English Language Teaching and Linguistics Studies
ISSN 2640-9836 (Print) ISSN 2640-9844 (Online)
Vol. 7, No. 3, 2025
www.scholink.org/ojs/index.php/eltls

Original Paper

Traditional Chinese Medicine Terminology Translation via Large Language Model: A DeepSeek-Based Study

Sili Wu[1], Yingqin Yang[1], Yuhang Yang[1] & Meng Xia[1*],

[1] Guangxi University of Chinese Medicine, Nanning city, Guangxi, China

[*] Corresponding author

Received: April 12, 2025 Accepted: May 06, 2025 Online Published: May 21, 2025

doi:10.22158/eltls.v7n3p7 URL: http://dx.doi.org/10.22158/eltls.v7n3p7

Abstract

The accuracy and standardization of TCM terminology translation have drawn significant scholarly attention during Traditional Chinese Medicine (TCM) context's internationalization. The advent of Deepseek, an advanced large language model, offers promising supports in addressing these critical linguistic and conceptual challenges, including interdisciplinary collaborative competence enhancement, rigorous translation quality control, comprehensive evaluation framework establishment, and standardized translation system development.From the perspective of large language model, Deepseek adopts a multi-strategy collaborative approach in the translation of TCM terminology to mitigate cultural barriers and domain-specific misinterpretations, thereby facilitating the global dissemination of TCM knowledge and culture.

Keywords

large language model, Deepseek, TCM translation, TCM terminology translation

1. Introduction

Traditional Chinese Medicine (TCM), a gem of traditional Chinese culture, is progressively gaining global prominence, bolstered by the dual forces of globalization and the Healthy China initiative. Deepseek, a large language model, leverages its advanced natural language processing (NLP) capabilities to offer unprecedented opportunities for the standardized translation of TCM terminology. However, significant challenges persist in translating TCM terminology with high domain-specific complexity and cultural-loaded connotations. Effectively leveraging Deepseek to achieve precise TCM terminology translation remains a critical issue requiring scholarly attention. Continued exploration of innovative methodologies and practical approaches is imperative to advance the accuracy and standardization of TCM terminology, ensuring both terminological accuracy and profound

cross-cultural comprehension.

1. Values of TCM Terminology Translation via Large Language Model

1.1 Advance the International Dissemination of TCM

TCM embodies profound cultural connotations and medical wisdom. The application of Deepseek's NLP capabilities, through the establishment of precise terminological paradigms, is capable of providing a robust foundation for the international dissemination of TCM knowledge. Deepseek's terminological accuracy helps mitigate cultural dilution effects and significantly enhances public comprehension and acceptance of TCM culture. The deep learning-optimized system of Deepseek demonstrates significant improvements in capturing both denotative and connotative dimensions of TCM terminology compared to rule-based translation systems, particularly in preserving culturally-embedded medical concepts.Leveraging its high-efficiency translation capabilities, Deepseek accelerates the translation process of TCM classics and academic literature. This technological advancement progressively expands the reach of TCM's dissemination, facilitating its integration into international academic, educational, and medical domains. By establishing a robust linguistic infrastructure for TCM, the system also substantially enhances the international influence for it.

1.2 Enhancement of TCM Terminology's Standardized Translation

The terminological system of TCM demonstrates considerable complexity and scope, encompassing foundational theories, clinical diagnostics, formulas and medical herbs, and other specialized domains. With its machine learning (ML) and deep learning (DL) capabilities, Deepseek systematically analyzes and synthesizes authoritative TCM translation corpora to identify recurrent translation patterns, ultimately extracting standardized lexical representations. Utilizing its big data statistical techniques, Deepseek conducts frequency analysis and semantic evaluation of variant translations for standardized translation, ultimately identifying renditions that align with both academic consensus and clinical application requirements. Its dynamic learning capability enables continuous integration of emerging TCM research findings and updated translation norms. By systematically incorporating novel terminology and revised renditions into its data base, the system ensures translation standards remain contemporaneous with disciplinary advancements.Specifically regarding the development of a TCM terminology knowledge base and automated verification system, Deepseek enables translators to adopt authoritative methodologies that mitigate arbitrariness in translation. This significantly enhances the standardization of TCM terminology translation processes, thereby facilitating the systematic advancement of international standardization efforts for TCM terminology (LEI, 2025).

