



# Research Status and Implications of Artificial Intelligence in the Field of Chinese Language and Characters –Based on CNKI Data Research

---

Shenglan Cai and Chatuwit Keawsuwan

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

February 27, 2025

# Research Status and Implications of Artificial Intelligence in the Field of Chinese Language and Characters

## —Based on CNKI Data Research

<sup>a</sup>*Shenglan CAI, Phetchaburi Rajabhat University, Thailand*

<sup>b</sup>*Chatuwit KEAWSUWAN, Kasetsart University, Thailand*

[\\*caishenglan7@gmail.com](mailto:caishenglan7@gmail.com)

**Abstract:** With the rapid development of artificial intelligence technology, its application in the field of natural language processing is increasingly extensive. As the largest Chinese speaker in the world, the complexity and diversity of Chinese language provide unique opportunities and challenges for the research of artificial intelligence. This paper takes the literatures related to artificial intelligence in the field of Chinese language and text collected by CSSCI in CNKI as the research object. This paper uses bibliometric method and Citespace visualization tool for quantitative and qualitative analysis, summarizes the general situation of research development in this field, and summarizes the application maturity of artificial intelligence in the field of Chinese language and text. Identify research subjects, research hotspots and evolutionary paths. After 2018, the field ushered in a period of rapid development; However, researchers and institutions are relatively scattered, and cooperative research needs to be strengthened. There are some differences in the application of artificial intelligence in the field of Chinese language and writing. The research focus mainly includes machine translation, international Chinese education, language services, natural language understanding, digital humanities, etc. Finally, based on the analysis, the corresponding enlightenment of this study is obtained.

**Keywords:** Artificial intelligence, Chinese language, CNKI, CSSCI

## 1. Introduction

The term artificial intelligence (AI) was introduced at the "Dartmouth Conference" in 1956, originally signifying the goal of enabling machines to think and learn like humans, or to simulate human intelligence using computers. Technically, AI is a scientific discipline focused on researching and developing theories, methods, technologies, and applications aimed at simulating, extending, and augmenting human intelligence.

With the rapid development of information technologies such as big data and cloud computing, AI is accelerating from research to application, becoming a strategic technology in the new wave of technological revolution and industrial transformation. The Wuzhen Institute's Global Artificial Intelligence Development Report (2018) notes that AI technology is currently applied in China's education, environment, healthcare, and agriculture sectors, with education becoming a prominent area of research in recent years. ChatGPT, in particular, offers advanced services to China's education, research, culture, and business sectors. This technology helps educational institutions in China cultivate more AI experts, advance scientific research, improve cultural innovation capabilities, accelerate business development, and enhance consumer services.

In today's digital, intelligent era, AI technology is penetrating various fields at an unprecedented rate, with particular significance in the field of Chinese language and writing research. AI applications in this domain not only foster the transmission and innovation of traditional culture but also bring transformative changes to fields such as education, media, and translation. This paper aims to review the current status of AI research in the field of Chinese language and writing, exploring its achievements, challenges, and potential insights for future research and practice. The goal is to promote deeper integration and development of AI in this field.

## 2. Literature Review

The rapid development of artificial intelligence is profoundly transforming the world and human life, swiftly gaining popularity globally due to its powerful language processing capabilities and intelligent human-computer interaction functions (Liu, 2023). Reviewing the progress of AI, its integration with Western languages has proceeded smoothly. However, in the development of computational linguistics in China, there is a tendency to over-rely on deep learning algorithms, neglect language ontological research, and acknowledge that AI has yet to achieve the necessary theoretical depth. As a result, the construction and development of linguistic theories in China have been relatively slow compared to other countries (Zheng Wei & Yin Jiayi, 2022). Chinese scholar Shen Jiaxuan (2022) thus calls for active participation in transforming paradigms in linguistic research. During the exploration phase of AI, Liu Yuzi (2005) emphasized that, unlike Western languages, Chinese lacks natural delimiters between words. Consequently, he applied natural language processing (NLP) to develop an "automatic Chinese word segmentation system," creating a new approach to address the challenges of automatic segmentation in Chinese.

