DeepSeek V3/R1 in International Chinese Language Education: Opportunities, Challenges, and Solutions

Runci Zhang[1] Ying Zhang[2]

Hangzhou Dianzi University[1-2]

Zhejiang Hangzhou, China

E-mail: 1528937348@qq.com

Abstract

The rapid development of artificial intelligence technology has injected new vitality into traditional international Chinese language education. This study focuses on the potential application of the DeepSeek V3/R1 large language models (LLMs) in the field of international Chinese language education. Through the combination of theory and data analysis, it systematically explores its value in empowering education, practical challenges, and corresponding strategies. The research results indicate that DeepSeek V3/R1 provides users with personalized learning and a visual learning platform, enhancing students' learning capabilities and improving teaching effectiveness. Simultaneously, DeepSeek V3/R1 possesses inherent technical limitations and poses challenges to the traditional education system and learners' autonomy. This study aims to provide new insights and practical references for the innovative development of international Chinese language education empowered by artificial intelligence.

Keywords: Artificial Intelligence (AI), large language models (LLMs), DeepSeek V3/R1, International Chinese Language Education, Educational

Transformation

Introduction

Launched by Hangzhou DeepSeek V3/R1 in January 2025, the generative AI tool DeepSeek V3/R1 demonstrates significant educational potential through its natural language processing (V3 model), reasoning (R1 model), and web search capabilities (Han et al., 2025), alongside advantages in performance, open-source accessibility, and cost efficiency. While GPT-4 shows established value in English education (Liu, 2024b), dedicated AI tools for international Chinese education remain underdeveloped, positioning DeepSeek V3/R1 as a pioneering solution.

With Chinese integrated into national education systems across 70 countries (25 million learners outside China) and over 40 million HSK test-takers (Dong, 2021), DeepSeek V3/R1's intervention addresses the global shortage of 44 million teachers (UNESCO, 2023) in international Chinese education.

Current challenges include the absence of ethical frameworks for AI tools (e.g., DeepSeek V3/R1) and standardized usage protocols in teaching contexts like Confucius Institutes. Establishing governance mechanisms is critical to leverage opportunities, mitigate risks, and achieve sustainable development.

Literature Review

The integration of large language models (LLMs) and education field has attracted widespread attention. With the precedent of ChatGPT-4's positive demonstration in enhancing the English proficiency of higher education students (Liu, 2024b), as the latest generation of domestically

produced large language models, it is highly anticipated whether DeepSeek V3/R1 can become a leader in the international Chinese education field.

Core advantage: DeepSeek V3/R1 has lower resource requirements and higher generation efficiency, render it more suitable for widespread educational applications (Kerimbayev et al., 2024). Furthermore, compared to traditional language-teaching LLMs, it has more advantages in data-driven personalized adaptability and provides 24/7 real-time feedback, breaking through temporal and spatial limitations. These features make DeepSeek V3/R1 as a highly efficient tool for advancing language acquisition (Liu, 2024a).

Risks and Challenges: AI systems may generate inaccurate or incorrect responses during operation (Zhang & Premy, 2025), potentially impacting users' cognitive development. Currently, DeepSeek V3/R1's functionality remains incomplete and cannot meet diverse learning needs. Additionally, the insufficient quantity and low quality of Chinese corpora (W3Techs, 2025) further diminish user experience. The extensive use of DeepSeek V3/R1 can also have a certain impact on educators: the finite cognitive capacity of the human brain cannot rival technology's computational power, which potentially undermining teachers' authority and triggering technology-related anxiety (Wang et al., 2022).

Technology is a double-edged sword, bringing us convenience but also causing certain adverse effects. The strong influence of AI enable us to rethink and innovate. However, AI has never been able to replace human unique qualities such as creativity, emotional intelligence, critical thinking, and social interaction (Zhang & Premy, 2025).

Methodology

This study adopts literature analysis and data integration to compare the performance and user experience of DeepSeek V3/R1 with other LLMs. Combining with current applications of DeepSeek V3/R1 in international Chinese language education, we analyze the opportunities, challenges, and future prospects of DeepSeek V3/R1 in empowering this field, and propose reference recommendations.