1.3 Support of TCM Cross-Border Clinical Application

In clinical practice of TCM, specialized domains such as syndrome differentiation and treatment, acupuncture and tuina, and formula's combination of medicinals require precise terminology translation to ensure the quality and safety of cross-border healthcare services. Deepseek, with its advanced semantic comprehension and generation capabilities, conducts in-depth analysis of the medical

connotations and practical orientations embedded in TCM clinical terminology.In cross-border medical consultations and international tele-medicine scenario, Deepseek demonstrates precise translation capabilities for specialized TCM clinical documentation, including medical records, prescriptions, and treatment protocols, and effectively mitigating medication inaccuracies caused by translation errors. This technological advancement significantly enhances the real-time delivery of TCM cross-border clinical application. Through the establishment of a standardized TCM terminology translation framework, Deepseek constructs a solid linguistic bridge for the international sharing and application of TCM clinical expertise, thereby substantiating the role and value of TCM within global healthcare systems.

2. Challenges of TCM Terminology Translation via Large Language Model

2.1 Risk of Cultural Connotation Deviation

As a data-driven AI model, Deepseek demonstrates capability in learning linguistic patterns at scale, yet faces inherent limitations in comprehending the profound cultural connotations embedded within TCM terminology. The cultural schemata of TCM terms often lack conceptual equivalents in other culture’s medical paradigms, frequently resulting in cross-cultural deviations. Furthermore, the metaphorical expressions characteristic of TCM terminology—rooted in traditional Chinese symbolic thinking —may induce literal misinterpretations when the model fails to grasp their underlying cognitive foundations.Furthermore, the significant linguistic disparities between ancient Chinese expressions in TCM classics and modern language present notable challenges for Deepseek. When processing archaic vocabulary and syntactic structures, if the contextual historical-cultural analysis cannot be effectively conducted, the resulting deviations in cultural comprehension will be exacerbated. This may ultimately compromise the accuracy of TCM’s international dissemination and its appropriate recognition within international academic discourse.

2.2 Restriction of Lacking Professional Corpus

As a medical system with millennia of historical evolution, TCM encompasses multiple specialized domains featuring terminologies with significant technical complexity. However, the existing TCM translation corpora are constrained by three fundamental limitations: limited scale, inconsistent quality, and delayed updates, which substantially fail to meet Deepseek's critical need for high-quality and domain-specific data. The current deficiency in specialized TCM corpora is primarily characterized by a substantial lack of dynamic knowledge of clinical guidelines, updated formulas, and modern research findings, within classics and TCM fundamental theoretical textbooks.It also exhibit inadequate terminological annotation frameworks, resulting in significant translation discrepancies when identical terms appear across disparate textual sources—a phenomenon that may induce cognitive dissonance during model training. Furthermore, existing bilingual parallel corpora overemphasize isolated lexical representations while lacking essential contextual metadata and domain-specific annotations, impairing computational model' capacity to discern context-dependent semantic variations in TCM terminology.

The structural constraints of specialized TCM corpora significantly limit Deepseek's operational capacity in TCM terminology translation (FAN & GAO, 2025).