Subsequently, other scholars have employed AI techniques to develop new programs aimed at improving Chinese text recognition. For instance, Li Lei (2013) focused on the automatic recognition and input of medical records, designing a classifier based on multi-angle recognition and cross-validation concepts. This classifier laid the groundwork for error self-detection and self-correction functions in text recognition systems. The relationship between language and information is fundamental and central to modern society. To strengthen our position in the international community, China must actively advance Chinese language computing software, promote the information industry, enhance automation in industrial sectors, and facilitate the application of AI, as the digitization of Chinese characters holds significant implications and value for China's information society (Chen Yaoxi & Chen Honggen, 2013). Zhao Shiju (2019) observed that society is being reshaped in the age of intelligence, and China should focus on the qualities and capabilities required for a "human-machine symbiotic society." Under the "broad language perspective," it is necessary to formulate and implement a national strategy and action plan for language development that addresses current needs and anticipates future developments to comprehensively enhance national language capacity. On an individual level, people should cultivate the mindset, qualities, and skills needed to adapt to and thrive in a human-machine symbiotic society, enabling a graceful transition.

In addition to exploring AI applications in the Chinese language, numerous scholars have focused on China's minority languages and ancient scripts. For instance, Xiao Weilin (2018) used data from the China National Knowledge Infrastructure (CNKI) and applied bibliometric methods to analyze research trends, institutional development, thematic studies, and author contributions in Tibetan natural language processing (NLP). Xiao found that while early insights were made regarding the potential of this field, meaningful progress started relatively late. Furthermore, factors such as dialectal influence and the lack of corpus resources have slowed development. Due to the focus of many scholars on the structural characteristics and evolution of ancient scripts, Tian Yuan (2020) employed tools like PyCharm, PyQt5, and Python to conduct cross-disciplinary research on Warring States bamboo scripts and computerized text recognition. This allows non-specialists to quickly recognize Warring States characters, providing significant support to researchers in the field. He Danni (2022) applied AI evaluation benchmarks to construct a Zhuang language word vector semantic evaluation dataset, facilitating research on semantic similarity and correlation assessments, model selection and improvement, and downstream task performance enhancement. This research significantly contributes to the intelligent processing and advancement of minority languages and scripts in China.

Beyond applications in textual fields, Wang Xi and Chen Yang (2020) used a case study approach to examine the differences between machine-translated and human-translated texts. They found that

machine translation serves as a powerful assistant to translators, functioning as a complement rather than a replacement for human translation. Their study demonstrated that linguistic analysis of machine-translated texts can help us better assess translations. Liu Huijun (2023) found through investigative research that AI voiceovers provide better sound fidelity, enhancing the efficiency of news reporting. In the field of international Chinese education, numerous experts have conducted exploratory trials with ChatGPT for tasks such as pronunciation practice, sentence formation, grammar explanations, article drafting and editing, course design, syllabus preparation, and resource recommendations (Liu Li, 2023). Li Dongtong, Gao Ruijing, and Tian Jia (2023) tested 50 Chinese academic journal abstracts and analyzed differences before and after editing with ChatGPT 3.5. They found that ChatGPT effectively corrected grammar errors, simplified language, and standardized academic terminology, though it occasionally provided incorrect information or misused technical terms. These findings underscore the need for enhanced digital literacy in academic editing among China's science journal editors, along with the development of domestically produced AI-assisted editing tools.

In summary, research on artificial intelligence in the field of Chinese language and writing is still in its early stages, with studies primarily focused on theoretical exploration, case analysis, and software development and application. There remains a lack of comprehensive research on key themes, evolutionary paths, and application maturity in AI for Chinese language studies, with certain areas still underexplored. This paper, therefore, aims to use bibliometric and visualization analyses to present a comprehensive overview of AI research in this field within China. By identifying the main researchers, research hotspots, and evolutionary paths, and analyzing the maturity of AI applications in this field, this study provides valuable insights for institutions and researchers seeking to advance theoretical research and practical exploration of AI in Chinese language and writing studies.

