Black Sea Journal of Artificial Intelligence
doi: 10.5281/zenodo.18213819

Open Access Journal
e-ISSN: 3108-4281

Research Article
Volume 1 - Issue 2: 40-52 / December 2025

PROMPT ENGINEERING: A BIBLIOMETRIC ANALYSIS

Alperen KOÇYİĞİT[1]*, Nurettin ŞENYER[1]

[1]*Samsun University, Faculty of Engineering, Department of Software Engineering, 55060, Samsun, Türkiye*

Abstract: Prompt engineering addresses the design and optimization of natural language instructions for controlling large language model behavior, representing a critical domain within artificial intelligence and human-computer interaction. Despite explosive research growth following widespread adoption of large language models, systematic analyses of the field's intellectual structure, publication patterns, and collaborative networks remain limited. This study conducted a comprehensive bibliometric analysis of prompt engineering research from 2020 to 2025 using Web of Science Core Collection as the data source. A systematic search strategy retrieved 4,890 publications from 1,538 sources authored by 16,052 researchers across computer science and artificial intelligence domains. Analysis employed the Bibliometrix package (version 4.3.5) in R (version 4.5.1) to examine publication trends, author productivity, institutional contributions, geographic distribution, thematic structure, collaboration patterns, and citation impact through performance analysis, keyword co-occurrence networks, and science mapping techniques. The field demonstrated explosive expansion with 125.1% annual growth rate, exhibiting three developmental phases: emergence phase (2020-2021), acceleration phase (2022-2023), and explosion phase (2024-2025) when output reached 2,093 publications annually. The Chinese Academy of Sciences led institutional productivity with 345 publications, while China dominated national output with 1,584 documents representing 32.67% of the corpus. Geographic analysis revealed quality-quantity trade-offs with Singapore achieving the highest average citation impact (37.37 citations per document) despite modest volume. Author analysis identified Zhang Y as most productive (43 publications) while collaboration metrics indicated 4.9 co-authors per document and 26.44% international co-authorship rate. Keyword analysis revealed "large language models" (946 occurrences) and "prompt engineering" (733 occurrences) as dominant themes with three distinct thematic clusters: core prompting methodologies, machine learning foundations, and application domains. Network visualization confirmed integration of few-shot learning, chain-of-thought prompting, and in-context learning techniques into large language model applications. IEEE Access dominated publication venues with 176 articles, while natural language processing conferences (ACL, EMNLP, NeurIPS) emerged as primary dissemination channels. Citation analysis identified foundational contributions in instruction following and chain-of-thought reasoning alongside contemporary methodological innovations. The findings reveal prompt engineering's rapid crystallization as a distinct research domain emphasizing practical techniques over theoretical foundations, while concentration around specific models indicates potential fragmentation risks requiring unified frameworks transcending particular implementations.

Keywords: In-context learning, Natural language processing, Publication trends, Research collaboration, Thematic analysis

***Corresponding author:** Samsun University, Faculty of Engineering, Department of Software Engineering, 55060, Samsun, Türkiye
E mail: muh.alperenkocyigit@gmail.com (A. KOÇYİĞİT)
Alperen KOÇYİĞİT https://orcid.org/0009-0004-2781-3543
Nurettin ŞENYER https://orcid.org/0000-0002-2324-9285

Received: November 02, 2025
Accepted: December 05, 2025
Published: December 15, 2025
Cite as: Koçyiğit, A., & Şenyer, N. (2025). Prompt engineering: A bibliometric analysis. *Black Sea Journal of Artificial Intelligence*, *1*(2), 40–52.

1. Introduction

The rapid evolution of artificial intelligence has transformed how we interact with computational systems, with large language models emerging as powerful tools capable of understanding and generating human-like text. These sophisticated neural networks, trained on vast corpora of textual data, have demonstrated remarkable versatility across diverse natural language processing tasks. At the heart of effectively harnessing these models lies a critical skill: prompt engineering—the art and science of crafting precise instructions to guide model behavior and optimize output quality.

Prompt engineering represents a fundamental paradigm shift in human-computer interaction, where natural language becomes the primary interface for controlling and directing artificial intelligence systems. The field's significance stems from a well-documented phenomenon: large language models exhibit substantial sensitivity to the specific wording, structure, and context provided in prompts, with seemingly minor modifications often yielding dramatically different results (Sasson Lazovsky et al., 2024). Research has demonstrated that effective prompt design can substantially enhance model performance, while poorly constructed prompts may produce irrelevant or nonsensical responses (Sasson Lazovsky et al., 2024). The emergence of prompt engineering as a distinct research domain parallels the rapid advancement of transformer-based language models, particularly following the introduction of GPT-3 in 2020 (Brown et al., 2020) and the subsequent release of ChatGPT in November 2022 (Vaira et al., 2023).

These developments catalyzed widespread public and academic interest in large language models, transforming them from specialized research tools into ubiquitous technologies accessible to millions of users. This democratization of access has simultaneously amplified both the potential impact and the critical need for systematic understanding of effective prompting strategies. Contemporary prompt engineering encompasses several sophisticated techniques that extend beyond simple instruction-giving. Zero-shot prompting provides task descriptions without examples, relying on the model's pre-existing knowledge (Jovanovic and Voss, 2025). Few-shot learning incorporates demonstration examples within prompts to guide model behavior, with research showing that even the ordering of these examples can significantly influence final outputs (Cheong, 2025). Chain-of-thought prompting breaks complex problems into intermediate reasoning steps, enabling models to tackle multi-step logical challenges more effectively (Wang et al., 2025). These diverse approaches reflect the field's growing methodological sophistication and its recognition that different tasks may require fundamentally different prompting strategies. Despite the proliferation of prompt engineering applications across domains ranging from software development (Nguyen-Duc et al., 2025) to medical diagnostics (van Diessen et al., 2024), the field's rapid growth has outpaced comprehensive scholarly analysis. While individual studies have explored specific prompting techniques or evaluated model performance on particular tasks, the broader landscape of prompt engineering research remains incompletely mapped. Understanding the field's intellectual structure, key contributors, dominant themes, and evolutionary trajectory requires systematic bibliometric investigation.

