# Human vs. Machine : Assessing Translation Quality of Four-Character Terms in the Classical Chinese Medical Text *Huangdi Neijing*

# Abstract

With the rapid advancement of technology, machine translation has been increasingly applied in the field of medical translation. As the foundational text of Traditional Chinese Medicine (TCM), the Huangdi Neijing plays a critical role in shaping TCM theory and clinical practice. This study focuses on the translation of four-character terms in the Huangdi Neijing and compares the translation quality between human and machine translation. A mixed-methods approach, incorporating both qualitative and quantitative analysis, was employed. A total of 463 samples were selected, and four mainstream machine translation systems—Youdao, DeepL, ChatGPT-4o, and DeepSeek were utilized.By analyzing BLEU and TER scores to evaluate the outputs against human translations, it was found that ChatGPT-4o achieved the highest BLEU score of 0.60, indicating the greatest lexical similarity to human translation, and the lowest TER score of 0.99, suggesting the fewest required edits. These results suggest that ChatGPT-4o is currently the most effective among the tested systems in handling this specialized translation task. Despite these advancements, machine translation still exhibits notable limitations when applied to TCM texts. Human translation still outperforms machine-generated outputs, particularly in conveying semantic precision and preserving the cultural and conceptual subtleties embedded in medical discourse. This study contributes to the field by providing empirical evidence on the performance of large language models in translating culturally and linguistically complex four-character expressions from classical TCM texts.

**Key Words**: Comparison between human and machine translation, four-character terms, translation quality assessment, traditional Chinese medical classics, *Huangdi Neijing*

# 1. Introduction

The global spread of Traditional Chinese Medicine (TCM) has grown quickly in recent years. This is because of China’s “Health Silk Road” plan under its Belt and Road project and because more people worldwide want alternative healthcare options. Because TCM is now part of many countries’ healthcare systems, sharing its core ideas clearly across cultures has become a key goal and a challenge for experts.

Translating medical terms have special challenges that show the bigger problems in connecting Chinese medical ideas with Western science. The *Huangdi Neijing* (*Yellow Emperor’s Internal Classic*), written over 2000 years ago, has been credited to have laid the solid foundation for TCM (Wang & Yang, 2022). Among its distinctive features, four-character terms stand out as a highly condensed and frequently employed linguistic form (Zheng, 2013). These expressions possess inherent rhythm, are concise yet rich in meaning, and often involve complex philosophical and medical concepts. This study therefore focuses on the translation of four-character terms in the *Huangdi Neijing*.

With the advancement of large-scale language models and the continuous learning of AI systems from human-generated data, machine translation has made significant progress in various domains, including medical translation. While machine translation systems have achieved reasonable accuracy in general medical contexts (Wang & Xie 2023), they continue to face substantial challenges in translating culturally embedded and semantically dense TCM terminology. Recent studies have identified three major issues: inconsistent translations across platforms, the loss of symbolic and cultural meaning, and inadequate contextual interpretation (Tian, 2024; Zhu et al., 2023). These issues are particularly obvious in classical texts such as the *Huangdi Neijing*, which are characterized by archaic syntax, historical allusions, and implicit knowledge that often require expert interpretation.

Although scholarly attention to the translation of TCM has grown, there remains a notable gap in empirical studies comparing human and machine translations of four-character terms. This study seeks to address this gap by comparing human and machine translations of selected texts from the *Huangdi Neijing*. The study aims to answer the following two questions:

1) What are the linguistic and cultural characteristics of the four-character terms in *Huangdi Neijing*?

2) What are the main differences between the human and machine translations of these terms?

## 2. Literature Review

Current scholarship has primarily examined three interrelated areas: MT studies, and specifically the translation of medicine and *Huangdi Neijing*. Artificial Intelligence has improved machine translation a lot, making translations more accurate and natural (Siu, 2023; Wang & Xie, 2023). Combining large language models and natural language processing (NLP) has helped handle small details in context and create useful translations. Machine translation tries to match the original text as closely as possible. But even with these improvements, machine translation still has problems, especially with complicated sentences, unclear meanings, and multiple descriptions. This often leads to mistakes like using wrong terms or translating information incorrectly, mostly because the system sticks too closely to the original text (Yang, 2025).