### **3. Research Method**

#### *3.1 Data Sources and Processing*

This study uses Excel and Citespace visualization tools to conduct quantitative mapping analysis on research literature related to artificial intelligence in the field of Chinese language and writing, collected by CNKI from 1998 to 2024. To ensure the academic quality of the source literature, the Chinese Social Sciences Citation Index (CSSCI) database was selected as the data source. The search strategy involved keywords such as "artificial intelligence," "AI," and "ChatGPT," with the academic field set to "Chinese language and writing." No time constraints were applied, and the search date was September 1, 2024. After filtering, 420 relevant papers were identified. To better explain the research trends and application maturity of AI in Chinese linguistic studies, this paper analyzes the overall publication trends to examine development and evolution characteristics; studies authors and institutions to assess the main research entities and collaboration; evaluates the maturity of AI applications based on keywords and literature distribution; and explores recent research hotspots and evolutionary paths through keyword co-occurrence and knowledge mapping. These insights provide valuable references and experience for AI research and practice in the Chinese language and writing field.

#### *3.2 Analysis Tools*

CiteSpace is a powerful and user-friendly bibliometric analysis tool that helps researchers gain deeper insights into their research fields and improve research efficiency. It analyzes a large volume of literature, examining citation relationships, author collaboration, keyword co-occurrence, and other data to generate intuitive visual maps. These maps enable researchers to understand the development trends, knowledge structures, and research impact within a specific field. The visualizations produced by CiteSpace are straightforward and provide researchers with a comprehensive view of the research area. Due to its comprehensiveness and professionalism, CiteSpace is widely recognized in academia as a commonly used tool for bibliometric analysis.

## 4. Research Results

### 4.1 Trends in Research Literature on Artificial Intelligence

Through a statistical analysis of the annual distribution of 420 documents, it is observed that the research in this field does not exhibit a systematic trend, as illustrated in Figure 1.

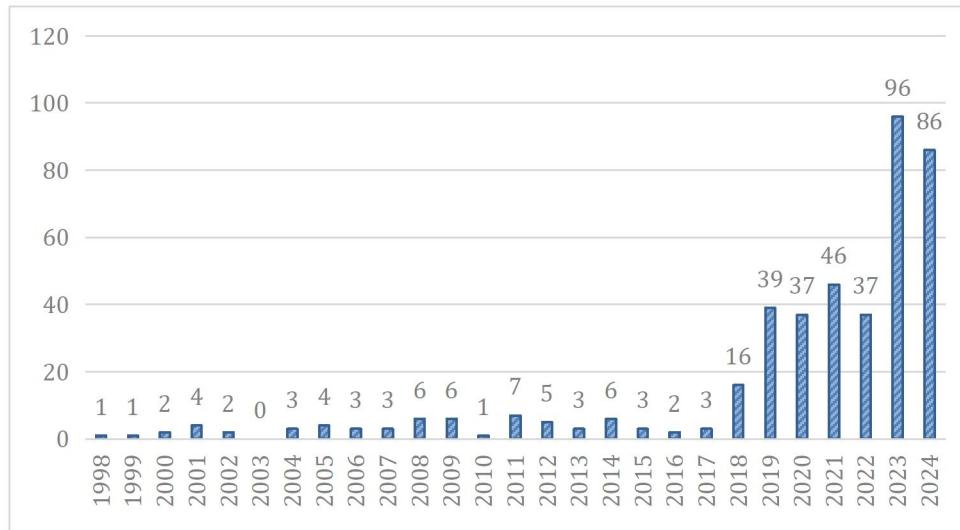


Figure 1. Annual Distribution of Research Literature on Artificial Intelligence in the Field of Chinese Language and Characters

The volume of published literature over the years reflects the theoretical level and developmental pace of artificial intelligence in the field of Chinese language and characters, as seen in Figure 1. The period from 1998 to 2017 marks the initial development stage, during which the annual publication volume remained below seven articles, with only a few scholars conducting related research. The content mainly focused on discussions about optimizing information retrieval using natural language and automatic word segmentation techniques to enhance the intelligence of information retrieval. The period from 2018 to 2022 represents a stable development phase. Although this stage is relatively shorter compared to the initial development phase, a greater number of scholars entered the field. From 2023 to the present is characterized as a rapid development phase. With the emergence of concepts like AI and ChatGPT in the public eye, along with the implementation of relevant policies and plans for artificial intelligence development in China, research in this area has experienced an explosion, resulting in a significant increase in annual publication volume and a rapid expansion of research themes.

