, and chatbot research. The fourth and smallest community contained programming language, software engineering, and software, representing a distinct software-development cluster. These labels are descriptive interpretations of the detected communities rather than externally validated taxonomic categories.

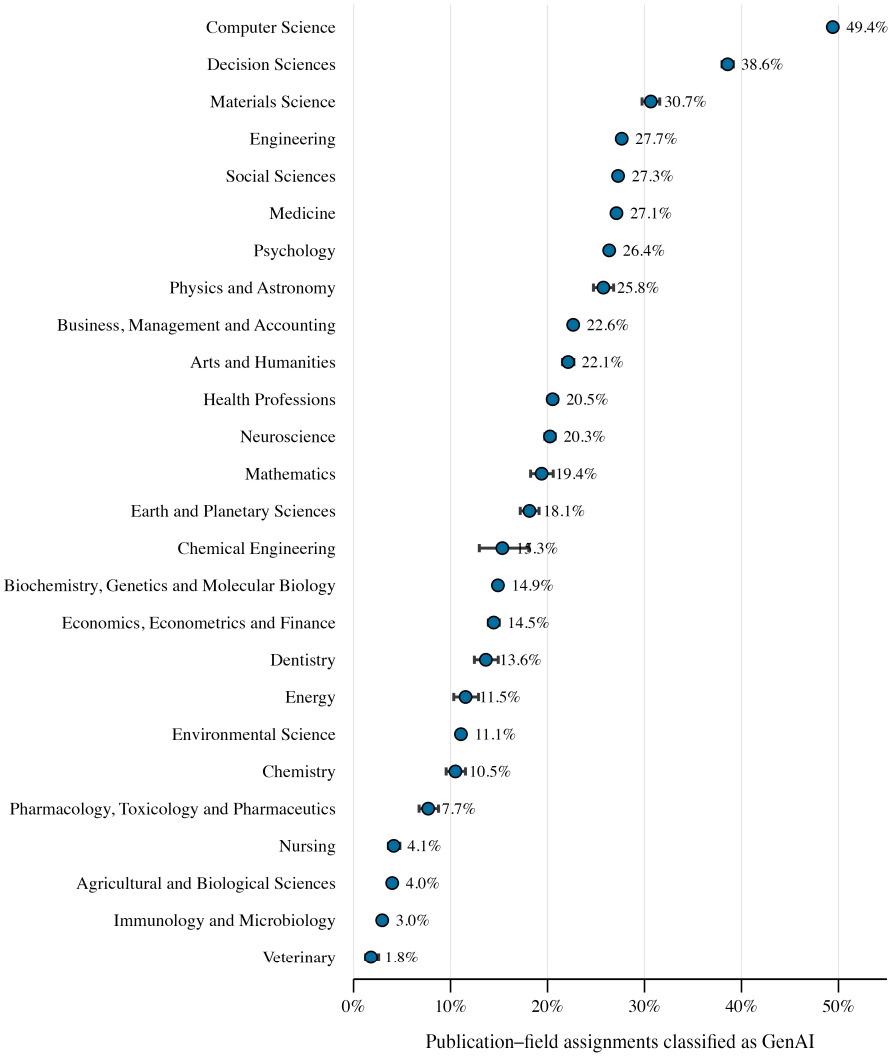

Figure 2. Relative prevalence of GenAI classification across OpenAlex disciplinary fields.

The complementary hierarchical analysis provided a reasonably faithful representation of the pairwise cosine distances, with a cophenetic correlation of 0.781 for average linkage. However, silhouette coefficients were low across the evaluated cluster solutions, and agreement between the four-cluster hierarchical partition and the four Louvain communities was moderate rather than high (adjusted Rand index = 0.378; normalized mutual information = 0.530). The dendrogram therefore supports the presence of local semantic proximities but does not indicate sharply separated hierarchical clusters or independently confirm the Louvain partition. Accordingly, the four-community solution is interpreted as a descriptive partition of overlapping thematic concentrations rather than as evidence of four sharply delimited semantic subfields.

Figure 3 presents the hierarchical clustering structure of the 50 highest-frequency eligible OpenAlex keywords associated with GenAI-classified publications. The figure is organized into four panels corresponding to the Louvain communities identified in the semantic network: (A) social, educational, and epistemic themes; (B) AI methods and information technologies; (C) software development; and (D) language models, LLMs, and deployment-related themes. Within each panel, the dendrogram represents the relative semantic proximity among terms based on cosine distance.

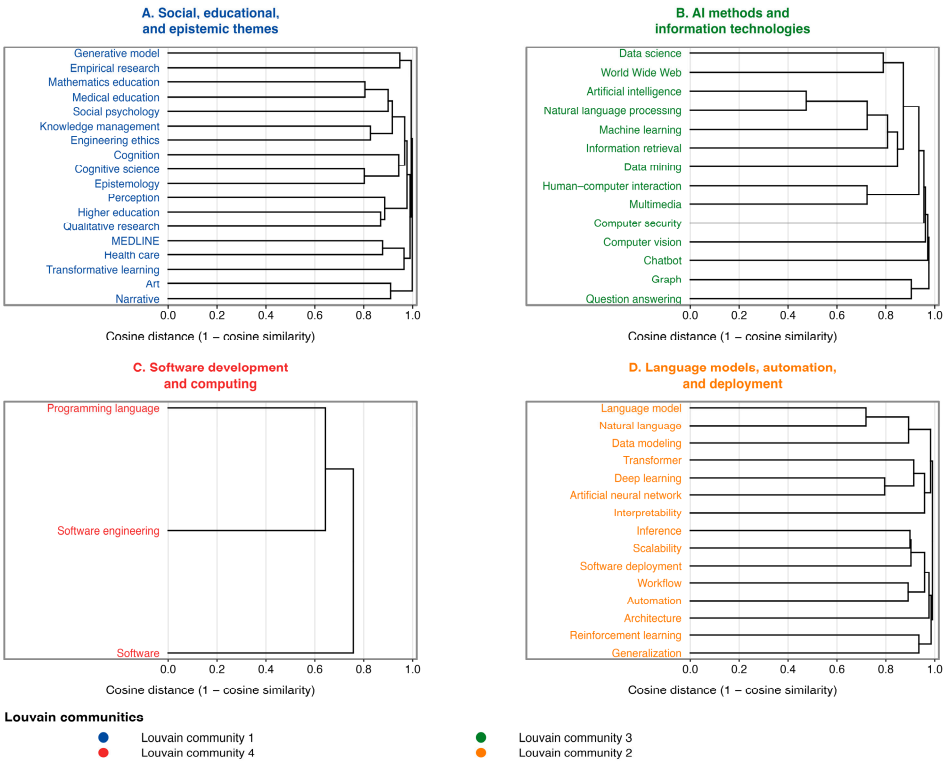

Figure 3. Hierarchical clustering of the 50 highest-frequency eligible OpenAlex keywords by Louvain community.