Opportunities for DeepSeek V3/R1 Empowering International Chinese Language Education

1. Empowering Personalized Chinese Learning

Global technological advancements are accelerating educational transformation, and pedagogical differentiation profoundly reshaping traditional classroom instruction. This shift necessitates adapting teaching content to learners' needs, preferences, and knowledge acquisition rates (Reddy et al., 2025). Liu et al. (2023a) empirically demonstrated that AI language models can adapt to individual learning requirements and paces, thereby promoting theoretical knowledge acquisition, motivation, and learning attitudes.

Based on a deep analysis of cross-linguistic transfer patterns and second language acquisition pathways, DeepSeek V3/R1 achieves exceptional pedagogical compatibility (9.0/10, significantly surpassing GPT-4's 8.0). It dynamically generates targeted materials for multimodal personalized learning needs (teacher-led, collaborative, and self-directed modalities). Furthermore, DeepSeek V3/R1 demonstrates robust iterative learning capabilities (Kerimbayev et al., 2024) through its cyclical refinement framework: Personalized Design - Teaching Implementation - Real time Evaluation

- Targeted Optimization.This establishes a precision-optimized personalized learning network for global learners.

2. Breaking the "Spatio-Temporal Constraints of Teacher Resources"

Traditional international Chinese language education has long been constrained by the spatiotemporal limitations of in-person instruction (Liu et al., 2023b). DeepSeek V3/R1 overcomes these barriers by reconstructing teacher supply models through AI technology.

At the level of teacher resource allocation efficiency, AI undertakes standardized tasks such as homework grading and Q&A support, enabling teachers to focus on creative teaching activities like cross-cultural communication strategy design and cognitive skill cultivation.

In terms of faculty distribution, the integration of AI technology with learning platforms delivers diverse teaching models: combining professional courses, intelligent tutoring, and digital training, to resource-deprived regions, thereby bridging the geographical education gap.

For cross-time-zone collaboration, DeepSeek V3/R1 as an always-accessible AI learning companion, eliminating time difference barriers. This allows global students to receive instant feedback anytime, anywhere while learning Chinese.

This human-AI collaborative model breaks physical constraints, constructing a dynamic pedagogical network that interconnects learners worldwide through diverse learning scenarios.

Challenges of Applying DeepSeek V3/R1 to International Chinese Language Education

1. Multidimensional Challenges of Technical Functional Limitations

DeepSeek V3/R1 operates primarily through text-based interactions. As illustrated in the model architecture diagram (see Figure 1), it is a Transformer-based model specifically designed and optimized for text tasks. While the current version lacks native support for speech input and image recognition, it relies on OCR (Optical Character Recognition) technology to extract legible text from images for subsequent processing (see Figure 2).
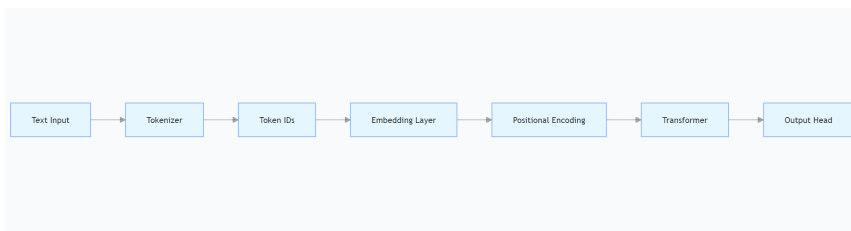

Figure 1

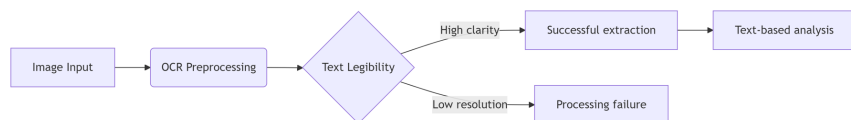*Model architecture diagram of DeepSeek V3/R1.* Created by the author, 2025.