And machine translation still deals with ongoing problems. A big issue is inconsistent terms, where technical words aren’t standardized across different uses (Tian, 2024). This can confuse readers, especially in fields like medicine or engineering. Another problem is literal translations, which often don’t work well because machines lack cultural or situational knowledge (Zhu et al., 2023). For example, idioms or metaphors might lose their real meaning. Also, translated texts often have wrong facts or repeat unnecessary details.

With today’s advanced AI models, studies point to problems that still need fixing. These include not fully understanding how the models work, no shared standards for testing them, and needing to explore more how understanding connects to how well they perform (Che et al., 2023). These problems show that we need to improve machine translation tools further, especially in areas like medicine where translations need to be very precise and accurate.

Medical term translation is a very specific field needing exact wording, cultural awareness, and context understanding. Medical terms are often technical and depend heavily on their situation. Translators need deep knowledge of both languages and the medical field itself. Getting these terms right is vital, because mistakes can cause serious problems in healthcare. For example, words with multiple meanings are a big challenge. A single medical term might have different meanings based on the context. Picking the correct one needs not just language skills but also medical expertise (Randhawa, 2013). Also, metaphors and subtle medical phrases are common in texts. Translating them well requires understanding both the original and target cultures deeply.Now some scholars have made research on the translation of medical terms, especially machine translation, to ensure the consistency of machine translation of medical terms(Richard & Lena, 2023).

Machine translation is now used more for medical texts and has many benefits. A major strength is its ability to use vast knowledge bases and create translations that fit the audience’s needs (Noll et al., 2023; Merx et al., 2024). These systems can quickly handle huge amounts of medical information, helping healthcare workers get fast results. But even with these benefits, machine translation still struggles to correctly understand complex medical terms, especially words with multiple meanings, metaphors, and cultural phrases.

To fix these problems, studies suggest mixing human skills with machine translation.Useful methods are editing translations and using medical-focused databases to improve term accuracy and cultural fit. Machine translation works well for basic medical terms, but tricky or detailed terms still need human help. Human translators add cultural and situational knowledge machines don*’*t have yet, making sure translations are correct and culturally suitable.

The *Huangdi Neijing* is a key text in TCM. More researchers are now studying its translation. Earlier, they focused on matching words. Now they look at how to share culture and context. Today, the *Huangdi Neijing*’s value comes from combining science, theory, culture, and medical practice. This makes it hard to translate (Wang, 2024). Some experts have studied its English translations, looking at ways to share its deep cultural and philosophical ideas worldwide (Zhao & Shi, 2025).

But few studies look at models for translating old Chinese texts. Research on translating ancient TCM works is even rarer. TCM’s specialized ideas make it hard, leading to poor translations, awkward wording, and unclear meanings (Song, 2024). These problems are bigger for *Huangdi Neijing*, which has many technical terms, cultural ideas, and old writing styles that don’t translate easily.

Translating *Huangdi Neijing* faces big challenges. First, many TCM terms don’t match words in other languages. Translators must invent new terms or use long explanations. Also, TCM ideas are tied to Chinese culture and philosophy. Translating these ideas needs deep knowledge of both Chinese and other cultures. Another thing, the text has a poetic style and rhythm that are hard to keep in translation. But classical Chinese grammar is complex, making this hard for both humans and machines.

Machine translation struggles with *Huangdi Neijing* for a few reasons. First, unclear meanings and phrases with hidden meanings—like words with multiple meanings or metaphors—are hard for machines because they don’t understand the context well. Also, there’s not enough good Chinese-English data for training TCM translation tools. This makes it hard for machines to create smooth and correct translations of old TCM texts. Another problem is classical Chinese’s complex grammar rules. Machines often can’t handle these rules, causing mistakes in translations.

But most studies focus on TCM translation theories, the existing studies on translating *Huangdi Neijing* are rare. To fix this issue, this study compares human and machine translations of four-character terms. Using both number-based and detailed analysis, it wants to find what machines do well and where they struggle with TCM terms. The study also creates ways to improve machine translations through editing. Finally, it aims to help people better understand TCM translation in theory and practice.

# 3. Methods

This study is a mixed-methods research that integrates both qualitative and quantitative approaches to compare human and machine translation in the field of medical terms. It aims to evaluate the performance of MT systems in rendering four-character terms from *Huangdi Neijing*, focusing on their linguistic accuracy and cultural fidelity. By combining quantitative metrics such as BLEU and TER scores with qualitative error analysis, the study provides a comprehensive understanding of the strengths and limitations of MT in handling specialized TCM terminology.