4.4. Citation Patterns and Citation-Based Recognition

Citation counts were strongly right-skewed across the corpus. Among the 487,753 retrieved records, 305,972 publications had received no citations at the extraction date, corresponding to 62.73% of the corpus. The overall mean was 3.89 citations, whereas the variance was 1292.10, producing a variance-to-mean ratio of 331.76 and supporting the use of an overdispersed count model.

Within the GenAI-classified subset, publications assigned to treatment_strict had a mean of 6.07 citations, compared with 2.88 citations for records assigned exclusively to treatment_broad. The proportion of uncited records was 60.87% in the strict category and 70.39% in the expanded-only category. These unadjusted differences should be interpreted cautiously because citation opportunity varied substantially across publication months. Table 3 summarizes the descriptive citation statistics by GenAI classification category.

Table 3. Descriptive citation statistics by GenAI classification category.

Classification Category	N	Mean Citations	Median	Zero-Citation Records, %
Expanded-only treatment_broad	9525	2.88	0	70.39
High-specificity treatment_strict	139,272	6.07	0	60.87

In the primary month-fixed-effects negative binomial model, high-specificity classification was positively associated with citation counts relative to expanded-only classification. The coefficient for treatment_strict was $\beta = 0.3218$, with a publication-month-clustered standard error of 0.0662 ($p < 0.001$). The corresponding incidence rate ratio was 1.38 (95% CI: 1.21–1.57), indicating that strict-classified publications had approximately 38% higher expected citation counts than expanded-only publications after controlling flexibly for publication month.

A sensitivity model using a spline representation of publication age and HC1 robust standard errors produced a similar estimate, $\beta = 0.3447$, IRR = 1.41 (95% CI: 1.29–1.54), supporting the substantive stability of the association across timing-control strategies. Table 4 reports the negative binomial model for citation counts.

Table 4. Negative binomial model comparing high-specificity and expanded-only GenAI classifications.

Variable	β	Clustered SE	95% CI for β	*p*	IRR	95% CI for IRR
High-specificity classification	0.3218	0.0662	[0.1900, 0.4537]	<0.001	1.38	[1.21, 1.57]

Across the full retrieved corpus, classifier-negative records had a mean of 3.03 citations, whereas GenAI-classified records had a higher unadjusted mean. However, these descriptive comparisons are strongly influenced by publication timing. Earlier publications had substantially greater citation exposure, and the pre- and post-intervention groups differed markedly in age. For this reason, interpretation focused on the month-adjusted negative binomial interaction model rather than on raw period means.

In the full-corpus month-fixed-effects negative binomial model, GenAI classification was positively associated with citation counts in the pre-intervention period ($\beta = 0.7038$, clustered $SE = 0.1413$, $p < 0.001$; IRR = 2.02). The interaction between GenAI classification and the post-intervention period was negative and statistically significant ($\beta = -0.3602$, clustered $SE = 0.1526$, $p = 0.018$; IRR = 0.70, 95% CI: 0.52–0.94). Table 5 reports the estimates from this GenAI-by-period interaction model.

Table 5. Negative binomial model of citation counts with GenAI-by-period interaction.

Variable	β	Clustered SE	95% CI for β	*p*	IRR	95% CI for IRR
GenAI classification	0.7038	0.1413	[0.4268, 0.9807]	<0.001	2.02	[1.53, 2.67]
GenAI × Post	−0.3602	0.1526	[−0.6592, −0.0611]	0.018	0.70	[0.52, 0.94]

The combined post-intervention GenAI coefficient was $\beta = 0.3436$, corresponding to an incidence rate ratio of approximately 1.41. Thus, GenAI-classified publications retained a citation advantage after December 2022, but the magnitude of that advantage was smaller than in the pre-intervention period. This pattern is compatible with attenuation of an early citation advantage as GenAI-related publishing expanded, although the associational design does not identify the mechanisms responsible for the change.

The age-spline sensitivity model yielded a similar qualitative pattern: the GenAI main coefficient was positive, the interaction was negative, and both were statistically significant. This consistency indicates that the attenuation finding was not dependent on a single timing-control specification.

4.5. Robustness Checks

For the primary December 2022 specification, the pre-intervention trend was positive ($\beta = 0.0444$, $p = 0.002$), the estimated level difference was positive and statistically significant ($\beta = 2.8927$, $p < 0.001$), and the change in slope was not statistically significant

($\beta = 0.0252$, $p = 0.131$). The dispersion estimate was $\alpha = 0.4441$, with a log-likelihood of −492.84 and an AIC of 995.68.

The September 2022 anticipatory specification also produced a positive pre-intervention trend ($\beta = 0.0467$, $p = 0.010$) and a positive level difference ($\beta = 2.4478$, $p < 0.001$; IRR = 11.56). The estimated change in slope remained non-significant ($\beta = 0.0341$, $p = 0.139$). This model had a dispersion estimate of $\alpha = 0.5761$, a log-likelihood of −503.88, and an AIC of 1017.76.

The February 2023 specification did not yield a finite converged negative binomial solution and was therefore not interpreted. Among the estimable specifications, the December 2022 model showed the lower AIC and higher log-likelihood. These results support December 2022 as the preferred prespecified visibility marker among the converged alternatives, without implying that it represents a unique or causal structural breakpoint.