2.3 Adaptation Difficulties of Terminology's Dynamic Updating

As a dynamically evolving medical system, TCM exhibits continuous terminological innovation through clinical practice advancement, integration with modern medical science, and international academic exchange. These developmental trajectories elevate the adaptability benchmarks and performance standards required of Deepseek's computational framework. A significant discrepancy exists between Deepseek's training paradigm and the dynamic evolution of TCM terminology. The model's reliance on pre-training and fine-tuning mechanisms for linguistic knowledge acquisition inherently creates temporal lags in data updates, rendering it incapable of synchronizing with the developmental trajectory of TCM terminological systems. The dynamic updating of TCM terminology maintains an intrinsic relationship with contemporary reinterpretations of traditional concepts. Failure to incorporate conventional translation paradigms within the model's architecture may precipitate significant interpretative inaccuracies, potentially compromising terminological fidelity.The cross-cultural dissemination of TCM is further complicated by region-specific adaptations of its terminology, shaped by varying local medical paradigms and cultural contexts. This heterogeneity fundamentally challenges Deepseek's capacity to simultaneously address the three critical dimensions of TCM terminology translation: professional accuracy, temporal synchronization, and cultural-linguistic adaptability.

2.4 Pending Validation of Deepseek's Output

The translation of TCM terminology necessitates a multidimensional approach that prioritizes not only precise linguistic conversion but, more critically, the preservation of medical expertise and cultural authenticity. This dual emphasis on technical accuracy and contextual fidelity serves as the foundation that ensures the reliability of TCM dissemination, the validity of clinical applications, and the rigor of cross-cultural academic dialogue. As an AI-based technological system, Deepseek's terminological learning and translation processes currently lack rigorous expert validation and authoritative certification mechanisms, thereby fundamentally compromising the epistemic reliability of its textual outputs in specialized medical domains.Leveraging its DL algorithms, Deepseek generates target translations through computational corpus probability modeling. While this data-driven approach enables predictive translation processes, the system's lack of theoretical comprehension and logical reasoning regarding TCM theories renders it susceptible to both data bias and semantic misinterpretation.Concurrently, Deepseek's training data suffers from two fundamental limitations: insufficient interdisciplinary integration and temporal obsolescence, rendering it incapable of meeting the authoritative requirements for clinical diagnostics and academic research applications. This deficiency is particularly pronounced in the absence of an established certification framework for AI model specializing in TCM terminology translation. Furthermore, the lack of standardized quality assessment protocols and validation procedures has resulted in the systematic exclusion of Deepseek's

outputs from both scholarly discourse and clinical practice (Li, 2025).

3. Effective Approaches of TCM Terminology Translation via Large Language Model

3.1 Data Resource Enhancement for Robust Translation Infrastructure

TCM terminology embodies unique conceptual frameworks encompassing yin-yang theory, meridian systems, and formula combination principles.The translation of TCM terminology necessitates both terminological precision and target-language idiomatic conformity.To establish a standardized and structured TCM terminology translation corpus, it is imperative to optimize data resources through comprehensive integration and cleansing of multidimensional sources, including classical TCM texts, modern medical literature, and authoritative translations. Enriched corpora will provide Deepseek's large language model with sufficient training samples, enabling the model to develop a nuanced understanding of TCM terminology's semantic and cultural dimensions, thereby generating translations that satisfy both linguistic-cultural contexts and domain-specific requirements. The refinement of "yin-yang" terminology translation serves as a representative case study in TCM terminology optimization. Following comprehensive data resource enhancement, the translator conducted systematic extraction and analysis of bilingual text pairs from classical TCM works including *Huang Di Nei Jing (Yellow Emperor's Canon of Medicine)* and *Shang Han Za Bing Lun (Treatise on Cold Damage and Miscellaneous Diseases)*, with particular emphasis on documenting the clinical applications of yin-yang theory. Through rigorous data purification and annotation processes, the translator achieved three key outcomes: (1) identification of context-dependent semantic variations, (2) development of a multidimensional classification framework encompassing theoretical, clinical, and linguistic dimensions, and (3) creation of culturally-grounded explanatory supplements to facilitate accurate conceptual transfer. Subsequently, during Deepseek's model training, the translator integrated comprehensive contextual metadata, enabling the model to develop a sophisticated comprehension of yin-yang's philosophical constructs. This enhanced understanding establishes a theoretical framework to support TCM diagnostic and therapeutic applications.In translating the TCM term “Yin Yang Liang Xu (阴阳两虚)”, Deepseek produced accurate target-language equivalents:“Deficiency of both yin and yang”, with a short annotation: “Refers to a pathological state where both yin and yang aspects of the body are in a weakened condition, affecting various physiological functions.”This annotation enables target-language readers to accurately comprehend the medical connotations and application contexts of the terminology, thereby achieving the dual objectives of enhancing both semantic transmission and cultural dissemination in TCM terminology translation.