## Sampling

This study employed the computer-assisted translation tool *MemoQ* to extract all four-character terms from an authoritative dictionary titled *Chinese-English Dictionary of Common Terms in Huangdi Neijing* (Wang & Yang, 2022). This dictionary includes commonly used terms from *Huangdi Neijing*, ranging from single-character entries to those with seven or even eight characters. The rationale for selecting this translation lies in the expertise of its editorial board, which consists of bilingual PhD holders in medicine who possess a strong command of both medical knowledge and English. Their professional background enables them to provide accurate and insightful interpretations of the terms found in the *Huangdi Neijing*. Moreover, the dictionary focuses on commonly used terms, thereby saving the current researcher considerable time and effort that would otherwise be required to manually align terms across the entire *Huangdi Neijing*. By using *MemoQ,* a total of 463 Chinese four-character terms and their translations were extracted , and those terms constitute the sample for the current study.

## Translation Tools

The translation tool includes popular translation engines like Youdao, DeepL, ChatGPT-4o, and DeepSeek. These were picked because they are widely trusted. Youdao is popular for Chinese-English translations, especially for everyday and technical texts. DeepL is known for accurate and smooth translations in many languages, so professionals often use it. ChatGPT-4o, a top AI model, understands context well and creates fitting translations. Lastly, DeepSeek, a newer tool, was added to see how recent AI advances work for translations.

## Data Collection Procedures

The data collection process began with the extraction of four-character terms from the*Chinese-English Dictionary of Common Terms in Huangdi Neijing* (Wang & Yang, 2022). This dictionary selects the terms from the classic translations of *Huangdi Neijing* and the international standards. Then, through our subsequent manual operation, we extract the four-character terms separately and convert them into *txt* files.When the terms were extracted, they were input into Youdao, DeepL, and ChatGPT-4o, Deepseek respectively. And their English translations were saved in different *txt* files. The reference human translation from the dictionary was served as reference file for comparison.

Data Analyses Procedures

The study first employed a Python script to automatically calculate BLEU and TER scores for evaluating machine translation quality. The BLEU score, which incorporates modified n-gram precision and a brevity penalty, was calculated using the *nltk.translate.bleu\_score*[[[1]](#footnote-0)] module from the Natural Language Toolkit (NLTK). The core of BLEU is n-gram matching. The caculation formula is *BLEU = BP \* exp（∑（w\_n \* log（p\_n）））*. In this caculation, *w\_n* represents the weight assigned to each n-gram precision (usually 0.25 for uniform weighting), and *w\_n* is the precision of the n-gram (Liu S.J., 2024). Hence, this study set the weighting as 0.25. For the TER score, the study utilized the *pyter 3*[[[2]](#footnote-1)], a Python library that provides a straightforward method for computing TER between machine-generated translations and reference human translations.

Following the automated evaluation, the meanings and implications of the BLEU and TER scores were analyzed in the context of translation adequacy and fluency. To enhance the validity of the interpretation, two TCM experts were invited to review and confirm the analysis of the results, ensuring that the findings reflected domain-specific translation challenges and expectations.

# 4. Results and Discussion

**4.1 Results**

To assess the performance of machine translation systems on four-character medical terms, both quantitative and qualitative evaluation methods were employed, with BLEU and TER scores providing the core metrics for quantitative analysis. BLEU score checks how many word groups match between machine and human translations. The BLEU score usually ranges from 0 to 1, where 0 indicates no match at all and 1 represents a perfect match. In some evaluation practices, BLEU scores are often converted to a percentage scale to make the evaluation results easier to understand and compare. Higher BLEU scores mean better word matches. TER counts changes like adding, removing, or swapping words needed to fix machine translations. Lower TER scores mean smoother sentences with better grammar. Together, BLEU checks word accuracy, and TER checks how smooth sentences read. Table 1 shows score differences between translation tools and AI models.