The pre-intervention placebo analysis did not identify a comparable positive level discontinuity at any of the four artificial cutoffs. Estimated placebo level IRRs were 0.92 (95% CI: 0.41–2.04; $p = 0.836$) for August 2020, 0.64 (0.20–2.03; $p = 0.449$) for February 2021, 1.77 (0.38–8.18; $p = 0.465$) for August 2021, and 0.34 (0.11–1.02; $p = 0.055$) for February 2022. Thus, none reproduced the large positive level discontinuity estimated for December 2022 (IRR = 18.04).

Table 6 summarizes the robustness checks across alternative intervention dates. Among the converged models, the December 2022 specification provided the best relative fit and retained the clearest substantive interpretation as the first complete month following the public release of ChatGPT.

Table 6. Sensitivity of the centered negative binomial interrupted time-series model to alternative cutoff dates.

First Post-Cutoff Month	β Time	*p*	β Post	*p*	IRR Post	β TimeAfter	*p*	α	Log-Likelihood	AIC	Converged
September 2022	0.0467	0.010	2.4478	<0.001	11.56	0.0341	0.139	0.5761	−503.88	1017.76	Yes
December 2022	0.0444	0.002	2.8927	<0.001	18.04	0.0252	0.131	0.4441	−492.84	995.68	Yes
February 2023	—	—	—	—	—	—	—	—	—	—	No

Taken together, the alternative-cutoff results support December 2022 as the preferred prespecified first post-intervention month among the converged specifications, while also suggesting that the observed discontinuity may reflect a broader late-2022 transition rather than a uniquely identifiable breakpoint.

Classification-error sensitivity analyses preserved the principal late-2022 level discontinuity. Deterministic reweighting reduced the post-period IRR from 18.04 in the observed series to 10.80 (95% CI: 4.97–23.47) under the original-validation scenario and 7.32 (95% CI: 3.50–15.29) under the adjudicated scenario; both remained statistically significant. Monte Carlo analyses were closely aligned, yielding median post IRRs of 10.81 and 7.25, respectively, with the level-shift IRR exceeding 1 in all converged simulations. In contrast, the estimated post-period slope change became positive and statistically significant in the reweighted scenarios, indicating that inference about the subsequent trajectory is more sensitive to classification assumptions than the estimated late-2022 level discontinuity.

Citation results were also stable in direction under classification-error reweighting. The strict-versus-expanded-only IRR was 1.380 in the observed model and 1.378 in both reweighted scenarios. In the full-corpus model, the positive GenAI citation association and the negative GenAI-by-post interaction remained statistically significant under both the original-validation and adjudicated/hybrid scenarios. The precise magnitudes differed across specifications, but the direction of the citation associations was unchanged.

5. Discussion

5.1. From Publication Growth to a Visibility Discontinuity in Scholarly Communication

The interrupted time-series analysis identified a large positive level discontinuity in late 2022 against an already-increasing pre-intervention trajectory. In the primary specification, the subsequent slope change was not statistically significant, whereas classification-error sensitivity analyses preserved a substantial positive level shift but yielded greater sensitivity in the estimated post-intervention slope. The evidence is therefore more robust for a discrete late-2022 discontinuity than for the exact form of the subsequent growth trajectory.

This discontinuity should not be interpreted as the origin of GenAI research or as a causal effect attributable exclusively to the public release of ChatGPT. The single-series design identifies a temporal departure from the projected pre-intervention trajectory, while concurrent developments in language-model research, media attention, editorial priorities, terminology, and bibliographic indexing may also have contributed. The pattern is therefore best interpreted as a change in scholarly visibility, naming, retrieval, and bibliographic encoding of an already-developing research domain. This interpretation is consistent with science-of-science perspectives emphasizing that emerging domains evolve through changes not only in publication volume but also in categorization, institutional visibility, and the cognitive organization of knowledge (Fortunato et al., 2018; Krauss, 2024).

5.2. Changes in Lexical Classification and Terminological Concentration

The proportion of GenAI-classified publications assigned to the high-specificity lexical category increased from 86.65% before December 2022 to 93.64% afterward. Although the association was statistically significant, Cramér's V = 0.020 indicates a small effect. Moreover, the strict and expanded-only categories represent operational lexical rule sets rather than independently validated stages of linguistic standardization or field maturity.

The observed shift is therefore best interpreted as greater lexical concentration around recognizable expressions such as ChatGPT, large language models, generative artificial intelligence, and foundation models. Such concentration may facilitate retrieval, indexing, and classification and is compatible with greater terminological convergence, consistent with research showing that emerging scientific areas often develop increasingly specialized and standardized vocabularies as they become more institutionally visible (Hamilton et al., 2016; Sun et al., 2021; Rauch et al., 2024). However, while the present evidence is compatible with greater terminological concentration, it does not demonstrate semantic stabilization or full consolidation of GenAI as an autonomous scientific field.

5.3. Disciplinary Diffusion and a Prominent Computational Component

The disciplinary analysis indicates that GenAI-related scholarly communication has achieved broad but heterogeneous cross-field distribution. Computer Science showed the highest relative prevalence, followed by Decision Sciences, with substantial shares also observed in Materials Science, Engineering, Social Sciences, Medicine, Psychology, and Physics and Astronomy. Because OpenAlex field assignments overlap, these findings describe bibliographic distribution rather than interdisciplinarity as a multidimensional property involving collaboration, cognitive integration, or reference diversity.

The semantic network complements this disciplinary pattern. Artificial intelligence occupied the most central position, while natural language processing, machine learning, data science, information retrieval, human–computer interaction, and related computational terms formed a highly connected thematic region. Language-model architectures, inference, deployment, automation, and model-development workflows constituted another technical concentration. Together, these patterns suggest that computational terms occupy a prominent and highly connected position within the keyword network, while the

moderate agreement between clustering solutions cautions against treating this structure as a uniquely defined thematic partition.