Figure 2

Illustration of OCR text extraction process. Created by the author, 2025.

Secondly, technical constraints exist in ancient character recognition. When recognizing ancient Chinese bronze inscription characters

(Jinwen), current models can only recognize characters that have already been deciphered. They struggle to effectively handle the large number of undeciphered characters prevalent in the field of bronze inscriptions. Utilizing artificial intelligence models to decipher and interpret difficult and undeciphered characters in bronze inscriptions remains a major challenge in current research within this field (Xu, 2024).

DeepSeek V3/R1 still faces technical bottlenecks in processing phonetic loan characters (tongjiazi). For example, in the passage "乃筑台于章华之上，阙为石郭，陂汉，以象帝舜。罢弊楚国，以间陈、蔡"from the seventh year of Duke Zhao in the Zuo Zhuan, due to the model's lack of understanding of Warring States period character loan practices, it failed to correctly interpret the character "间"（jiān）as "县"（xiàn, meaning "county/district"). Instead, it understood it as "离间"（líjiàn, meaning "to sow discord"). Phenomena unique to Warring States period script, such as phonetic loans, variant writings, miscopyings, and synonymous character substitutions, pose severe challenges to the ability of large language models to deeply comprehend textual semantics (Lin et al., 2025).

Figure 3, the overall volume of Chinese corpora is insufficient, and there is a gap in high-quality Chinese language resources. According to real-time data from W3Techs on August 12, 2025, Chinese accounts for only 1.1% of global internet content, far below English's 49.2% share (W3Techs, 2025). This disparity in the total volume of Chinese and English corpora results in a severe shortage of open-source Chinese data. While English open-source data supports the operation of large models like GPT, which are built primarily on English resources, Chinese models such as DeepSeek V3/R1 have access to significantly limited web data (Zhang, 2024).

Crucially, there is a serious lack of data reserves for areas like Classical Chinese, dialects, polyphonic and polysemous characters/words, and ancient scripts. As of April 15, 2025, China's Jishou University is preparing to launch the country's first digitized Qin bamboo slips database, the "DeepSeek V3/R1-Qinjian Database V1.0". This online database contains 17,269 text images, annotating 115,996 characters and covering 2,847 categories of ancient characters (Song & Hu, 2025). In contrast, the Shuowen Jiezi (說文解字 - Explaining Graphs and Analyzing Characters), compiled by Xu Shen during the Han dynasty, included 9,353 Chinese characters (Xun et al., 2018). This comparison highlights the limited coverage of existing Classical Chinese corpora in China.

This scarcity leads to limited recognition and processing capabilities in models for both ancient and modern Chinese languages, increasing the likelihood of biases and errors. Consequently, it amplifies the difficulties faced by second language learners of Chinese in understanding and identifying linguistic elements.

Technologies > Content Languages

Usage statistics of content languages for websites

This diagram shows the percentages of websites using various content languages. See technologies overview for explanations on the methodologies used in the surveys. Our reports are updated daily.