3.2 Domain-Specific Knowledge Integration to Enhance Translation Accuracy

TCM is a complex discipline integrating philosophy, medicine, and literature. TCM terminology exhibits specialized characteristics, encompassing both medical principles and cultural values. Integrating TCM foundational theories, clinical practices, and formula expertise into Deepseek's large language model enables the development of a TCM-aware translation system. The system achieves

deep semantic comprehension of terminological constructs while generating medically standardized and culturally adapted translations, thereby ensuring precise information transfer (LIU & TAN, 2025). The translation of “Liuwei Dihuang Wan (六味地黄丸)”, a commonly prescribed pill in TCM practice, employs as a paradigmatic case to demonstrate how integrating specialized TCM formula expertise into Deepseek's model enhances terminological precision in clinical translation.The translator identified this pill as a classical formula consisting of six herbs: Shu Dihuang (Rehmanniae Radix Praeparata),Shan Zhuyu (Fructus Corni), Shan Yao (Rhizoma Dioscoreae), Ze Xie (Rhizoma Alismatis), Mudanpi (Cortex Moutan), and Fu Ling (Poria), with demonstrated effects in nourishing yin and tonifying kidneys.The translator employed a structured data approach to integrate domain-specific knowledge into the model, encompassing both formula combination principles and standardized Latin nomenclature of medicinal materials. This knowledge incorporation explicitly prioritized three core translation objectives: accurate transmission of (1) formula composition, (2) therapeutic efficacy, and (3) cultural attributes.Following the training phase, the model generated translation of “Liuwei Dihuang Wan (六味地黄丸)”: “Liuwei Dihuang Wan (Six-Ingredient Rehmannia Pill), a classic formula in traditional Chinese medicine composed of six herbs for nourishing kidney yin, widely used to treat kidney deficiency syndromes.” The translation output systematically includes the formula’s nomenclature, component herbs, primary therapeutic effects, and supplementary clinical application information, providing readers with a multidimensional understanding.Additionally, the model is able to adapt its translation strategies for different contexts. For example, while focusing on ingredients and efficacy in drug instructions, in cultural texts it elaborates on historical origins and philosophical content.

3.3 Optimizing Interdisciplinary Collaboration for Translation Quality Control

The translation of TCM terminology constitutes a systematic endeavor encompassing descriptive, theoretical, and applied research dimensions, necessitating interdisciplinary expert collaboration to ensure clinical-practice renditions. Specifically, Deepseek can facilitate the establishment of a collaborative network integrating TCM specialists, translation scholars, clinical practitioners, and technical developers to synthesize domain-specific knowledge and practical expertise.TCM specialists provide theoretical elaboration and terminology standardization; translation scholars ensure linguistic accuracy and cultural adaptability; clinical practitioners verify the feasibility of translations in practical applications; technical developers optimize model algorithms. With multilateral collaboration, a dynamic feedback mechanism is formed. Through repeated proofreading, evaluation, and revision, the limitations of individual participants can be addressed, thereby tangibly improving translation quality.

The optimization translation of the term"Shi Xie Kun Pi (湿邪困脾)" serves as a representative case where an interdisciplinary team employed Deepseek's collaborative translation project.TCM specialists elucidated the etiology, pathological mechanisms, and clinical manifestations of the term based on *Huang Di Nei Jing (Yellow Emperor’s Canon of Medicine)* and the Warm Disease Theory. Drawing upon functionalist translation principles, translation scholars rendered the translation "Dampness