Applied educational, cognitive, ethical, social, and health-related terms were also represented, suggesting diffusion into socially situated domains while maintaining links to technical foundations. This configuration is consistent with networked accounts of innovation in which emerging technologies diffuse through asymmetric relationships between technically central and adjacent applied domains (Acemoglu et al., 2016), and with diffusion-of-innovation perspectives emphasizing uneven circulation across pre-existing disciplinary structures (Rogers, 2003; Cheng et al., 2023; Takahashi et al., 2024). GenAI can therefore be interpreted as a broadly circulating research object with a prominent computational component, although the network evidence does not support sharply bounded or uniquely identifiable thematic subfields.

The hierarchical analysis reinforces this caution: cluster separation was weak, and concordance with the Louvain partition was only moderate. The network and dendrogram therefore capture related but non-equivalent dimensions of semantic organization, supporting an interpretation of overlapping thematic concentrations rather than rigidly bounded subfields.

5.4. Citation-Based Recognition and Attenuation of an Early Citation Association

The citation analysis indicates that publications captured by the high-specificity lexical rules had higher expected citation counts than records captured only by the expanded rules after controlling for publication month. In the full-corpus interaction model, GenAI-classified publications also showed a stronger positive citation association before the late-2022 visibility marker, while the negative interaction indicated that this relative association became smaller afterward, although the combined post-intervention association remained positive.

This pattern is compatible with attenuation of an early visibility advantage as the GenAI literature expanded and academic attention became distributed across a larger number of publications. Such a transition is consistent with cumulative-advantage perspectives, according to which initially visible contributions may attract disproportionate recognition before a field expands and attention becomes more diffuse (Merton, 1968; Fortunato et al., 2018).

These associations should nevertheless be interpreted cautiously. Citation accumulation is influenced by publication age, journal visibility, author reputation, open-access status, field-specific citation cultures, strategic behavior, topical centrality, and other unobserved characteristics (Bornmann & Daniel, 2008; D'Ippoliti, 2021; Fire & Guestrin, 2019; Larivière & Sugimoto, 2019). The findings therefore represent changing citation associations rather than causal effects of lexical wording or definitive evidence of field normalization.

Classification-error sensitivity analyses reinforced the directional stability of these findings: the strict-versus-expanded-only association was virtually unchanged after reweighting, and the positive GenAI association together with the negative post-period interaction remained statistically significant. However, the magnitude of the full-corpus coefficients varied across error scenarios and should therefore not be interpreted as a uniquely identified effect size.

5.5. Robustness and Interpretation of Late 2022 as a Visibility Marker

The alternative-cutoff analysis supports interpretation of a broader late-2022 transition rather than a uniquely identifiable breakpoint. Both the anticipatory September 2022 specification and the primary December 2022 specification produced positive and statistically significant level discontinuities, whereas the February 2023 model did not yield a finite

converged solution. December 2022 remains the preferred prespecified cutoff because it corresponds to the first complete month following ChatGPT's public release and provided better relative fit than the September alternative among the converged specifications.

The similarity between the September and December models cautions against attributing the observed discontinuity to a single date or mechanism. Concurrent technological, editorial, media, terminology, and indexing developments may also have contributed. Publication lags may also blur the precise timing of the observed discontinuity, because articles published in late 2022 may reflect research, review, and editorial processes initiated months earlier. Accordingly, November 2022 is treated as a substantively motivated marker of increased public visibility rather than as the sole causal origin of the observed change.

Pre-intervention placebo tests provided additional support for the temporal distinctiveness of the late-2022 level change: none of the four artificial pre-2022 cutoffs produced a statistically significant positive level discontinuity comparable to the primary estimate. However, this diagnostic cannot separate changes in the underlying literature from contemporaneous changes in OpenAlex indexing, metadata coverage, or terminology standardization.

5.6. Implications for Research on Scientific Publishing and Emerging Domains

Taken together, the findings suggest that the emergence of GenAI in scholarly communication can be examined through several complementary dimensions: temporal visibility, lexical classification and concentration, disciplinary distribution, thematic organization, and citation associations. Rather than relying only on publication volume or citation counts, combining interrupted time-series analysis, lexical classification, disciplinary mapping, co-occurrence networks, and citation modeling provides a broader framework for studying rapidly developing technological domains.

From this perspective, academic communication can be understood not simply as a channel for reporting scientific advances but also as a system in which emerging research objects are named, classified, connected across disciplinary contexts, retrieved through bibliographic infrastructures, and differentially recognized through citation practices. The present findings are therefore best interpreted as evidence of ongoing discursive and bibliographic organization around GenAI rather than as definitive evidence of semantic stabilization, interdisciplinarity, or full field consolidation.

6. Conclusions

Based on 487,753 research and review articles retrieved from OpenAlex, this study identified four main empirical patterns.

First, the interrupted time-series analysis showed a large late-2022 level discontinuity in the visibility of GenAI terminology that remained robust to classification-error sensitivity analyses, although inference regarding the subsequent slope was more sensitive to classification assumptions.

Second, the share of GenAI-classified publications assigned to the high-specificity lexical category increased from approximately 87% to 94%, indicating greater lexical concentration rather than direct evidence of semantic stabilization.

Third, GenAI-classified publications were distributed across all 26 OpenAlex disciplinary fields, although their relative prevalence remained highest in Computer Science and Decision Sciences. Because field assignments overlapped, these findings represent broad but heterogeneous disciplinary diffusion rather than a direct measurement of interdisciplinarity. The keyword network revealed overlapping thematic concentrations around computational, language-model, applied, and software-related topics. Artificial intelligence occupied a central network position, but only moderate agreement between Louvain and

hierarchical clustering indicates that these concentrations should not be interpreted as sharply separated subfields.

Fourth, month-adjusted citation models showed positive associations for high-specificity and GenAI classifications, with a smaller relative GenAI citation association after the late-2022 visibility marker. Overall, the study provides a reproducible framework for examining the emergence of rapidly developing technological domains through complementary temporal, lexical, disciplinary, semantic-network, and citation dimensions. The findings characterize an ongoing process of organization within scholarly communication while remaining cautious about causal disruption, full terminological stabilization, interdisciplinarity, and field consolidation.