**Table 1 BLEU scores and TER scores**

|  |  |  |
| --- | --- | --- |
| Translation systems | BLEU scores | TER scores |
| Youdao | 0.46 | 1.26 |
| DeepL | 0.57 | 1.18 |
| ChatGPT-4o | 0.60 | 0.99 |
| Deepseek V3 | 0.60 | 1.01 |

The BLEU and TER scores for the four translation tools show clear differences. Youdao scores 0.46 in BLEU and 1.26 in TER. This shows Youdao’s translations don’t match the human references well and need many edits. This happens because Youdao uses literal translations and doesn’t handle TCM terms well.DeepL does a bit better, with a BLEU score of 0.57 and TER of 1.18. Though it translates words more accurately, the high TER score shows grammar and sentence structure problems, especially with old Chinese structures and cultural details.DeepSeek and ChatGPT-4o perform better. DeepSeek scores 0.6 (BLEU) and 1.01 (TER). ChatGPT-4o scores 0.6 (BLEU) and 0.99 (TER). Both handle TCM terms better and produce smoother translations. ChatGPT-4o needs slightly fewer edits, likely because its advanced AI understands context and culture better.

# 4.2 Discussion

# 4.2.1 The Linguistic and Cultural Characteristics of the Four-Character Terms in *Huangdi Neijing*

The four-character terms in Huangdi Neijing exhibit unique linguistic and cultural traits. Linguistically, they are highly condensed, often relying on metaphors and classical syntax. Culturally, these terms are rooted in TCM philosophy. Machine translationsstruggle to capture these nuances due to a lack of contextual and cultural awareness, often producing literal but incomplete interpretations. At the linguistic level, machine translation struggles to accurately render the deeper meanings or variant Chinese character in TCM four-character terms, often only translating their surface-level meaning. At the cultural level, machine translation may misinterpret or fail to grasp the intended references to human organs, physiological functions, or TCM cultural concepts embedded in these terms due to insufficient training in TCM knowledge. As a result, the translated outputs are often unsatisfactory.

### 4.2.2 The Main Differences Between the Human and Machine Translations of the Four-character Terms in *Huangdi Neijing*

Looking at BLEU and TER scores together helps explain translation quality. If BLEU is high and TER is low, the translation uses correct words and reads smoothly. If BLEU is high but TER is also high, the words match but the sentences sound awkward and need manual fixing. If BLEU is low and TER is low, the translation flows well but uses wrong terms. If both BLEU and TER are low, the translation is both wrong and hard to read, often because it’s overly literal or broken into pieces.

The scores show ChatGPT-4o comes out as the best model. It scores 0.6 in BLEU (word accuracy) and 0.99 in TER (smoothness), meaning it keeps terms consistent and handles cultural details better than others. This makes it good for translating TCM texts. Youdao, however, does the worst—low BLEU (0.46) and high TER (1.26)—meaning its translations use wrong words and sound awkward.This happens because Youdao relies too much on word-for-word translations and isn*’*t trained enough on medical terms.

When translating four-character terms from *Huangdi Neijing*, humans and machines work very differently. For language: humans change phrases to fit meanings and handle old Chinese grammar better.For example, humans translate “调和” as “Harmonization”, but machines stick to word-for-word versions like “Balance”. Machines also often choose less accurate words because they don’t know enough about medicine.For culture: humans change terms to match TCM ideas. For example, “天人相应” becomes “Unity of Heaven and Human” in human translations. Machines give direct translations like “Correspondence between Heaven and Human” and miss the deeper ideas.

**4.2.2.1 Differences at Linguistic Level**

TCM four-character terms tend to employ variant Chinese characters or polysemous expressions, which may pose comprehension challenges for machine translation systems. The following examples detail the differences of the semantic meanings between human and machine translation.

**Example 1**

**Source Text:**五色命藏

**Human Translation:** Five complexions serve to identify the viscera

**GPT-4o Translation:**The five colors correspond to the organ

**DeepSeek Translation:** The five colors correspond to the organ

**Youdao Translation:** Five color live hidden

**DeepL Translation:** The five colors of live

The term 五色命藏 (wǔ sè mìng zàng) is a core concept in TCM, first introduced in the *Huangdi Neijing*. It refers to the clinical observation of five specific facial or bodily hues—blue-green (青), red (赤), yellow (黄), white (白), and black (黑)—each corresponding to the functional states of the five zang organs: the liver, heart, spleen, lungs, and kidneys, respectively. This diagnostic approach goes beyond the surface meaning of “colors” as mere visual attributes; rather, it involves a nuanced interpretation of complexion changes as physiological or pathological indicators. Therefore, translating “五色” as “complexions” rather than “colors” more accurately captures its diagnostic function and clinical connotation in TCM.