### 4.2 Analysis of Research Themes in the Field of Artificial Intelligence

#### 4.2.1 Core Author Analysis

Statistical analysis reveals that the 420 documents in this field involve 678 authors. According to Price's Law, core authors are defined as those with a publication volume exceeding  $M=0.749*\sqrt{N_{max}}$ , where  $N_{max}=9$ . This allows us to calculate that the threshold for core authorship is approximately 2.24 articles, which we round to 3 articles. Authors with three or more publications are classified as core authors, totaling 24 individuals, or 3.5% of all authors. These core authors have collectively published 102 articles, accounting for 24.3% of the total publication volume.

This proportion significantly deviates from the theoretical value of 50% proposed by Price's Law, suggesting that research scholars in this field are relatively dispersed and have not yet formed a stable core author group. The clustering effect is not prominent. Most researchers operate independently, and there is a lack of broad consensus on many research themes and directions, indicating that the field is still in its early stages of development.

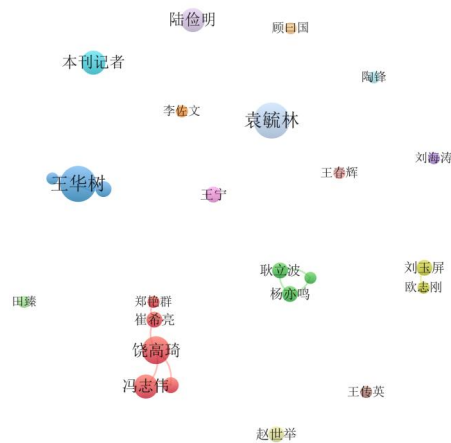


Figure 2: Author Collaboration Network Diagram

The analysis of the author collaboration network has identified 16 nodes and 11 connections. As illustrated in Figure 2, the author collaboration network is relatively sparse, primarily forming three groups of research collaboration: Feng Zhiwei, Rao Gaoqi, Cui Xiliang, Zheng Yanqun, Geng Libo; Yang Yiming and Liu Yuping; and Ou Zhigang. Researchers Wang Huashu and Yuan Yulin stand out, as shown in Table 1, with both scholars having published 9 articles each. Rao Gaoqi follows, ranking third with 7 articles. Overall, the research community appears to be quite dispersed, lacking a distinct collaboration network.

Table 1

*Selected Core Authors and Their Publication Status*

Number	Core Author	Number of Publications	Affiliation
1	Wang Huashu	9	Beijing Foreign Studies University
2	Yuan Yulin	9	University of Macau
3	Rao Gaoqi	7	Beijing Language and Culture University
4	Feng Zhiwei	6	Xinjiang University
5	Lu Jianming	6	Peking University

4.2.2 *Analysis of Publishing Institutions*

Statistical analysis reveals that research in this field involves a total of 156 institutions, including 29 non-university institutions and 127 universities. As shown in Table 2 and Figure 3, the institutional distribution indicates the emergence of core research forces led by institutions such as Beijing Foreign Studies University, Shanghai International Studies University, Peking University, and Shanghai Jiao Tong University. Among non-university institutions, the European Academy of Sciences and the Collaborative Innovation Center for Language Competency—a national science and technology information center in the field of Chinese language—hold a leading position in conducting AI-related research.

In terms of research collaboration, a cluster of institutions, including Beijing Language and Culture University, Capital Normal University, East China Normal University, Renmin University of China, and Beijing Normal University, has begun to form, representing an initial scale of cooperative research. However, collaboration intensity remains limited, with most institutions still conducting research independently. This suggests that while core institutions and collaborative groups in AI research in the field of Chinese language have started to take shape, inter-institutional cooperation is still in its early stages, and the advantages of collective research efforts are only partially realized.