Limitations and Future Work

This study has several limitations. First, the corpus was defined through a specific OpenAlex retrieval strategy and is therefore conditioned by database coverage, indexing practices, metadata quality, and search behavior. Records excluded by the initial query could not enter the classifier, and classifier-negative retrieved records should not be interpreted as representative of all non-GenAI scholarship. Although the pre-intervention placebo analysis did not reproduce the late-2022 positive discontinuity at alternative artificial cutoffs, this does not rule out contemporaneous changes in OpenAlex indexing or metadata-generation practices. Record-level indexing lag could not be examined because the analytical dataset retained publication dates but not OpenAlex record-creation or update timestamps.

Second, the lexical classifier depended on predefined lexical dictionaries and exclusion patterns. Independent double coding showed moderate inter-rater agreement ($\kappa = 0.529$), indicating that substantive GenAI relevance remains partly judgment-dependent in borderline records. Although adjudication provided a consensus reference for the independently coded subsample, classification uncertainty remains, particularly for records not captured by the positive lexical rules. The strict and expanded-only categories should therefore be interpreted as operational lexical classifications rather than definitive ontological categories. The classification-error sensitivity analyses also assume that validation-derived error probabilities are approximately transportable across publication periods and analytical strata.

Third, the interrupted time-series analysis used a single bibliographic series without a credible unexposed comparison group. The late-2022 cutoff was substantively motivated, but contemporaneous technological, editorial, institutional, media, terminology, and indexing developments may also have contributed. The ITS estimates therefore identify temporal discontinuities rather than causal effects attributable to ChatGPT or any single event.

Fourth, OpenAlex permits multiple disciplinary field assignments per publication. Although dependence was addressed statistically through publication-clustered standard errors, the field analysis measures the relative distribution of GenAI-classified records across overlapping bibliographic categories rather than interdisciplinarity through collaboration, cognitive integration, reference diversity, or related multidimensional constructs.

Fifth, the semantic-network analysis depends on the availability and quality of OpenAlex keywords and on several analytical decisions. The 50-keyword network represents an interpretable high-frequency subset of the 915 eligible terms rather than a complete thematic map; one isolated top-frequency term was replaced by the next eligible keyword to obtain a connected visualization; and Louvain community labels were assigned post hoc. The resulting communities should therefore be interpreted as exploratory thematic concentrations rather than externally validated taxonomies.

Sixth, citation counts were measured at a single extraction date and remain affected by unequal accumulation periods and unobserved determinants such as journal visi-

bility, open-access status, author reputation, collaboration structure, field-specific citation practices, and topical centrality. The independently coded subsample contained no expanded-only records; consequently, the adjudicated citation sensitivity scenario retained the expanded-only probability estimated from the original validation sample and should be interpreted as a hybrid scenario.

Future research could extend the classifier through multilingual, supervised, or hybrid approaches; use controlled or synthetic-comparison time-series designs; examine interdisciplinarity through co-authorship, institutional affiliation, reference diversity, and cross-field citation networks; and model citation accumulation longitudinally. Qualitative and mixed-method studies could further examine how researchers, journals, reviewers, and institutions adopt and stabilize terminology around emerging technologies.

Author Contributions: Conceptualization, C.H.S.-R. and A.M.G.-G.; methodology, C.H.S.-R. and E.L.-A.; software, C.H.S.-R.; validation, C.H.S.-R., A.M.G.-G. and E.L.-A.; formal analysis, C.H.S.-R. and E.L.-A.; investigation, C.H.S.-R., A.M.G.-G. and E.L.-A.; resources, C.H.S.-R. and A.M.G.-G.; data curation, C.H.S.-R. and E.L.-A.; writing—original draft preparation, C.H.S.-R.; writing—review and editing, C.H.S.-R., A.M.G.-G. and E.L.-A.; visualization, C.H.S.-R. and E.L.-A.; supervision, A.M.G.-G.; project administration, C.H.S.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The bibliographic records analyzed in this study were retrieved from the OpenAlex Works API. The data-retrieval and preprocessing scripts, lexical-classification dictionaries, exclusion rules, manual-validation materials, analysis code, model outputs, and documentation required to reproduce the reported results are available in the GitHub repository genai-scholarly-communication-analysis (https://github.com/suarezrodcarlos/genai-scholarly-communication-analysis, accessed on 1 September 2026). Because the study uses dynamically updated OpenAlex metadata, exact citation counts and some metadata fields may differ in later retrievals. The repository includes the extraction date, query parameters, and derived analytical outputs used in the present study.

Acknowledgments: During the preparation of this manuscript, the authors used Claude Sonnet 4.5 (Anthropic) to assist with code debugging and code documentation, Grammarly (version 1.192.0.0) and DeepL (version 26.6.14916780) to assist with language editing. All AI-assisted and language-editing outputs were reviewed, verified, and revised by the authors. These tools did not determine the study design, semantic classifications, statistical estimates, interpretation of the results, or conclusions. The authors take full responsibility for the content, accuracy, and integrity of the publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Appendix A.1. OpenAlex API Configuration

The complete OpenAlex request configuration used for corpus retrieval is shown below. The query covered 2 August 2019–30 March 2026 and was restricted to research and review articles, excluding paratextual and retracted records.

Cursor-based pagination retrieved up to 100 records per request. Automatic retries handled connection errors and HTTP 429 responses, and records and cursor states were checkpointed every 25 pages.

```
search_text = (
    '"chatgpt" OR '
    '"generative ai" OR '
    '"generative artificial intelligence" OR '
    '"large language model" OR '
    '"large language models" OR '
    'llm OR '
    '"foundation model" OR '
    '"foundation models"'
)
params = {
    "filter": (
        "from_publication_date:2019-08-02,"
        "to_publication_date:2026-03-30,"
        "type:article | review,"
        "is_paratext:false,"
        "is_retracted:false"
    ),
    "search": search_text,
    "per-page": 100,
    "cursor": cursor,
    "mailto": CONTACT_EMAIL
}
```

Appendix A.2. Variables Retrieved from OpenAlex

The following variables were extracted when available and are summarized in Table A1.

Table A1. Variables retrieved from the OpenAlex Works API.