The term “命藏” poses an additional challenge. While a literal translation might render it as “life storage”, such a rendering fails to reflect its embedded medical significance. In classical TCM theory, “命藏” refers to the visceral system’s role in sustaining life and maintaining internal balance. A more appropriate translation would be “viscera identification” , which aligns with TCM’s organ-based diagnostic logic and avoids misleading biomedical interpretations. Furthermore, the term “viscera” is preferred over “organs” in this context, as it preserves the distinct theoretical framework of TCM, where “zang-fu” (viscera and bowels) encompass both physiological functions and metaphysical attributes not found in modern anatomy.

**Example 2**

**Source Text:**二十五人

**Human Translation:** 25 types of people (referring to 25 constitutions in TCM)

**GPT-4o Translation:** 25 people

**DeepSeek Translation:** 25 people

**Youdao Translation:** 25 people

**DeepL Translation:** 25

The concept of “二十五人” (èr shí wǔ rén) represents a foundational model in the constitutional theory of TCM. This framework systematically categorizes human physiological, psychological, and pathological characteristics through a hierarchical integration of the Five Elements (五行 wǔ xíng) and the Five Tones (五音 wǔ yīn). Importantly, the term “二十五人” does not refer to “twenty-five individuals” in a literal sense, but rather to twenty-five distinct constitutional types of people. As such, the appropriate translation is “Twenty-Five Constitutional Types” or “Twenty-Five Types of Constitution”, rather than the misleading literal translation “twenty-five people”. This nuanced interpretation better reflects the theoretical and diagnostic depth of the term within the TCM system.

**Example 3​​**

​​**Source Text:**心部于表

**​​Human Translation:** Heart governs the surface

**​​GPT-4o Translation:** The heart is located on the exterior

**​​DeepSeek Translation:** The heart governs the exterior

**​​DeepL Translation:** Heart at the surface (of the brain)

**​​Youdao Translation:** The heart is on the table

The study found big differences in how human and machine translations handle the TCM term “心部于表” (xīn bù yú biǎo). Human translators correctly rendered it as “Heart governs the surface”, using “governs” to show the heart’s role in controlling body functions and “surface” to mean visible symptoms like skin or pulse changes. Machines struggled badly. DeepSeek’s “The heart governs the exterior” came close but used “exterior”, which is too vague for medical use. ChatGPT-4o’s “The heart is located on the exterior” mistakenly treated it as a physical position, not a functional role. DeepL made it worse by adding “(of the brain)”, which has nothing to do with TCM. Youdao’s “The heart is on the table” was completely wrong, missing both the medical meaning and common sense. These mistakes show that translating TCM terms needs deep understanding of medical ideas, not just word-for-word conversion. Human translations stay the best because they balance accurate terms and real medical meaning.

**4.2.2.2 Differences at Cultural Level**

At the cultural level, TCM four-character terms frequently employ functional metaphors of human organs to denote specific anatomical structures, or establish pathological definitions based on the theories of Yin-Yang, Five Elements, and Five Tones-Six Pitches.These culturally embedded conceptual frameworks often elude the comprehension of machine translation systems

**Example 4**

**Source Text**:下实上虚

**Human Translation**: Lower excess and upper deficiency

**GPT-4o Translation**: Fullness below and empty above

**DeepSeek Translation**: Excess below and deficiency above

**Youdao Translation**: Deficiency at the bottom and excess at the top

**DeepL Translation**: Lower part of the body is solid, upper part empty

The TCM term “下实上虚” means an imbalance of qi and blood between the upper and lower body. “下实” (lower excess) means too much active or built-up qi in the lower body (waist, belly, legs). “上虚” (upper deficiency) means not enough healthy qi in the upper body (head, face, heart, lungs). This imbalance shows a problem with the rising and falling flow of qi, called in *Huangdi Neijing* “clear qi sinking down while dirty qi rises up”.

The terms exactly explain that the lower body has more qi than needed (so “excess” instead of “full”), while the upper body lacks qi (“deficiency” instead of “empty”). This is about balance in the body’s systems, not total presence or absence. Words like “upper” and “lower” refer to body areas (upper/lower body), not positions like “top/bottom”. This careful wording keeps the medical meaning accurate by focusing on how qi is spread in the body, not exact amounts.