How to read the diagram:
English is used by 49.2% of all the websites whose content language we know.
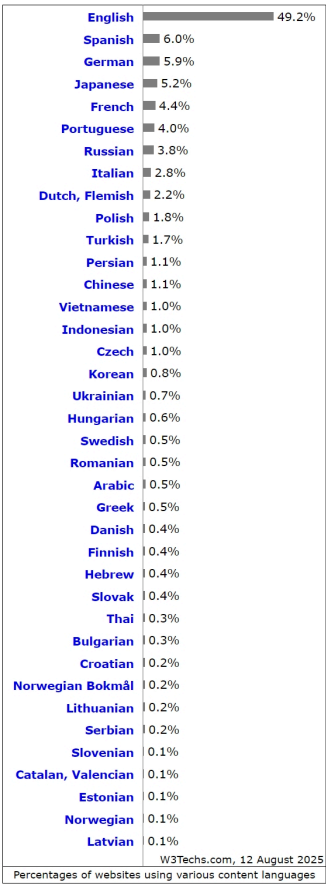
Language	Percentage
English	49.2%
Spanish	6.0%
German	5.9%
Japanese	5.2%
French	4.4%
Portuguese	4.0%
Russian	3.8%
Italian	2.8%
Dutch, Flemish	2.2%
Polish	1.8%
Turkish	1.7%
Persian	1.1%
Chinese	1.1%
Vietnamese	1.0%
Indonesian	1.0%
Czech	1.0%
Korean	0.8%
Ukrainian	0.7%
Hungarian	0.6%
Swedish	0.5%
Romanian	0.5%
Arabic	0.5%
Greek	0.5%
Danish	0.4%
Finnish	0.4%
Hebrew	0.4%
Slovak	0.4%
Thai	0.3%
Bulgarian	0.3%
Croatian	0.2%
Norwegian Bokmål	0.2%
Lithuanian	0.2%
Serbian	0.2%
Slovenian	0.1%
Catalan, Valencian	0.1%
Estonian	0.1%
Norwegian	0.1%
Latvian	0.1%

W3Techs.com, 12 August 2025

Percentages of websites using various content languages

Figure 3

Proportion and limitations of Chinese corpora compared to English resources.
Created by the author based on data from W3Techs (2025); Zhang (2024); Song and Hu (2025) & Xun et al. (2018). Image from a website: W3Techs

2. Structural Reconstruction Pressure on the Educational Authority System

Some voices in the education sector express concern that second language learners' overreliance on generative AI may impede their language proficiency development (Sangeethapriya, 2024). Research led by Stanford University showed that within their sample, 17.5% of computer science papers and 16.9% of peer review comments contained AI-generated content (Zhang & Premy, 2025). When students become accustomed to using keyword prompts to request definitions, text polishing, or even assignment answers, the phenomenon of "algorithm dependence" erodes their autonomy in language acquisition, leading to "learning stagnation".

Furthermore, DeepSeek V3/R1 may generate "plausible yet erroneous responses" during text generation (Zhang & Premy, 2025). Errors in data collection, processing, and application can result in misleading outputs for users. As an inevitable outcome of technological advancement (Jensen et al., 2025), it is necessary to establish effective regulatory mechanisms by the education sector. This will ensure both the utilization of its technical advantages and the safeguarding of academic integrity and learning outcomes.

3. Ethical Challenges to Learner Autonomy

DeepSeek V3/R1's profound integration is fundamentally restructuring the international Chinese education system. Under AI's sustained impact, the traditional pedagogical framework, where teachers derive authority from knowledge mastery and instructional dominance, now faces systemic reconstruction. This requires international Chinese language teachers to master the foundational understanding and practical skills of

DeepSeek V3/R1's operational mechanisms. Such booming technological demands, when unsupported by systematic training, risk becoming pedagogical burdens that trigger technostress (Wang et al., 2022). And when students can use AI to instantaneously access specialized knowledge surpassing teachers' expertise (e.g., etymology of classical Chinese particles, dialectal sound shift patterns), educators' core function as "knowledge arbiters" becomes significantly diminished.

Actually, DeepSeek V3/R1 challenges not merely teachers'centrality but humanity's century-old educational paradigm. Balancing technological efficiency with humanistic values while resolving educators' adaptation crisis may prove pivotal to the modernization of international Chinese language education.

Countermeasures for DeepSeek V3/R1 Empowering International Chinese Education Development

In May 2025, the Ministry of Education of the People's Republic of China published the "China Smart Education White Paper". This document reviews the development trajectory of educational digitization, formulates macro development strategies, focuses on practical exploratory measures, and outlines a blueprint for future prospects. It holds significant guiding implications for the digital-intelligent transformation of international Chinese language education.

1. Consolidate Resource Capabilities and Strengthen the Technological Foundation

To overcome bottlenecks in DeepSeek V3/R1's speech interaction, multimodal recognition, and corpus construction, a systematic technological

upgrades must be pursued. This hinges on establishing an intelligent chain for "multimodal information reception ⟶ analysis ⟶ interaction".