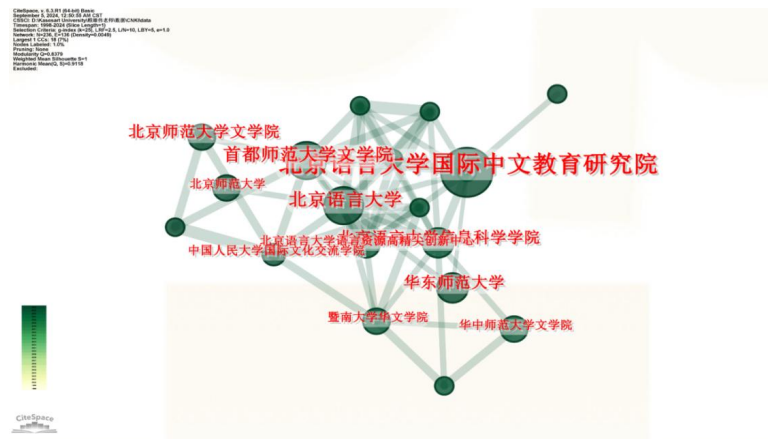


Figure 3. Diagram of Publishing Institutions

Table 2

*Selected High-Publication Institutions and Their Publication Volume*

Rank	University-based Publishing Institution	Number of Documents	Non-university-based Publishing Institution	Number of Documents
1	Beijing Foreign Studies University	18	European Academy of Sciences	2
2	Shanghai International Studies University	18	Collaborative Innovation Center for Language Ability	2
3	Institute of International Chinese Education, Beijing Language and Culture University	8	Jiangsu Key Laboratory of Language and Cognitive Neuroscience	2
4	Zhejiang University	6	Shanghai PopCap Games Co., Ltd.	1
5	School of Foreign Languages, Shanghai Jiao Tong University	5	International Society for Chinese Language Teaching	1

#### 4.2.3 Distribution of Publications by Journal

In terms of journal distribution, artificial intelligence research in the field of Chinese language primarily appears in journals related to linguistics, translation studies, education, and foreign languages. Among these, foreign language journals dominate, with high publication counts in Foreign Language Audio-Visual Teaching by Shanghai International Studies University (36 articles), Foreign Language World by Shanghai International Studies University (13 articles), Journal of Foreign Languages by Heilongjiang University (7 articles), Foreign Language Research by the University of National Defense Technology's School of International Relations (7 articles), and Foreign Language and Foreign Language Teaching by Dalian University of Foreign Languages (6 articles).

For linguistics journals, *Research on Language Strategy* by Commercial Press Ltd. has a high publication count (23 articles), as does *Applied Linguistics* by the Ministry of Education's Institute of Applied Linguistics (14 articles) and *Contemporary Rhetoric* by Fudan University (14 articles). In translation studies, key journals include *Chinese Translators Journal* by the Translators Association of China (20 articles) and *Shanghai Journal of Translators* by the Shanghai Association of Science and Technology Translators (16 articles).

These findings suggest a strong demand for publications on artificial intelligence within foreign language and linguistics journals, with a relatively lower demand in translation-focused journals. AI research supports work in linguistics through data collection, data processing, associative analysis, and semantic mining, with research intensity largely driven by this demand.

#### 4.2.4 Analysis of the Application Maturity of Artificial Intelligence in the Field of Chinese Language and Characters

Keywords represent a concise summary of article content by authors, reflecting the core focus of the literature. In this study, 1,085 keywords were first automatically identified using Citespace software, followed by manual data cleaning. To illustrate, a co-occurrence map was created (Figure 4) using the top 10% of keywords appearing in each yearly time slice. In the map, the node size represents keyword frequency.

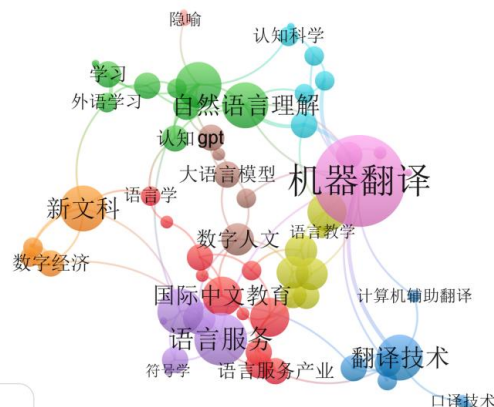


Figure 4. Keyword Co-occurrence Map

As shown in Table 3, "machine translation" first appeared in 2018, with a total frequency of 22 occurrences. "Foreign language education" and "language services" appeared for the first time in 2021 and 2019, respectively, each with 10 occurrences.