This study addresses this gap through a comprehensive bibliometric analysis of prompt engineering literature spanning 2020 to 2025. By analyzing 4,890 publications indexed in Web of Science, we examine the field's exponential growth trajectory, identify leading contributors and institutions, map collaborative networks, and trace the evolution of research themes. Our analysis reveals an annual growth rate exceeding 125%, with publication volumes increasing from 30 papers in 2020 to over 2,000 in 2024, reflecting the field's explosive expansion. This quantitative assessment provides researchers, practitioners, and policymakers with empirical insights into prompt engineering's development as a scholarly discipline and its position within the broader artificial intelligence landscape. The following sections present our methodology for data collection and analysis, results encompassing publication trends, authorship patterns, geographic distribution, and thematic evolution, and discussion of implications for future research directions in this rapidly evolving field.

2. Materials and Methods

2.1. Research Design

This study employed a quantitative bibliometric analysis approach to systematically examine the scholarly landscape of prompt engineering research. Bibliometric analysis represents an established methodology for quantitative assessment of scientific literature through statistical and mathematical techniques (Aria and Cuccurullo, 2017). The research design encompasses data collection from a bibliographic database, application of bibliometric indicators, and visualization of research patterns through network analysis and performance metrics.

2.2. Data Source

Web of Science (WoS) Core Collection served as the sole data source for this analysis. WoS was selected based on its comprehensive coverage of high-impact publications in computer science and artificial intelligence disciplines, rigorous quality control through selective journal indexing, and robust citation indexing capabilities enabling citation-based analyses (Pranckutė, 2021). The use of a single database ensures methodological consistency and data compatibility across all bibliometric indicators.

2.3. Search Strategy and Query Development

A comprehensive search query was developed through iterative refinement to capture the multifaceted nature of prompt engineering research while maintaining domain specificity. The search was executed on November 27, 2025, using the following query structure:

((TS=("prompt engineering" OR "prompt design" OR "prompt optimization" OR ("prompt" AND ("tuning" OR "learning" OR "construction") AND ("AI" OR "LLM" OR "large language model*" OR "NLP" OR "machine learning")) OR "in-context learning" OR "few-shot prompting" OR "zero-shot prompting" OR "chain of thought prompting" OR "LLM prompting" OR "GPT prompting" OR "natural language prompting" OR "prompt-based learning" OR ("prompting" AND ("technique*" OR "method*") AND ("AI" OR "NLP")))) AND PY=(2020-2025)) AND WC=(Computer Science OR Computer Science, Artificial Intelligence)

The search terms were selected to encompass core dimensions of prompt engineering. The query includes primary field terms such as "prompt engineering", "prompt design", and "prompt optimization" to capture direct references to the field. Technical process terms including "prompt tuning", "prompt learning", and "prompt construction" represent key methodological approaches within the domain. Paradigm-level terms including "in-context learning", "few-shot prompting", "zero-shot prompting", and "chain of thought prompting" capture fundamental techniques and methodologies. System-level terms including "LLM prompting", "GPT prompting", and "natural language prompting" represent implementation perspectives. The wildcard operator (*) was applied to "large language model*", "technique*", and "method*" to retrieve morphological variations.

Boolean operators AND and OR were strategically employed to combine related concepts while maintaining query precision and avoiding retrieval of irrelevant publications. The temporal scope was restricted to 2020-2025 to focus on the contemporary emergence and rapid development of prompt engineering following the introduction of GPT-3 and subsequent large language models. The WoS category restriction to Computer Science and Computer Science, Artificial Intelligence ensures domain relevance while capturing interdisciplinary contributions at the intersection of natural language processing, machine learning, and human-computer interaction.

2.4. Inclusion and Exclusion Criteria

Publications were included if they met the following criteria: publication type classified as Article, Proceedings Paper, or Review Article as these represent peer-reviewed research contributions; publication year between 2020 and 2025 inclusive; indexed in Web of Science Core Collection with complete bibliographic metadata; and classification within Computer Science or Computer Science, Artificial Intelligence WoS categories. Publications were excluded if they met any of the following criteria: document type classified as Editorial Material, Letter, Book Chapter, Book Review, Correction, or Meeting Abstract as these do not represent original peer-reviewed research contributions; identification as duplicate records through DOI and title matching to avoid distortion of frequency-based metrics; or incomplete essential bibliographic fields compromising analysis reliability. No duplicate records were detected in the final dataset. The search query yielded 4,890 publications meeting all inclusion criteria.

2.5. Data Collection and Processing

Bibliographic data were exported from WoS on November 27, 2025, in plain text format. The export included complete records containing bibliographic information such as title, authors, source, publication year, volume, issue, pagination, and DOI; author information including names, affiliations, and corresponding author address; abstract and keywords comprising author keywords and Keywords Plus; citation data including times cited and citation count per year; and complete reference lists for each publication. Exported data underwent preprocessing to ensure analytical quality through four automated procedures. Author name normalization disambiguated author name variants through algorithmic matching based on affiliation and co-author patterns. Institutional affiliation standardization harmonized institutional name variants using the Bibliometrix institution cleaning algorithm. Keyword harmonization standardized British and American spelling variants, performed lowercase conversion, and removed punctuation inconsistencies. Duplicate detection identified and removed duplicate records based on DOI and title matching, though no duplicates were found in the final dataset. All preprocessing utilized automated Bibliometrix functions to ensure reproducibility without manual intervention.