Develop a high-robustness speech interaction system integrating accent-adaptive speech recognition and emotional speech synthesis, supporting tone calibration and immersive scenario dialogues to enhance oral training quality.

Building specialized text recognition systems, ancient script recognition primarily relies on conventional AI image recognition (Xu, 2024), combining DeepSeek V3/R1's high computational capacity with imaging technology will significantly boost efficiency, making exceptional contributions to classical Chinese cultural research.

Construct databases for dialects, classical Chinese, and modern Chinese with reinforced polyphonic/polysemic character annotation to overcome semantic misinterpretation challenges, thereby making incremental contributions to linguistic phenomena research.

These technological implementations will propel DeepSeek V3/R1's evolution from a text processor to a multimodal language acquisition assistant, delivering precise pedagogical information through multisensory channels to resolve the "high-input-yet-low-return" dilemma in comprehensive international Chinese skills teaching.

2. Deepen Intelligent Integration and Optimize the Educational Ecosystem

To address the risk of teacher marginalization, a "Teacher-Led Technology Enhancement (TLTE)" pedagogical system should be established. In 2022, the Professional Standards for International Chinese Language

Teachers incorporates educational technology competency into its core competency system, creating an institutional pivot for reasserting teacher agency (Jin, 2023). Through practical AI tool certification programs, educators can transition from technology users to pedagogical conductors.

DeepSeek V3/R1 undertakes the responsibility of knowledge delivery, while teachers focus on embodied instruction.Taking calligraphy teaching as an example, AI provides character structure analysis. Teachers guide practical skills like brushstroke dynamics. This synergy establishes a pedagogical loop of "unity of knowing and acting" (知行合一), enabling complementary collaboration (see Figure 4).
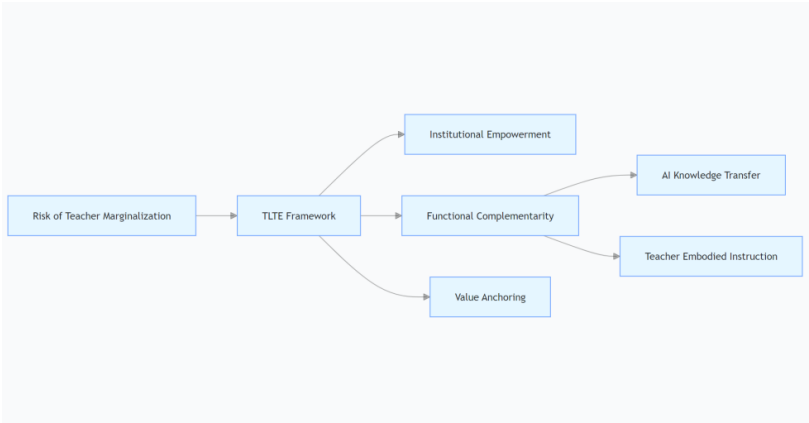

Figure 4

Collaborative pedagogical loop between AI and teacher in calligraphy instruction. Created by the author, 2025.

Only by establishing clear functional demarcation between technological empowerment and cultural perpetuation can DeepSeek V3/R1 become a sustaining force for educational innovation rather than an existential threat.

3. Safeguard Cognitive Autonomy and Resist Technological Alienation

In the process of DeepSeek V3/R1 becomes deeply embedded in international Chinese education, the tension between technological convenience and academic ethics demands an adaptive ethical framework. This requires clearly demarcating boundaries for AI usage, mandating explicit labeling of AI-generated content while prohibiting its direct submission as original academic work. Simultaneously, academic integrity education must be thoroughly integrated into curricula, establishing that appropriating AI-generated material constitutes scholarly misconduct equivalent to human plagiarism.

Assessment systems should be reconfigured through multidimensional approaches: reducing standardized testing weight while implementing qualitative evaluations as supplements, conducting integrated assessments of listening, speaking, reading, writing, and translation competencies, with particular emphasis on process-focused evaluation. Institutionally, intelligent management systems incorporating tools like Turnitin and Grammarly (Liu, 2024a), which have demonstrated capabilities in identifying AI-generated text, should be deployed to preserve academic originality and prevent DeepSeek V3/R1-enabled misconduct.