Table 3

Frequency of Selected Keywords

number	Frequency	Year	Keyword
1	22	2018	Machine Translation
2	10	2021	Foreign Language Education
3	10	2019	Language Services
4	9	2019	Translation Technology
5	8	2011	Semantic Network
6	6	2017	Big Data
7	6	2011	Language



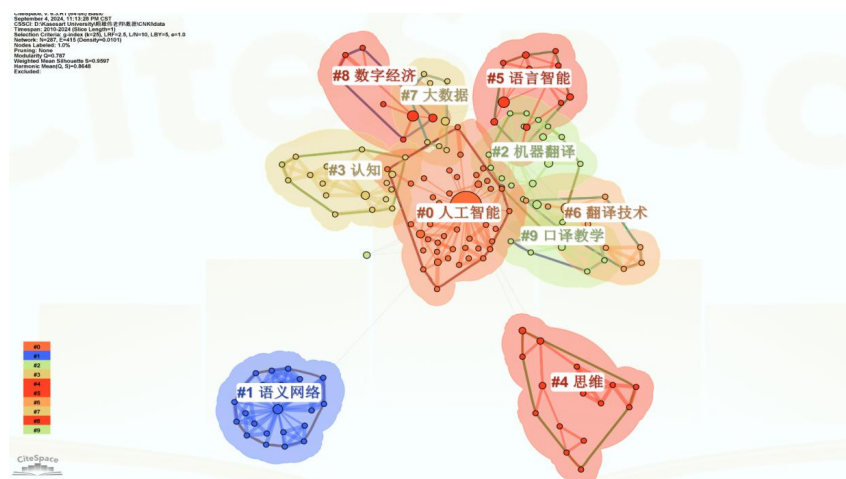


Figure 5. Keyword Clustering Map

The keyword clustering map reveals the diverse research focuses within this field. Observing Figure 5, there are 10 labels, representing 10 clusters. Each cluster label consists of keywords from the co-occurrence network, with cluster numbers ranging from #0 to #9. The higher the cluster number, the fewer keywords it contains, while a lower number indicates a larger collection of keywords within the cluster.

Most connections in the map are within clusters, though some cross-cluster connections are present. The clusters for language intelligence, machine translation, and interpreting pedagogy show a notable number of cross-cluster connections, indicating a high degree of co-citation among these research areas.

## 5. Discussion

With the rapid advancement of artificial intelligence technology, its applications in the field of language and linguistics are becoming increasingly widespread. This study examines AI-related papers in the Chinese language field, providing a comprehensive analysis of research trends, key contributors, journal distribution, research hotspots, and their evolution. Reviewing the current state of research in this field in China, the author explores the implications of AI for innovative developments in Chinese language and linguistics from the following three perspectives.

(1)Enhancing Collaboration Among Higher Education Institutions: Research reveals a high level of engagement from universities in this field. Given that the application of AI in Chinese language and linguistics is still developing, with practical application and implementation yet to be fully realized, there is a need for greater focus on applied research and practical exploration of AI technology. Although China's large language models have shown significant performance improvements, particularly with advantages in Chinese language processing, efforts should be made to achieve broader and higher-level integration of AI with Chinese language and linguistics. Beyond furthering research on key technologies such as deep learning and natural language processing, there should be active efforts to promote new AI technologies and innovations in this field.

(2)Industry Applications: Driven by the “Internet + Education” policy, applications such as intelligent educational environments and intelligent learning process support are transforming traditional educational models, enhancing teaching efficiency and personalized learning experiences, particularly

in language education. For instance, international Chinese education is currently undergoing a critical transformation period; thus, the deep integration of artificial intelligence into international Chinese education is likely to become a new research trend. The rapid development of educational technology has also emerged as a significant driving force for this transformation. Therefore, the government should cultivate talent with interdisciplinary backgrounds, which will help promote further development of AI in the field of language and linguistics, ultimately fostering more innovation and transformation in the processing and application of the Chinese language.

(3) Translation Technology: Machine translation technology has rapidly developed in China, playing a crucial role in cross-language communication and cultural dissemination. Research on subtitle translation technology has provided technical support for the international dissemination of film and television works. The application of artificial intelligence in speech recognition and synthesis has become quite mature, with speech-to-text and speech synthesis technologies being widely utilized in various fields, including intelligent assistants and customer service systems.