Dimension	Variables
Identification	OpenAlex ID, DOI, PMID, and PMCID
Bibliographic metadata	Title, display name, publication date, publication year, document type, and language
Citation and access	cited_by_count, open-access status, and open-access category
Publication venue	Journal or source, source type, publisher, and host organization
Authorship	Author names, institutional affiliations, and countries
Semantic metadata	Topics, fields, domains, concepts, and keywords
Textual proxy	Abstract reconstructed from the OpenAlex inverted index
Access links	Landing-page URL and PDF URL

Abstracts represented as inverted indexes were reconstructed by ordering each token according to its positional indices.

Appendix A.3. Deduplication Procedure

Records were first deduplicated by OpenAlex ID. Among records with a normalized DOI, one record per DOI was retained. Records without a DOI remained identifiable through their OpenAlex ID. The implemented procedure was as follows:

```
df = df.drop_duplicates(subset=["openalex_id"])
with_doi = df[df["doi"].notna()].drop_duplicates(subset=["doi"])
without_doi = df[df["doi"].isna()]
df = pd.concat([with_doi, without_doi], ignore_index=True)
```

Appendix A.4. Text Representation and Normalization

The unified text field concatenated title, display name, abstract, keywords, concepts, topics, and disciplinary fields. Missing values were replaced with empty strings. The normalization function was as follows:

```
def normalize_text(text):
    text = strip_accents(text).lower()
    text = re.sub(r"http\S+|www\S+", " ", text)
    text = re.sub(r"[^a-z0-9\s\-\./]", " ", text)
    text = re.sub(r"\s+", " ", text).strip()
    return text
```

Appendix A.5. Strict GenAI Regular-Expression Dictionary

The exact regular expressions used in the strict specification were:

```
GENAI_STRICT_PATTERNS = [
    r"\bchatgpt\b",
    r"\blarge language model\b",
    r"\blarge language models\b",
    r"\bfoundation model\b",
    r"\bfoundation models\b",
    r"\bgenerative artificial intelligence\b"
]
is_genai_strict = (
    (genai_hits_strict >= 1)
    & (~has_exclusion)
)
```

Appendix A.6. Expanded GenAI Regular-Expression Dictionary

The expanded specification incorporated all strict expressions and the following additional technical terms:

```
GENAI_EXPANDED_ADDITIONAL_PATTERNS = [
    r"\bgenerative ai\b",
    r"\bllm\b",
    r"\bllms\b",
    r"\bgpt[- ]?3(\.5)?\b",
    r"\bgpt[- ]?4o?\b",
    r"\btransformer\b",
    r"\btransformers\b",
    r"\bretrieval[- ]augmented generation\b",
    r"\brag\b",
    r"\bfine[- ]tuning\b",
    r"\binstruction tuning\b",
    r"\bpredictive prompting\b",
    r"\bprompt engineering\b"
]
```

To reduce overclassification, expanded membership required at least two lexical signals, unless the record explicitly mentioned ChatGPT or a large language model. GPT-4 and GPT-4o were represented through a unified regular expression to prevent double counting.

```
is_genai_broad = (
    (
     (genai_hits_broad >= 2)
     | text_full.str.contains(r"\bchatgpt\b", regex=True)
     | text_full.str.contains(r"\blarge language models?\b", regex=True)
    )
    & (~has_exclusion)
)
```

Appendix A.7. Exclusion Rules

The following expressions were used to reduce false positives caused by legal uses of the acronym LLM:

```
EXCLUSION_PATTERNS = [
    r"\bmaster of laws\b",
    r"\blegal studies\b"
]
```

Records triggering an exclusion expression were not assigned to either GenAI category.

Appendix A.8. Exclusive Semantic Labels

Because the expanded indicator incorporated the strict expressions, the two binary indicators were not mutually exclusive. Exclusive labels were assigned hierarchically:

```
def assign_semantic_label(row):
    if row["is_genai_strict"]:
        return "treatment_strict"
    elif row["is_genai_broad"]:
        return "treatment_broad"
    else:
        return "other"
```

Appendix A.9. Manual Validation Protocol

The original 299-record validation set was coded by one researcher. To assess the reproducibility of the human reference judgments, a reproducible random subsample of 100 records was subsequently coded independently by a second researcher who was blinded to both the original manual labels and classifier predictions. Both coders applied the same binary operational criterion: whether the publication substantively addressed generative artificial intelligence. Independent coding yielded 79.0% observed agreement and Cohen's $\kappa = 0.529$ (bootstrap 95% CI: 0.367–0.681). The 21 disagreements were adjudicated only after calculation of inter-rater agreement, using the predefined coding criteria. Sixteen disagreements were resolved as GenAI-related and five as not GenAI-related.

Table A2. Classifier performance against the adjudicated 100-record reference.

Classifier	TP	FP	FN	TN	Precision	Recall	Specificity	F1
Strict	17	4	19	60	0.810	0.472	0.938	0.597
Expanded	17	3	19	61	0.850	0.472	0.953	0.607

Metric Calculation

Classifier performance was evaluated separately for the strict and expanded specifications. For each specification, the manually assigned reference label was compared with the corresponding algorithmic prediction. True positives (TP) were records correctly classified as GenAI-related; false positives (FP) were records classified as GenAI-related by the algorithm but not by the manual reference coding; false negatives (FN) were manually confirmed GenAI-related records not detected by the classifier; and true negatives (TN) were records correctly identified as not GenAI-related.

Precision, recall, and the F1-score were calculated from the confusion-matrix components following standard classification-performance definitions (Sokolova & Lapalme, 2009), as shown in Equations (A1)–(A3):

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{A1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{A2}$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{A3}$$

The manually coded reference set contained 68 GenAI-related publications and 231 publications not substantively related to GenAI. Tables A3 and A4 present the confusion matrices for the strict and expanded classifiers, respectively.

Table A3. Confusion matrix for the strict classifier.

Manual Reference	Predicted GenAI-Related	Predicted Not GenAI-Related	Total
GenAI-related	TP = 51	FN = 17	68
Not GenAI-related	FP = 23	TN = 208	231
Total	74	225	299

Table A4. Confusion matrix for the expanded classifier.