This dual-track strategy combining institutional constraints with ethical cultivation deliberately avoids extreme technological prohibition. By enhancing student agency through these measures, it achieves the essential synthesis of technological empowerment and academic integrity within evolving educational paradigms.

Conclusion

As an emerging discipline, international Chinese education bears the critical mission of disseminating language and culture in the digital era. Large language models (LLMs), represented by DeepSeek V3/R1, are injecting new momentum into innovation within this field through their powerful capabilities in generation, comprehension, and interaction. By offering personalized learning support, AI-assisted teaching, and innovations in cross-cultural communication, they significantly enhance learning efficiency and teaching experiences, providing new models for the propagation of Chinese and even global cultural exchange.

In the process of technological empowerment, we must fully leverage technical advantages to unlock AI's potential while remaining vigilant against the risks of dehumanizing effects that may weaken learning autonomy. Upholding the principle of "education as the foundation, technology as the tool", we should skillfully utilize intelligent technology while preserving the essence of humanistic education. This approach will drive high-quality development in international Chinese education, enabling the Chinese language and culture to radiate new vitality in the digital age and advancing the global dissemination of Chinese language and culture.

References

Dong, L. W. L. (2021, June 15). *Iron shoulders bear the responsibility of national revitalization: A century of exploration in education for national strength and prosperity* [铁肩担起兴邦任——教育强国富民的百年探索]. *China Education Daily.*
http://www.moe.gov.cn/jyb_xwfb/xw_zt/moe_357/2021/2021_zt15

/fjl/fjl_zongshu/202106/t20210615_538052.html

Han, X., Ma, R., & Xu, J. (2025). Exploring the technological path of empowering international Chinese teaching resource construction with DeepSeek V3/R1: Taking hierarchical reading text generation as an example. *Research on International Chinese Language Teaching,* (1), 30–40.

Jensen, L. X., Buhl, A., Sharma, A., Bekker, T., Hofkens, T., & Douliopoulos, A. (2025). Generative AI and higher education: A review of claims from the first months of ChatGPT. *Higher Education, 89*(4), 1145–1161. https://doi.org/10.1007/s10734-024-01265-3

Jin, X. (2023). The ideal functions, practical dilemmas, and application strategies of ChatGPT's deep integration into international Chinese language education. *Journal of Yunnan Normal University (Philosophy and Social Sciences Edition), 55*(4), 52–58.

Kerimbayev, N., Menlibay, Z., Garvanova, M., Djaparova, S., & Jotsov, V. (2024). A comparative analysis of generative AI models for improving learning process in higher education. In *Proceedings of the 2024 International Conference Automatics and Informatics (ICAI)* (pp. 271–276). IEEE. https://doi.org/10.1109/ICAI63388.2024.10851491

Lin, H., & Fan, C. (2025, July 18). Large language models and Warring States script research [大语言模型与战国文字研究]. *Social Sciences in China Press.* https://www.cssn.cn/skgz/bwyc/202507/t20250718_5886817.shtml

Liu, C. C., Wang, H. J., Wang, D., Tu, Y. F., Hwang, G. J., & Wang, Y. (2023a). An interactive technological solution to foster preservice teachers'

theoretical knowledge and instructional design skills: a chatbot-
based 5E learning approach. *Interactive Learning
Environments, 32*(10), 6698–6716.
https://doi.org/10.1080/10494820.2023.2277761

Liu, D. (2024a). Opportunities, challenges, and countermeasures of ChatGPT
in international Chinese language education. *Journal of
Shijiazhuang University of Applied Technology, 36*(3), 68–76.