Research on artificial intelligence in the field of Chinese language and linguistics indicates that technological innovation is key to driving development in this area. Continuous research and algorithm improvement are essential for enhancing language processing capabilities. Additionally, the government should emphasize the necessity of interdisciplinary integration; research in the field of language and linguistics requires collaboration among computer science, linguistics, education, and other disciplines to foster a more comprehensive and in-depth research perspective.

However, in line with principles of data security and privacy protection, issues of data security and privacy cannot be overlooked during the application of artificial intelligence. This is particularly important in language processing, where personal privacy information is often involved, necessitating effective measures for its protection. Finally, the application of artificial intelligence in the field of language and linguistics not only promotes the inheritance of traditional culture but also drives innovation in language and linguistics, contributing to cultural diversity.

## 6. Conclusion

Through bibliometric and visualization analysis methods, this study conducted an in-depth investigation into the development trends, key contributors, journal distribution, research hotspots, and evolution of artificial intelligence in the field of Chinese language and linguistics. The main findings indicate that a stable core group of authors has not yet formed; research intensity is guided by demand; and while collaborative research among institutions is beginning to take shape, the intensity of collaboration requires strengthening. These findings not only enrich the understanding of the application of artificial intelligence in the field of Chinese language and linguistics but also provide insights for subsequent research.

However, this study has some limitations. Firstly, the data was sourced exclusively from CSSCI papers in CNKI, resulting in a relatively small dataset. Secondly, the research focused solely on papers published within China. Future studies could undertake comparative analyses of research focuses both within and outside of China, or conduct further investigations into the development trends of artificial intelligence in the field of Chinese language and linguistics beyond China's borders.

## References

- Deng, Q. Z. (2024). Empowering Chinese as a Second Language Teaching with ChatGPT-like Generative Artificial Intelligence: Opportunities and Challenges. *China Educational Informatization*, (03), 121-128. doi:CNKI:SUN.0.2024-03-014.
- He, D. N. (2022). Research on the Construction Method of the Zhuang Language Word Vector Semantic Evaluation Set (Master's thesis, Guangxi Normal University). Retrieved from <https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD202301&filename=1022564117.nh>

- Liu, H. J. (2023). An Exploration of the Advantages and Disadvantages of AI Dubbing: A Case Study of the Narration Dubbing in Innovative China. *News Research Guide*, (07), 7-9. doi:CNKI:SUN.0.2023-07-003.
- Li, D. T., Gao, R. J., & Tian, J. (2023). Practicality Testing and Analysis of ChatGPT in Editing Abstracts for Chinese Scientific Journals. *Chinese Journal of Scientific Journals*, (08), 1014-1019. doi:CNKI:SUN.0.2023-08-005.
- Shao, J. M., & Hu, Y. W. (2024). Innovation Requirements and Teaching Responses for "New Technology Chinese." *Journal of Yunnan Normal University (Foreign Language Teaching and Research Edition)*, (02), 71-75. doi:10.16802/j.cnki.ynsddw.2024.02.010.
- Shen, J. X. (2023). ChatGPT, Zhao Yuanren, and the New Liberal Arts: Reflections of a Linguist. *Chinese Language Strategy*, (01), 1-15. doi:CNKI:SUN.0.2023-01-001.
- Tian, Y. (2020). Research on Recognition Technology for Bamboo Slips from the Warring States Period Based on Deep Metric Learning (Master's thesis, Henan University). Retrieved from <https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD202101&filename=1020108601.nh>
- Wu, F., & Wang, Y. Y. (2024). Exploration and Practice of "Internal Integration and External Collaboration" Talent Training Paths in International Chinese Education at Local Universities. *Science and Education Wenhui*, (17), 154-158. doi:10.16871/j.cnki.kjwh.2024.17.035.
- Xiao, W. L. (2018). A Bibliometric Analysis of Tibetan Natural Language Processing Research. *Qinghai-Tibet Plateau Forum*, (01), 97-102. doi:CNKI:SUN.0.2018-01-013.
- Yu, G. W. (2019). Summary of the Eighth Academic Discussion on Research Methods and Methodology in Linguistics. *Language Science*, (04), 447-448. doi:CNKI:SUN.0.2019-04-013.