2.6. Analytical Framework and Tools

Analysis was conducted using a Bibliometric Analysis and Paper Creator Agent system developed as part of the first author's Master's thesis. The system operates on R programming language version 4.5.1 (R Core Team, 2025) with Bibliometrix package version 4.3.5 (Aria and Cuccurullo, 2017) and supporting packages including dplyr version 1.1.4 for data manipulation (Wickham et al., 2023), ggplot2 version 3.5.0 for visualization (Wickham, 2016), and xtable version 1.8-4 for LaTeX table generation (Dahl et al., 2019). All analyses were executed in R environment ensuring reproducibility through documented code and version-controlled packages. The analysis applied five bibliometric techniques. Performance analysis evaluated productivity and impact of research actors (authors, institutions, countries, sources) through publication counts, citation metrics, and h-index values. Science mapping techniques visualized intellectual structure through keyword co-occurrence networks revealing thematic clusters and their interrelationships. Co-citation analysis identified intellectual linkages between publications frequently cited together, indicating shared conceptual foundations. Bibliographic coupling analysis grouped publications sharing common references, revealing thematic communities. Collaboration network analysis examined co-authorship patterns at author, institution, and country levels, quantifying international collaboration rates and identifying key collaborative actors. Network visualizations employed Fruchterman-Reingold force-directed layout algorithm with node size representing frequency, edge thickness representing co-occurrence strength, and color coding representing community clusters identified through modularity optimization. Keyword co-occurrence networks were constructed using author keywords and Keywords Plus terms with minimum frequency threshold of 5 occurrences to filter noise while preserving meaningful patterns. Citation-based metrics include total citations, average citations per document, and normalized citation scores accounting for publication age and field citation density. All metrics follow standard bibliometric definitions as established in Aria and Cuccurullo (2017).

2.7. Visualization

Network visualizations were generated to illustrate thematic and collaboration structures. Keyword co-occurrence networks employed node size to represent keyword frequency, edge thickness to represent co-occurrence strength, and force-directed layout algorithm for node positioning. Word clouds provided visual representation of keyword frequency through font size scaling. Temporal plots depicted publication trends over time through line graphs. All visualizations maintain minimum 600 dpi resolution for publication quality with network visualizations employing Fruchterman-Reingold layout algorithm for optimal node positioning.

2.8. Study Limitations

This study acknowledges six methodological limitations. The exclusive use of WoS Core Collection excludes publications indexed only in Scopus, IEEE Xplore, ACL Anthology, or other databases, prioritizing methodological consistency over absolute comprehensiveness. The English-only restriction may underrepresent research contributions from non-Anglophone regions, particularly Chinese research published in domestic journals. The temporal scope beginning in 2020 excludes earlier foundational works in natural language processing and machine learning that preceded the modern prompt engineering era, though these works' influence is acknowledged qualitatively. The WoS category restriction may exclude relevant interdisciplinary work published outside selected categories, particularly publications in computational linguistics, human-computer interaction, or domain-specific application venues. Despite comprehensive term selection, the search query may not capture all terminology variations used across different research communities, conferences, and time periods, particularly newly emerging terms in this rapidly evolving field. Recently published papers from 2024-2025 have limited citation accumulation time, potentially underestimating their eventual impact compared to earlier publications. These limitations are inherent to bibliometric studies and do not compromise the validity of findings within the defined scope.

3. Results

3.1. Dataset Overview and Descriptive Statistics

The search query retrieved 4,890 publications from Web of Science Core Collection spanning 2020 to 2025. Table 1 presents the main descriptive statistics of the dataset. The corpus comprised publications from 1,538 sources authored by 16,052 researchers. The average document age was 0.939 years with an average of 10.1 citations per document. The annual growth rate reached 125.1%, indicating explosive expansion of the field. Author collaboration averaged 4.9 co-authors per document with 26.44% of publications involving international collaboration.

Table 1. Main descriptive statistics of the dataset

Description	Results
Timespan	2020:2025
Sources (Journals, Books, etc)	1538
Documents	4890
Annual Growth Rate %	125.1
Document Average Age	0.939
Average citations per doc	10.1
Average citations per year per doc	3.544
Documents per Author	0.305
Co-Authors per Doc	4.9
International co-authorships %	26.44
International co-authorships %	23.22

3.2. Temporal Publication Trends

Figure 1 presents the temporal evolution of prompt engineering publications from 2020 to 2025. The field exhibited dramatic exponential growth characterized by three distinct phases: emergence phase (2020-2021) with 99 publications, acceleration phase (2022-2023) witnessing expansion from 145 to 819 publications coinciding with ChatGPT's release, and explosion phase (2024-2025) reaching 2,093 publications in 2024 and 1,734 through November 2025.

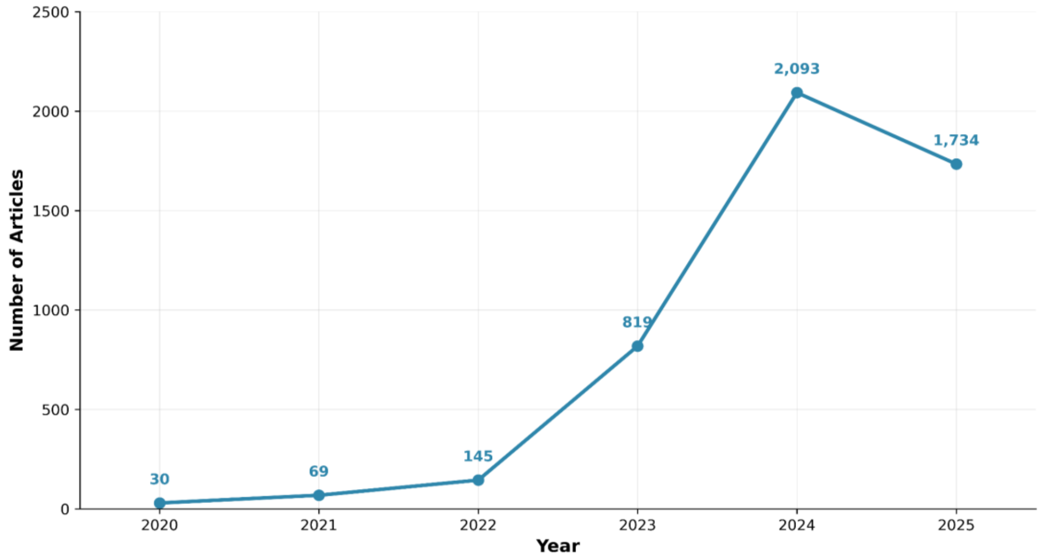

Figure 1. Annual publication output from 2000 to 2025.