impairing spleen function," which maintains the essential medical components of the original term while conforming to standard English medical terminology conventions.Based on actual clinical practice, clinical practitioners recommended incorporating detailed symptom descriptions into the translation to optimize its functional application in medical contexts.Technical developers synthesized multidisciplinary recommendations to comprehensively optimize Deepseek's training datasets and parameters, ensuring the inclusion of medically complete translation exemplars.After repeated revisions, deepseek confirmed its output as: “Dampness Impairing Spleen Function: A pathological state characterized by the obstruction of spleen qi by dampness pathogens, leading to symptoms such as abdominal distension, loose stools, and heaviness in the limbs.”In this case, through interdisciplinary collaboration, Deepseek leveraged the strengths of various specialties, thoroughly explored the connotations of TCM terminology, ensured that the translations complied with standards and met application requirements across different scenarios, and effectively guaranteed the quality of translation.

3.4 Facilitating the Evaluation System and Promoting Standardization

Although large language model can assist in translation, and Deepseek-generated translations exhibit characteristics of high efficiency, the translated texts may still encounter issues such as semantic ambiguity and cultural misinterpretation. Furthermore, due to the lack of a precise evaluation system, it remains difficult to determine whether the translations align with the professional requirements of TCM and the cultural context of the target language. In this regard, after ensuring accuracy, professionalism, cultural adaptability, and readability, an evaluation system should be established. Only by integrating a hybrid evaluation model combining manual review and machine scoring can the quality of translations be quantified, deviations in translation be accurately identified, and the development of TCM translation toward standardization and normalization be promoted, thereby facilitating the international dissemination of TCM (MAO, 2025).

A case in point is the translation assessment and standardization of “Zi Wu Liu Zhu (子午流注)”, for which a multidisciplinary team may be convened to launch a dedicated terminology evaluation and standardization initiative. This team should rigorously observe translation protocols while remaining grounded in TCM theoretical principles, thereby establishing an exacting assessment framework to systematically evaluate prevailing translation variants through iterative review cycles.After iterative team deliberations and revisions, Deepseek confirmed “Meridian Qi Circulation Based on Midnight-Noon Rhythm” as the recommended translation. To validate the applicability of the translation, targeted surveys employing this terminology could be conducted during international acupuncture academic conferences to systematically collect expert feedback. Subsequently, the translated term should be integrated into TCM translation corpora, while machine scoring systems are employed to monitor its adoption frequency and contextual suitability in academic papers and textbooks. Ultimately, by synthesizing case study data, a comprehensive evaluation framework can be established to identify high-quality translations, thereby significantly contributing to the development

of standardized TCM terminology translation systems and the enhancement of international dissemination and exchange of TCM culture.

4. Conclusion

In summary, Deepseek exhibits considerable potential for TCM terminology translation. Through coordinated implementation of multiple strategies, Deepseek effectively addresses the cross-cultural and linguistic challenges in TCM terminology translation, ensuring terminological precision while facilitating TCM's international dissemination. Looking ahead, sustained exploration into the integration approaches between large language model and TCM terminology translation—coupled with continuous technological updates—will enable Deepseek to further propel the internationalization process of TCM.

References

FAN, W. Q., & GAO, X. (2025). To Reconstruct the Conceptual Relations in the Standardized Translation of Traditional Chinese Medicine Terminology. *Shanghai Journal of Translators*, (02), 71-77.

LEI, S. (2025). TCM Translation Research in China: Status Qua and Prospects. *Shanghai Journal of Translators*, (03), 34-38.

Li, L. (2025). DeepSeek Enabling academic library Subject service from the perspective of large language model[J/OL]. *Library Development*, 1-13[2025-05-16].

LIU, Q., & TAN, J. H. (2025). Personalized Learning of Foreign Languages Empowered by Large Language Models from Learners' Perspective—Opportunities and Challenges. *Journal of University of Shanghai for Science and Technology(Social Sciences Edition)*, *47*(01), 15-21.

MAO, H. R. (2025). A Study on the English Translation of Tibetan Medical Terminology “Three Humors” Based on the Principles of Traditional Chinese Medicine Terminology Translation. *Journal of Li-Shizhen Traditional Chinese Medicine*, *36*(02), 315-320.