Manual Reference	Predicted GenAI-Related	Predicted Not GenAI-Related	Total
GenAI-related	TP = 53	FN = 15	68
Not GenAI-related	FP = 26	TN = 205	231
Total	79	220	299

The strict classifier correctly identified 51 of the 68 manually confirmed GenAI-related publications and produced 23 false-positive classifications. The expanded classifier detected 53 of the 68 positive records, reducing the number of false negatives from 17 to 15, but increased false positives from 23 to 26. This pattern reflects the expected trade-off between the slightly higher precision of the strict specification and the higher recall of the expanded specification.

Appendix A.10. Monthly Intervention Coding

The monthly intervention variables were constructed as follows:

```
monthly = monthly.sort_values("month_start").reset_index(drop=True)
monthly["time"] = np.arange(len(monthly))
first_post_month = pd.Timestamp("2022-12-01")
first_post_index = monthly.index[monthly["month_start"] == first_post_month][0]
monthly["post"] = (
    monthly["month_start"] >= first_post_month
).astype(int)
monthly["time_after"] = np.where(
     monthly["post"].eq(1),
     monthly["time"] - first_post_index,
     0
)
```

The centered time_after variable equals zero throughout the pre-intervention period and in December 2022 and increases by one in each subsequent month. This parameterization allows the post coefficient to represent the modeled level difference at the beginning of the post-intervention period.

Appendix A.11. Interrupted Time-Series Estimation and Diagnostics

The monthly series contained 80 consecutive observations from August 2019 to March 2026. The Poisson model exhibited extreme overdispersion, with a Pearson dispersion statistic of approximately 649.76. Substantive inference was therefore based on a negative binomial NB2 model estimated by maximum likelihood. The centered primary specification produced a dispersion estimate of $\alpha = 0.4441$, a log-likelihood of -492.84, and an AIC of 995.68. Heteroskedasticity- and autocorrelation-consistent standard errors were calculated using a Newey–West covariance estimator with a maximum lag of three months and the Bartlett kernel. Full coefficient estimates are reported in Table 1.

Appendix A.12. Robustness Specifications

Robustness checks re-estimated the centered negative binomial ITS using September 2022 as an anticipatory cutoff and February 2023 as a delayed cutoff. The September specification converged and produced a positive level coefficient, whereas the February specification did not yield a finite converged negative binomial solution. Alternative models were evaluated using coefficient stability, convergence, finite likelihood and dispersion estimates, log-likelihood, and AIC. Results are reported in Table 6 of the main text.

Appendix A.13. Disciplinary Field Analysis

A publication assigned to multiple OpenAlex fields contributed one observation to each publication–field combination. Of the 487,753 publications, 487,438 had at least one field assignment, generating 865,123 publication–field observations. Field-specific GenAI prevalence was calculated as the proportion of assignments associated with records classified as either treatment_strict or treatment_broad.

Because publication–field observations were not independent, field differences were evaluated using a binomial regression with field indicators and standard errors clustered by publication. The omnibus cluster-robust Wald statistic was used as the principal inferential test. A conventional Pearson chi-square statistic was retained only as a descriptive comparison. Field-specific observed proportions and cluster-robust 95% confidence intervals were used in Figure 2.

Appendix A.14. Curated Keyword Network and Hierarchical Clustering

The principal semantic analysis used OpenAlex keywords associated with records classified as treatment_strict or treatment_broad. OpenAlex concepts were audited but excluded from the primary analysis because they contained broad disciplinary labels, ontology artifacts, contextual ambiguities, and substantial overlap with keywords.

Keywords were normalized and deduplicated within publications. Broad disciplinary labels, malformed expressions, confirmed artifacts, parenthetical disambiguation labels, and terms appearing in fewer than 100 GenAI-classified publications were excluded. The final eligible vocabulary contained 915 keywords, and 137,476 GenAI-classified publications contained at least one eligible keyword.

Binary publication–keyword occurrence profiles were used to calculate cosine similarity between terms. The principal network included 50 high-frequency eligible keywords. The isolated term benchmarking in the initial top-50 set was replaced by the next eligible term, data modeling, to obtain a connected visualization; this decision was recorded in the replacement log.

An undirected edge was retained when cosine similarity was at least 0.05. Sensitivity analyses evaluated thresholds of 0.04, 0.05, 0.06, 0.075, and 0.10. The final network contained 50 nodes, 193 edges, one connected component, no isolated nodes, and a density of 0.1576. Louvain community detection used resolution 1 and random seed 42, yielding four communities and a modularity of 0.3535.

Hierarchical clustering used the same 50-keyword set, cosine distance defined as $1-$ cosine similarity, average linkage, and optimal leaf ordering. Average linkage produced a cophenetic correlation of 0.781. Agreement with the four-community Louvain solution was assessed using the adjusted Rand index and normalized mutual information, which were 0.378 and 0.530, respectively. The dendrogram was interpreted as a complementary representation of local semantic proximity rather than validation of four sharply separated clusters.

Appendix A.15. Supplementary Robustness Material

Table A5 summarizes the sensitivity of the principal ITS and citation estimates across the observed, original-validation, and adjudicated/hybrid scenarios. Results are reported as incidence rate ratios (IRRs) with 95% confidence intervals.

Table A5. Classification-error sensitivity of principal ITS and citation estimates.

Outcome/Model	Observed	Original Validation	Adjudicated/Hybrid
ITS post IRR	18.04 (7.10–45.86)	10.80 (4.97–23.47)	7.32 (3.50–15.29)
ITS time-after IRR	1.026 (0.993–1.060)	1.048 (1.019–1.077)	1.056 (1.028–1.084)
Strict vs. expanded-only citation IRR	1.380 (1.212–1.571)	1.378 (1.211–1.569)	1.378 (1.210–1.568)
GenAI × post citation IRR	0.698 (0.517–0.941)	0.518 (0.312–0.858)	0.483 (0.276–0.844)

References

Acemoglu, D., Akcigit, U., & Kerr, W. R. (2016). Innovation network. *Proceedings of the National Academy of Sciences*, *113*(41), 11483–11488. [CrossRef]

Agresti, A. (2013). *Categorical data analysis* (3rd ed.). Wiley.