3.3. Author Productivity and Impact

Table 2 presents the top 20 most productive authors measured by both absolute publication counts and fractionalized contributions. Zhang Y emerged as the most prolific author with 43 publications, followed by Liu Y (42 publications) and Wang Y (40 publications). When accounting for fractional contributions, Wang Y achieved the highest score (9.06), followed by Zhang Y (7.40) and Liu Y (7.26).

3.4. Institutional Productivity

Table 3 identifies the top 20 most productive institutions. The Chinese Academy of Sciences dominated institutional output with 345 publications, substantially exceeding the second-ranked University of California System with 171 publications. Chinese institutions occupied nine positions among the top 20, including Tsinghua University (149), Zhejiang University (133), and Peking University (125). Major technology corporations demonstrated substantial research contributions with Alphabet Inc. (125), Google Incorporated (112), and Microsoft (96) appearing among top producers, reflecting industry leadership in large language model development. United States institutions showed strong representation with University of California System, Carnegie Mellon University (85), University of Texas System (85), Harvard University (71), and Stanford University (69). Singapore maintained a notable presence through Nanyang Technological University (79) and National University of Singapore (74).

Table 2. Top 20 most productive authors

Authors	Articles	Authors	Articles Fractionalized
Zhang Y	43	Wang Y	9.06
Liu Y	42	Zhang Y	7.4
Wang Y	40	Liu Y	7.26
Chen X	33	Kim J	6.65
Kim J	31	Chen X	6.27
Wang X	30	Lee S	6.06
Li X	29	Anonymous	6
Zhang C	29	Liu H	5.69
Li J	28	Li J	5.32
Liu H	28	Wang X	5.21
Li Y	27	Zhang C	5.12
Chen L	26	Li L	5.04
Li L	26	Li Y	4.84
Lee S	25	Li X	4.84
Liu X	24	Chen L	4.8
Kim S	23	Wang L	4.71
Wang C	23	Wang H	4.64
Wang Xy	23	Wang Xy	4.35
Wang H	22	Wang C	4.29
Zhang J	22	Zhang J	4.25

Table 3. Top 20 most productive institutions

Institution	Frequency
Chinese Academy of Sciences	345
University of California System	171
Tsinghua University	149
Zhejiang University	133
Alphabet Inc.	125
Peking University	125
Google Incorporated	112
University of Chinese Academy of Sciences, Cas	106
Microsoft	96
Carnegie Mellon University	85
University of Texas System	85
Harbin Institute of Technology	81
University of Science and Technology of China, Cas	81
Nanyang Technological University	79
National University of Singapore	74
Harvard University	71
University of London	70
Stanford University	69
University of Illinois System	68
University System of Georgia	66

3.5. Country-Level Productivity and Collaboration

Table 4 presents country-level publication statistics including single-country publications (SCP), multiple-country publications (MCP), and international collaboration ratios. China produced the most publications with 1,584 documents representing 32.67% of the corpus. The United States ranked second with 1,162 publications (23.96%). India, Germany, and the United Kingdom completed the top five contributors.

Table 4. Top 20 countries by productivity and collaboration patterns

Country	Articles	Freq	SCP	MCP	MCP Ratio
China	1584	0.32667	1209	375	0.237
USA	1162	0.23964	940	222	0.191
India	199	0.04104	160	39	0.196
Germany	183	0.03774	139	44	0.24
United Kingdom	168	0.03465	96	72	0.429
Korea	153	0.03155	114	39	0.255
Italy	142	0.02928	111	31	0.218
Canada	130	0.02681	72	58	0.446
Japan	111	0.02289	83	28	0.252
Australia	102	0.02104	52	50	0.49
Singapore	89	0.01835	47	42	0.472
Spain	58	0.01196	45	13	0.224
France	49	0.01011	33	16	0.327
Switzerland	47	0.00969	33	14	0.298
Netherlands	39	0.00804	26	13	0.333
Brazil	35	0.00722	26	9	0.257
Saudi Arabia	34	0.00701	25	9	0.265
U Arab Emirates	29	0.00598	15	14	0.483
Israel	25	0.00516	15	10	0.4
Greece	24	0.00495	18	6	0.25

SCP: single country publications, MCP: multi-country publications, Freq: frequency ratio.

Geographic distribution analysis reveals distinct regional research ecosystems. East Asian countries (China, Korea, Japan, Singapore) collectively contributed 1,937 publications (39.6% of corpus), establishing the region as a major prompt engineering research hub. European nations demonstrated substantial but more distributed contributions across multiple countries, with no single European nation matching the output concentration of China or the United States. Middle Eastern countries (Saudi Arabia, United Arab Emirates, Israel) showed emerging research activity, while Latin American and African contributions remained limited. The Single Country Publications (SCP) metric identifies research conducted entirely within national boundaries. China produced 1,209 single-country publications, indicating robust domestic research capacity and potentially limited international research mobility. Similarly, the United States generated 940 domestic publications, reflecting its large-scale independent research infrastructure. Smaller nations like Singapore (47 SCP) and Switzerland (33 SCP) relied more heavily on international collaboration, consistent with their research strategies emphasizing global partnerships to overcome limited domestic researcher populations.

3.6. Country-Level Citation Impact

Table 5 presents citation impact across countries, revealing distinct patterns between productivity and influence. The United States achieved the highest total citations (25,176) and substantial average citations per article (21.67), demonstrating both volume and quality leadership. Singapore, despite modest output (89 articles), attained the highest average citations per article (37.37), indicating exceptional research impact. This pattern suggests that research quality does not necessarily correlate with quantity, with smaller nations producing highly influential work through focused research programs and international collaborations

3.7. Most Influential Publications

Table 6 presents the top 20 most cited publications in prompt engineering research. Ouyang L et al. (2022) achieved the highest citation count (4,859 citations, 1,214.8 citations per year). Wei et al. (2022) ranked second (4,123 citations, 1,030.8 per year). Liu PF (2023) achieved the highest normalized citation score (119.57 NTC) with 2,124 total citations.