Liu, L., Shi, Z. Q., Cui, X. L., Da, J., Tian, Y., Liang, X., ... & Hu, X. Y. (2023b).
Opportunities and challenges of ChatGPT for international Chinese
language education: Expert perspectives from the joint forum of
Beijing Language and Culture University and the Chinese Language
Teachers Association, USA. *Chinese Teaching in the World, 37*(3),
291–315. https://doi.org/10.13724/j.cnki.ctiw.2023.03.006

Liu, Y. (2024b). Reshaping and transforming of English teaching in higher
education in the ChatGPT era: An empirical study based on big
data. In *Proceedings of the 2024 IEEE 24th International
Conference on Software Quality, Reliability, and Security
Companion (QRS-C)* (pp. 1302–1311). IEEE.
https://doi.org/10.1109/QRS-C63300.2024.00169

Reddy, B. L., Nirmala, V., Lathika, V., Vidhyasree, M., Manimegalai, V., &
Kavya, K. (2025). Personalized learning systems using AI for
adaptive educational content delivery based on reinforcement
learning algorithms. In *Proceedings of the 2025 3rd International
Conference on Communication, Security, and Artificial Intelligence
(ICCSAI)* (pp. 960–964). IEEE.
https://doi.org/10.1109/ICCSAI64074.2025.11064584

Sangeethapriya, S. (2024). AI based ChatGPT impacts on second language learners. In *Proceedings of the 2024 10th International Conference on Communication and Signal Processing (ICCSP)* (pp. 790–794). IEEE. https://doi.org/10.1109/ICCSP60870.2024.10543703

Song, J. H., & Hu, L. Y. (2025, April 18). *Our university's research team has built China's first Qin bamboo slips character and text database*. Jishou University News. https://news.jsu.edu.cn/xyjj/5eb158bf0d96429794f7c44735ee3eac.htm

UNESCO, & International Task Force on Teachers for Education 2030. (2023). *Global report on teachers: Addressing teacher shortages; highlights* (ED-2023/WS/12). UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000387400

W3Techs. (2025, August 12). *Usage statistics of content languages for websites*. W3Techs. https://w3techs.com/technologies/overview/content_language

Wang, T. P., & Li, Z. (2022). The forms, causes, and countermeasures of teachers' technological anxiety in the intelligent era. *e-Education Research, 43*(10), 110–115, 128. https://doi.org/10.13811/j.cnki.eer.2022.10.015

Xu, J. Y. (2024, May 24). Application of artificial intelligence in bronze inscription research [人工智能在金文研究中的应用]. *Social Sciences in China Press.* https://cssn.cn/skqns/skqns_bsslt/202405/t20240524_5754595.shtml

Xun, L., Li, Z. Z., & Zhao, H. Y. (2018). *Postscript collation: Shuowen Jiezi*

[《说文解字》后记整理]. National Center for Ancient Book Conservation.

https://baike.sogou.com/appeal/snapshot?link=AuNbQGl3ceFEV-IKLkQkWKECh168mSK2_68m6Tpm7gm8hTrq7m5g6KozmSWg6TWL_TLg6tzi410h41U8_1Ql4Lily10siW02yWTiyQ8y_TqQ&originRef=http%3A%2F%2Fwww.nlc.cn%2Fpcab%2Fzlyj%2Fzzlj_tbzl%2F201904%2Ft20190416_177725.htm&lid=70100&title=%5Bobject%20HTMLHeadingElement%5D

Zhang, C., & Premy, T. (2025). Integrating large language models in foundational university courses: A teaching reform approach. In *Proceedings of the 2025 7th International Conference on Computer Science and Technologies in Education (CSTE)* (pp. 585–591). IEEE. https://doi.org/10.1109/CSTE64638.2025.11092138

Zhang, L. H. (2024). Accelerate the construction of Chinese training data corpus of AI large models. *People's Tribune Academic Front, 2024*(13). https://www.rmlt.com.cn/2024/0723/708072.shtml

Name: Runci Zhang

Highest Education: Undergraduate students majoring in Teaching Chinese to Speakers of Other Language

Affiliation: Hangzhou Dianzi University

Name: Dr. Ying Zhang

Highest Education: Doctor of Linguistics and Applied Linguistics

Affiliation: Hangzhou Dianzi University