Alaqrabi, M., & Alshagi, N. (2025). Navigating the evolving landscape of generative AI: Reality and challenges. *Academy Journal for Basic and Applied Sciences*, *7*(2), 1–6. [CrossRef]

Arumugam, J., & Gunavathi, M. (2025). OpenAlex and the future of libraries: Enhancing research discoverability, access, and innovation. *Journal of Advances in Library and Information Science*, *14*(4), 283–287. [CrossRef]

Bawden, D. (2008). Scholarship in the digital age: Information, infrastructure and the Internet. *Journal of Documentation*, *64*(4), 630–631. [CrossRef]

Bernal, J. L., Cummins, S., & Gasparrini, A. (2021). Corrigendum to: Interrupted time series regression for the evaluation of public health interventions: A tutorial. *International Journal of Epidemiology*, *50*(4), 1045. [CrossRef]

Blei, D. M., & Smyth, P. (2017). Science and data science. *Proceedings of the National Academy of Sciences*, *114*(33), 8689–8692. [CrossRef]

Bornmann, L., & Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*(1), 45–80. [CrossRef]

Buhari, B., & Kumala, S. A. (2024). Semantic change in historical linguistics: Theories, evidence, and contexts. *Lingua: Journal of Linguistics and Language*, *2*(2), 102–115. [CrossRef]

Cao, L., Chen, Z., & Evans, J. A. (2022). Destructive creation, creative destruction, and the paradox of innovation science. *Sociology Compass*, *16*(11), e13043. [CrossRef]

Cheng, M., Smith, D. S., Ren, X., Cao, H., Smith, S., & McFarland, D. A. (2023). How new ideas diffuse in science. *American Sociological Review*, *88*(3), 522–561. [CrossRef]

Culbert, J. H., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P., & Mayr, P. (2025). Reference coverage analysis of OpenAlex compared to Web of Science and Scopus. *Scientometrics*, *130*, 2475–2492. [CrossRef]

Demeter, M., Háló, G., & Rajkó, A. (2023). The capital-labor problem in academic knowledge production. *RAE-IC, Revista de la Asociación Española de Investigación de la Comunicación*, *10*(20), raeic102001. [CrossRef]

Ding, L., Lawson, C., & Shapira, P. (2025). Rise of generative artificial intelligence in science. *Scientometrics*, *130*, 5093–5114. [CrossRef]

D'Ippoliti, C. (2021). Many-citedness: Citations measure more than just scientific quality. *Journal of Economic Surveys*, *35*(5), 1271–1301. [CrossRef]

Dong, S., Mao, J., & Pei, L. (2023). Comparing the writing styles of multiple disciplines: A large-scale quantitative analysis. *Proceedings of the Association for Information Science and Technology*, *60*(1), 941–943. [CrossRef]

Fire, M., & Guestrin, C. (2019). Over-optimization of academic publishing metrics: Observing Goodhart's law in action. *GigaScience*, *8*(6), giz053. [CrossRef]

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., & Barabási, A.-L. (2018). Science of science. *Science*, *359*(6379), eaao0185. [CrossRef]

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th annual meeting of the association for computational linguistics* (Volume 1, pp. 1489–1501). Association for Computational Linguistics. [CrossRef]

Hou, J., Zheng, B., Li, H., & Li, W. (2025). Evolution and impact of the science of science: From theoretical analysis to digital-AI driven research. *Humanities and Social Sciences Communications*, *12*, 316. [CrossRef]

Khan, N., Khan, Z., Koubaa, A., Khan, M. K., & bin Salleh, R. (2024). Global insights and the impact of generative AI-ChatGPT on multidisciplinary: A systematic review and bibliometric analysis. *Connection* Science, *36*(1), 2353630. [CrossRef]

Kozlowski, D., Dusdal, J., Pang, J., & Zilian, A. (2021). Semantic and relational spaces in science of science: Deep learning models for article vectorisation. *Scientometrics*, *126*(7), 5881–5910. [CrossRef]

Krauss, A. (2024). Science of science: A multidisciplinary field studying science. *Heliyon*, *10*(8), e36066. [CrossRef]

Larivière, V., & Sugimoto, C. R. (2019). The journal impact factor: A brief history, critique, and discussion of adverse effects. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 3–24). Springer. [CrossRef]

Linden, A. (2015). Conducting interrupted time-series analysis for single- and multiple-group comparisons. *The Stata Journal*, *15*(2), 480–500. [CrossRef]

Mabe, M. A. (2010). Scholarly communication: A long view. *Chandos Information Professional Series*, *4*(3), 132–144. [CrossRef]

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.

Merton, R. K. (1968). The Matthew effect in science. *Science*, *159*(3810), 56–63. [CrossRef]

Quintans-Júnior, L. J., Gurgel, R. Q., Araújo, A. A. de S., Correia, D., & Martins-Filho, P. R. S. (2023). ChatGPT: The new panacea. *Revista da Sociedade Brasileira de Medicina Tropical*, *56*, e0060-2023. [CrossRef]

Rauch, C. B., Choi, H. W., & Kelly, M. (2024). Beyond the 'Idols of the Marketplace': Managing semantic change in research. *Proceedings from the Document Academy*, *11*(2), 3. [CrossRef]

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, *34*(1), 1–47. [CrossRef]

Siino, M., Tinnirello, I., & La Cascia, M. (2024). Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers. *Information Systems*, *121*, 102342. [CrossRef]

Sokolova, M., & Lapalme, G. (2009). Performance measures for classification tasks. *Information Processing & Management*, *45*(4), 427–437. [CrossRef]

Sun, K., Liu, H., & Xiong, W. (2021). The evolutionary pattern of language in scientific writings: A case study of *Philosophical Transactions of Royal Society* (1665–1869). *Scientometrics*, *126*, 1695–1724. [CrossRef]

Takahashi, C. K., de Figueiredo, J. C. B., & Scornavacca, E. (2024). Investigating the diffusion of innovation: A comprehensive study of successive diffusion processes through analysis of search trends, patent records, and academic publications. *Technological Forecasting and Social Change*, *198*, 122991. [CrossRef]