3.8. Source Productivity

Table 7 presents the top 20 publication venues. IEEE Access dominated with 176 publications, followed by major natural language processing conference proceedings including ACL 2024 Findings (86 publications) and NeurIPS 2023 (76 publications).

3.9. Keyword Analysis and Thematic Structure

Table 8 presents the most frequent keywords. "Large language models" emerged as the dominant author keyword with 946 occurrences, followed by "prompt engineering" (733 occurrences) and "large language model" (474 occurrences).

Figure 2 presents a word cloud visualization of author keywords. "Large language models" and "prompt engineering" dominated the visual representation, consistent with frequency analysis. Terms such as "machine learning", "generative AI", "ChatGPT", "in-context learning", and "natural language processing" appeared prominently. Visualization confirmed the centrality of large language model concepts while highlighting the increasing importance of specific prompting techniques and application domains.

Table 5. Top 20 countries by citation impact

Country	Total Citations	Average Article Citations
USA	25176	21.666
China	8560	5.404
Singapore	3326	37.371
Japan	1595	14.369
Canada	1267	9.746
India	893	4.487
Germany	846	4.623
Korea	835	5.458
United Kingdom	752	4.476
Israel	745	29.8
Italy	704	4.958
Australia	698	6.843
U Arab Emirates	585	20.172
New Zealand	321	16.05
France	221	4.51
Greece	187	7.792
Spain	172	2.966
Denmark	156	15.6
Armenia	153	153
Macedonia	145	145

Table 6. Most Influential Publications

Paper	DOI	TC	TCperYear	NTC
Ouyang L, 2022	-	4859	1214.8	37.46
Wei JS, 2022	-	4123	1030.8	31.79
Liu PF, 2023	10.1145/3560815	2124	708	119.57
Zhou KY, 2022	10.1007/s11263-022-01653-1	1565	391.2	12.07
Kojima T, 2022	-	1310	327.5	10.1
Ruiz N, 2023	10.1109/CVPR52729.2023.02155	1085	361.7	61.08
Jia ML, 2022	10.1007/978-3-031-19827-4_41	1014	253.5	7.82
ZHOU KY, 2022, -A	10.1109/CVPR52688.2022.01631	957	239.2	7.38
Gao TY, 2021	-	765	153	20.51
Min BN, 2024	10.1145/3605943	633	316.5	122.3
Zamfrescu-Pereira JD, 2023	10.1145/3544548.3581388	470	156.7	26.46
Khattak MU, 2023	10.1109/CVPR52729.2023.01832	456	152	25.67
Reynolds L, 2021	10.1145/3411763.3451760	417	83.4	11.18
Gao P, 2024	10.1007/s11263-023-01891-x	384	192	74.19
Liu VV, 2022	10.1145/3491102.3501825	357	89.2	2.75
Liu JC, 2022	-	350	87.5	2.7
Liang J, 2023	10.1109/ICRA48891.2023.10160591	302	100.7	17
Tumanyan N, 2023	10.1109/CVPR52729.2023.00191	296	98.7	16.66
Wang ZF, 2022	10.1007/978-3-031-19809-0_36	256	64	1.97
Abid A, 2021	10.1145/3461702.3462624	227	45.4	6.09

TC: total citations, TCperYear: citations per year, NTC: normalized total citations.

Table 7. Top 20 most productive sources

Sources	Articles
IEEE ACCESS	176
FINDINGS OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: ACL 2024	86
ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 36 (NEURIPS 2023)	76
PROCEEDINGS OF THE 62ND ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS VOL 1: LONG PAPERS	68
ELECTRONICS	63
PROCEEDINGS OF THE 2024 CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: HUMAN LANGUAGE TECHNOLOGIES VOL 1: LONG PAPERS	53
FINDINGS OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS - EMNLP 2023	52
2024 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING EMNLP 2024	48
2023 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING EMNLP 2023	47
KNOWLEDGE-BASED SYSTEMS	44
2023 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING (EMNLP 2023)	41
FRONTIERS IN ARTIFICIAL INTELLIGENCE	41
FINDINGS OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (EMNLP 2023)	40
EXPERT SYSTEMS WITH APPLICATIONS	38
FINDINGS OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: NAACL 2024	38
2024 IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR)	37
NEUROCOMPUTING	36
2024 INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS IJCNN 2024	35
PROCEEDINGS OF THE 18TH INTERNATIONAL WORKSHOP ON SEMANTIC EVALUATION SEMEVAL-2024	32
PROCEEDINGS OF THE 33RD ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT CIKM 2024	30

Table 8. Top 20 most frequent author keywords

Author Keywords (DE)	Articles	Keywords-Plus (ID)	Articles
Large Language Models	946	Classification	51
Prompt Engineering	733	Model	35
Large Language Model	474	Network	31
Machine Learning	278	System	28
Generative AI	222	Networks	23
In-Context Learning	211	Artificial-Intelligence	22
Chatgpt	210	Algorithm	21
Natural Language Processing	206	Diagnosis	20
Artificial Intelligence	192	Models	20
LLM	165	Prediction	20
Deep Learning	163	Challenges	19
Prompt Learning	141	Language	19
LLMS	123	Design	18
Training	111	Framework	18
Few-Shot Learning	98	Neural-Networks	17
Large Language Models (LLMS)	93	Information	16
Data Models	66	Recognition	16
Fine-Tuning	65	Risk	15
Accuracy	62	Internet	14
Transformers	62	Management	14

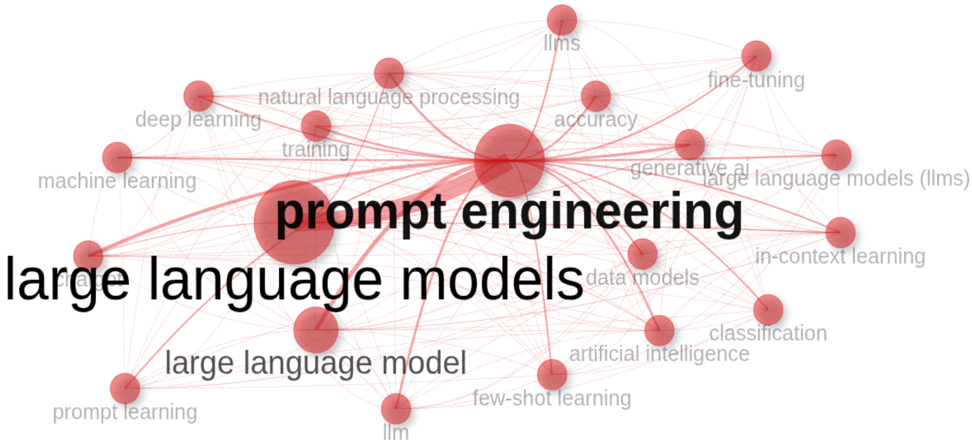

Figure 2. Keyword co-occurrence network based on keywords analysis.

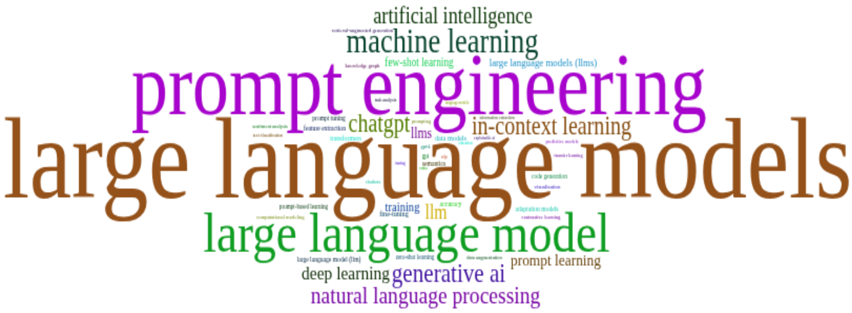

Figure 3. Word cloud of title-based keywords.

Figure 3 visualizes the keyword co-occurrence network based on author keywords. The network revealed three primary thematic clusters. The largest cluster centered on "large language models" and "prompt engineering" representing foundational concepts. A second cluster emphasized machine learning approaches including "deep learning", "neural networks", and "transformers". A third cluster focused on prompting techniques including "few-shot learning", "in-context learning", and "chain-of-thought prompting". The network structure indicated strong interconnections between theoretical prompting concepts and applied machine learning techniques.

4. Discussion and Conclusion

4.1. Publication Growth and Field Maturation

The 125.1% annual growth rate observed in prompt engineering research substantially exceeds typical expansion patterns in computer science subfields, indicating the domain's explosive emergence following widespread adoption of large language models. The temporal trajectory reveals three distinct developmental phases corresponding to major technological milestones. The emergence phase (2020-2021) produced limited output (99 publications) reflecting nascent exploration following GPT-3's introduction. The acceleration phase (2022-2023) witnessed exponential expansion from 145 to 819 publications, coinciding with ChatGPT's November 2022 release and subsequent mainstream recognition. The explosion phase (2024-2025) sustained unprecedented growth exceeding 2,000 annual publications, establishing prompt engineering as a permanent research domain within artificial intelligence. This growth pattern differs markedly from context engineering's 14.22% annual growth rate, reflecting prompt engineering's more concentrated temporal emergence within a compressed five-year window rather than gradual evolution over two decades. The field's rapid maturation compressed typical domain development cycles, with foundational methodologies, application domains, and theoretical frameworks emerging nearly simultaneously rather than through sequential accumulation. This accelerated development trajectory creates both opportunities and challenges for knowledge consolidation, quality assurance, and theoretical integration.

4.2. Geographic Distribution and Research Leadership

China's dominance with 32.67% of publications and the United States' 23.96% contribution collectively represent 56.63% of global prompt engineering research output. This concentration mirrors patterns observed in broader artificial intelligence research while exceeding context engineering's geographic distribution where China contributed 28.4% and USA 8.0%. The increased concentration in prompt engineering reflects these nations' strategic investments in large language model development and deployment, with major corporations

(OpenAI, Google, Meta, Baidu, Alibaba) driving both technological advancement and research priorities. The quality-quantity disparity between high-productivity and high-impact nations reveals distinct research strategies. Singapore's exceptional average citation impact (37.37 citations per article) from modest output demonstrates focused research programs emphasizing breakthrough contributions over volume production. Conversely, China's high productivity coupled with modest average citations (5.40) suggests rapid scaling potentially outpacing citation accumulation, consistent with national strategies prioritizing technological sovereignty and research output volume. International collaboration patterns reveal regional research ecosystems. Australia (49.0% MCP ratio), Canada (44.6%), and Singapore (47.2%) demonstrate high international integration, while China (23.7%) and USA (19.1%) maintain lower collaboration rates despite substantial productivity. These patterns reflect infrastructure capacity, policy incentives, and research culture differences, with smaller nations leveraging international partnerships to access resources and expertise while larger nations sustain independent research programs.

4.3. Author Productivity Patterns and Collaboration

The dominance of researchers with East Asian names (Zhang, Liu, Wang, Chen, Li, Kim, Lee) among top-ranked authors reflects China and South Korea's concentrated research activity in artificial intelligence and natural language processing. The relatively modest fractionalized scores even among leading authors (highest 9.06) indicate that prompt engineering research occurs predominantly through collaborative teams rather than individual efforts. This collaborative pattern aligns with the field's interdisciplinary character spanning machine learning, natural language processing, human-computer interaction, and domain-specific applications. The average of 4.9 co-authors per document exceeds context engineering's 3.83, suggesting increased collaborative complexity in prompt engineering research. This pattern may reflect the domain's requirement for diverse expertise including model architecture understanding, application domain knowledge, human factors considerations, and evaluation methodology design. The 26.44% international co-authorship rate, while substantial, remains below levels observed in mature interdisciplinary domains, suggesting continued potential for expanded global research networks.

4.4. Thematic Structure and Research Focus

Keyword analysis reveals prompt engineering's intellectual structure organized around three thematic clusters. The core cluster centers on large language models, prompt engineering methodologies, and specific techniques (few-shot learning, zero-shot learning, chain-of-thought prompting), representing the field's technical foundation. The secondary cluster emphasizes machine learning and natural language processing foundations (transformers, neural networks, training), situating prompt engineering within broader artificial intelligence contexts. The tertiary cluster focuses on application domains (classification, sentiment analysis, question answering), demonstrating practical deployment across diverse tasks. The dominance of "large language models" (946 occurrences) and "prompt engineering" (733 occurrences) as primary keywords confirms the field's clear conceptual identity despite rapid emergence. The prominence of methodological terms (in-context learning, few-shot learning, prompt learning) indicates research emphasis on fundamental techniques rather than purely application-driven investigation. However, the presence of model-specific keywords (ChatGPT with 210 occurrences) suggests potential fragmentation around particular platforms rather than unified theoretical frameworks. Trend analysis reveals temporal concentration in 2024-2025 for high-frequency terms, reflecting the field's recent maturation. Earlier topics (neural networks, IoT, edge computing) from 2021-2022 represent conceptual antecedents predating modern prompt engineering's crystallization. This pattern suggests the field synthesized existing concepts rather than emerging entirely de novo, building upon prior work in neural language models, few-shot learning, and human-AI interaction.

4.5. Source Concentration and Publication Venues

The dominance of conference proceedings over traditional journals reflects computer science publishing culture where conferences serve as primary venues for rapid research dissemination. IEEE Access's leadership (176 publications) combined with strong representation from ACL, EMNLP, and NeurIPS proceedings establishes natural language processing and machine learning conferences as principal dissemination channels. This pattern differs from context engineering where journal publications maintained a stronger presence, suggesting prompt engineering's faster development cycle favoring conference submission timelines. The prevalence of "Findings" tracks from major conferences (ACL, EMNLP, NAACL) indicates substantial submission volumes exceeding main conference acceptance capacity. These supplementary tracks provide additional publication opportunities while maintaining peer review standards, accommodating the field's explosive growth without compromising quality control. The presence of computer vision venues (CVPR 2024) reflects prompt engineering's expansion into multimodal models and vision-language applications, demonstrating disciplinary boundary crossing. Journal representation concentrates in applied artificial intelligence venues (Knowledge-Based Systems, Expert Systems with Applications, Neurocomputing) rather than pure theory publications, suggesting research emphasis on practical implementation alongside methodological innovation. Open-access journals (Electronics, Frontiers in Artificial Intelligence) provide rapid dissemination channels consistent with the field's fast-paced development and community preference for accessibility over prestige-based gatekeeping.

4.6. Citation Patterns and Knowledge Base

The most influential publications reveal both contemporary breakthrough contributions and recognition of foundational works predating prompt engineering's formalization. Ouyang et al. (2022) and Wei et al. (2022) achieved exceptional citation counts (4,859 and 4,123 respectively) reflecting their roles establishing instruction following and chain-of-thought prompting as core techniques. Liu PF (2023) attained the highest normalized citation score (119.57 NTC), indicating exceptional impact relative to publication age and field norms, likely representing a comprehensive synthesis that codified early knowledge. The temporal concentration of highly-cited papers in 2022-2023 marks the field's formative period when foundational frameworks were established. The inclusion of computer vision conference papers among top-cited works demonstrates prompt engineering's expansion beyond pure natural language processing into multimodal applications. The presence of human-computer interaction venues reflects recognition that effective prompting requires understanding human communication patterns alongside technical model capabilities. The citation patterns reveal rapid knowledge turnover characteristic of fast-developing technological domains. Recent papers from 2024 achieving substantial citations despite limited accumulation time indicate continued methodological innovation and community recognition of emerging contributions. This pattern suggests the field has not yet stabilized around canonical works, with ongoing competition among frameworks and techniques for establishing dominant paradigms.

4.7. Implications for Theory and Practice

For researchers, the bibliometric patterns identify productive institutions, influential authors, and high-impact publication venues that can inform collaboration strategies and manuscript targeting decisions. The thematic structure revealed through keyword analysis provides a conceptual map guiding literature review scope and identifying underexplored intersections between established clusters. The geographic distribution highlights regions where resources, expertise, and collaborative opportunities concentrate, informing international partnership development. For practitioners, the analysis reveals dominant techniques (few-shot learning, chain-of-thought prompting, in-context learning) that have achieved research consensus regarding effectiveness. The application-oriented keyword cluster identifies domains where prompt engineering has demonstrated practical value, guiding technology adoption decisions. The source concentration in accessible conferences and open-access journals facilitates practitioner engagement with cutting-edge research without institutional subscription barriers. The rapid growth trajectory and thematic breadth indicate prompt engineering's establishment as a distinct research domain rather than transient technological trend. However, the concentration around specific models (ChatGPT) and platforms suggests potential fragmentation risk if research emphasizes model-specific optimization over generalizable principles. Future work should prioritize theoretical frameworks transcending particular implementations while maintaining practical relevance for diverse large language model architectures.

4.8. Limitations

This study acknowledges several methodological limitations inherent to bibliometric approaches. The exclusive reliance on Web of Science Core Collection excludes publications indexed solely in Scopus, IEEE Xplore, ACL Anthology, or other specialized databases. While this choice ensures methodological consistency, it may underrepresent contributions from conference proceedings and preprint repositories that play particularly important roles in rapidly evolving technological domains. The English-language restriction potentially underrepresents research from non-Anglophone regions, particularly Chinese publications in domestic journals that may not appear in international databases despite substantial research activity. The temporal scope beginning in 2020 excludes earlier foundational work in natural language processing and machine learning that preceded modern prompt engineering's crystallization. While these antecedent contributions are acknowledged qualitatively, they do not appear in quantitative metrics, potentially obscuring intellectual lineages and conceptual continuities. The limited citation accumulation time for recent publications (2024-2025) may underestimate their eventual impact compared to earlier works with longer citation windows. The keyword analysis relies on author-provided terms and algorithmically-generated Keywords Plus, both subject to terminology inconsistencies, disciplinary variations, and strategic keyword selection for discoverability rather than accurate conceptual representation. Network analysis reveals co-occurrence patterns but cannot definitively establish causal relationships, conceptual dependencies, or qualitative assessment of thematic integration depth. Institutional affiliation data exhibited inconsistencies preventing reliable productivity analysis, reflecting metadata quality challenges in rapidly evolving fields with substantial preprint and conference proceeding publications.

4.9. Future Research Directions

The bibliometric patterns suggest several productive research directions. First, the thematic analysis revealed limited integration between technical methodology development and domain-specific application research. Future work should investigate how domain characteristics influence optimal prompting strategies, developing taxonomies that map application requirements to appropriate techniques. Second, the modest international collaboration rate indicates opportunities for establishing multinational research networks addressing prompt engineering challenges requiring diverse linguistic, cultural, and application perspectives. Third, the keyword analysis identified emerging topics including multimodal prompting, automated prompt optimization, and adversarial robustness that warrant systematic investigation (Hsieh and Lee, 2025). The integration of classical artificial intelligence techniques with neural language models represents promising directions for developing more robust and interpretable systems (Lizarraga et al., 2025). Fourth, the concentration around specific models suggests need for research emphasis on generalizable principles transcending particular implementations, developing theoretical frameworks applicable across diverse large language model architectures. Fifth, addressing challenges identified in recent literature including data bias, hallucination mitigation, few-shot generalization, and real-time responsiveness requires continued methodological innovation (Jovanovic and Voss, 2025). The creative variability introduced by automated prompt engineering raises questions about balancing innovation with adherence to task specifications and evaluation rubrics (Xue et al., 2025). Finally, the field would benefit from comprehensive synthesis works integrating findings across technical development, application deployment, and human factors research to advance theoretical coherence and practical impact.

Author Contributions

The percentages of the authors' contributions are presented below. All authors reviewed and approved the final version of the manuscript.

	A.K.	N.Ş.
C	50	50
D	50	50
S	50	50
DCP	50	50
DAI	50	50
L	50	50
W	50	50
CR	50	50
SR	50	50
PM	50	50
FA	50	50

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The authors declared that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans.

References

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., . . . Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901. https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html

Cheong, M. (2025). ChatGPT's performance evaluation in spreadsheets modelling to inform assessments redesign. *Journal of Computer Assisted Learning*, *41*(3), e70035. https://doi.org/10.1111/jcal.70035

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., & Swinton, J. (2019). *xtable: Export Tables to LaTeX or HTML* (Version 1.8-4) [R package]. https://CRAN.R-project.org/package=xtable

Hsieh, P., & Lee, W. (2025). A particle swarm optimization-based approach coupled with large language models for prompt optimization. *Expert Systems*, *42*(6), e70049. https://doi.org/10.1111/exsy.70049

Jovanovic, M., & Voss, P. (2025). Towards incremental learning in large language models: A critical review. *Expert Systems*, *42*(10), e70127. https://doi.org/10.1111/exsy.70127

Lizarraga, A., Honig, E., & Wu, Y. N. (2025). From stochastic parrots to digital intelligence: The evolution of language models and their cognitive capabilities. *WIREs Computational Statistics*, *17*(3), e70035. https://doi.org/10.1002/wics.70035

Nguyen-Duc, A., Cabrero-Daniel, B., Przybylek, A., Arora, C., Khanna, D., Herda, T., Rafiq, U., Melegati, J., Guerra, E., Kemell, K.-K., Saari, M., Nguyen, C., Yousuf, O., Matthies, C., Rafiq, A., Becker, S., Muñoz, R. F., Chanin, R., Sales, A., . . . Abrahamsson, P. (2025). Generative artificial intelligence for software engineering—A research agenda. *Software: Practice and Experience*, *55*(11), 1806–1843. https://doi.org/10.1002/spe.70005

Pranckutė, R. (2021). Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. *Publications*, *9*(1), 12. https://doi.org/10.3390/publications9010012

R Core Team. (2025). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Sasson Lazovsky, G., Raz, T., & Kenett, Y. N. (2024). The art of creative inquiry—From question asking to prompt engineering. *The Journal of Creative Behavior*, *59*(1), e671. https://doi.org/10.1002/jocb.671

Vaira, L. A., Lechien, J. R., Abbate, V., Allevi, F., Audino, G., Beltramini, G. A., Bergonzani, M., Bernini, L., Boccaletti, R., Bonvin, N., Bossi, P., Bruschini, L., Calvanese, L., Canzi, P., Carobbio, A. L. C., Cazzador, D., Cocuzza, S., Colombo, G., D'Aguanno, V., . . . De Riu, G. (2023). Accuracy of ChatGPT-generated information on head and neck and oromaxillofacial surgery: A multicenter collaborative analysis. *Otolaryngology–Head and Neck Surgery*, *170*(6), 1492–1503. https://doi.org/10.1002/ohn.489

van Diessen, E., Amerongen, R. A., Zijlmans, M., & Otte, W. M. (2024). Potential merits and flaws of large language models in epilepsy care: A critical review. *Epilepsia*, *65*(4), 873–886. https://doi.org/10.1111/epi.17907

Wang, Y., Li, L., & Chen, L. (2025). Recent advances in finetuning multimodal large language models. *AI Magazine*, *46*(3), e70025. https://doi.org/10.1002/aaai.70025

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A grammar of data manipulation* (Version 1.1.4) [R package]. https://CRAN.R-project.org/package=dplyr

Xue, M., Liu, Y., Xiao, X., & Wilson, M. (2025). Automatic prompt engineering for automatic scoring. *Journal of Educational Measurement*. Advance online publication. https://doi.org/10.1111/jedm.70002