**Example 5**

**Source Text:**传道之府

**Human Translation:** transmission organ

**GPT-4o Translation:** the organ of transmission

**DeepSeek Translation:** the organ of transmission

**DeepL Translation:** the house of preaches

**Youdao Translation:** house of missionaries

In Traditional Chinese Medicine (TCM), the term “传道之府” is translated as “transmission organ” to explain its role in TCM’s ideas. “Transmission” means it actively collects and moves waste from the small intestine. “Organ” labels it as one of TCM’s six fu-organs. This matches how *Huangdi Neijing* describes it working through movement and change, not fixed body parts like in Western medicine. The translation keeps TCM’s focus on how parts of the body work together, not their shape. Using “transmission organ” instead of direct translations like “house of conveyance” makes it useful for real medical practice and keeps TCM’s ideas clear.

**Example 6**

**Source Text**:众之为人

**Human Translation**: people of the general yu tone

**GPT-4o Translation**: the making of a person

**DeepSeek Translation**: the Nature of the Masses

**DeepL Translation**: once a man is a man

**Youdao Translation**: All are human

The term “众之为人” describes a body type in Traditional Chinese Medicine (TCM). It is part of the Water group in the “25 Body Types” list from the Ling Shu (a key TCM text). This term refers to the most common type in the Yu-tone group, linked to a calm, water-like personality.

The translation “people of the general Yu tone” correctly explains its medical meaning as a body type system tied to the five-tone theory. This avoids direct translations like “once man is a man”, which miss the medical ideas. This term shows TCM’s detailed personality system, where body types were sorted based on music tones, elements (like water), and yin-yang ideas. The right translation keeps the original ideas and real medical use in TCM body type checks.

# 5. Conclusion

This study evaluated machine translation (MT) performance on four-character terms from *Huangdi Neijing* across word accuracy, term consistency, and cultural fit. While advanced AI models (e.g., GPT-4o, DeepSeek) demonstrated partial competence in capturing linguistic and cultural nuances, they lagged behind human translations in precision and depth. Tools like Youdao and DeepL scored significantly lower in BLEU and higher in TER than AI models, reflecting their weaker fluency and accuracy—a gap attributed to differences in training data (e.g., medical term databases) and architectural sophistication. Pre-translation curation of domain-specific term lists could enhance MT output for classical medical texts.

The challenges stem from the inherent complexity of TCM terminology: four-character terms often encode profound medical concepts that transcend literal meanings, incorporating archaic Chinese semantics and theoretical constructs absent in modern language. MT systems frequently produce erroneous or overly literal translations due to their inability to contextualize these terms clinically, whereas human translators excel at reconciling granular and conceptual layers. Key limitations include MT’s lack of specialized TCM corpora, insensitivity to classical Chinese grammar, and context-dependence of terms. Future work should prioritize building TCM knowledge graphs and embedding theoretical frameworks into MT training to bridge this gap.

**Funding Project:** 2024 Guangxi Philosophy and Social Science Research (Project number (24WYL008); Key Project of the 2024 Open Research Program of the Guangxi Research Institute for the Consciousness of the Chinese National Community (2024GXMGY0101)

**COMPETING INTERESTS DISCLAIMER:**

Authors have declared that they have no known competing financial interests OR non-financial interests OR personal relationships that could have appeared to influence the work reported in this paper.

Disclaimer (Artificial intelligence)

Author(s) hereby declare that ChatGPT-4 was used to proofread the language and correct the APA format in this manuscript. The input prompts were as such: “You are a professor in translation, please proofread the following text, making them fluency and readable in academic domain” “Please correct the following references into APA formats”.

**REFERENCE**

1. Che, W.X., Dou, Z.C., Feng, Y.S., et al. (2023) Towards a comprehensive understanding of the impact of large language models on natural language processing: challenges, opportunities and future directions (in Chinese). *Sci Sin Inform*, 53(9), 1645–1687. doi: 10.1360/SSI-2023-0113
2. Chen, Z. (2024a). Research on machine translation of TCM classics based on example corpora: A case study of four-character terms in Huangdi Neijing [基于实例语料库的中医典籍机器翻译研究——以《黄帝内经》四字格术语为例]. *English Square*, (18), 15–18. DOI:10.16723/j.cnki.yygc.2024.18.001
3. Chen, Z. (2024b). Comparative analysis and post-editing of human and machine translations of TCM classics: A case study of four-character terms in Huangdi Neijing [中医典籍人机译文对比分析与译后编辑——以《黄帝内经》四字格术语为例]. *English Square*, (15), 18–21.DOI:10.16723/j.cnki.yygc.2024.15.030
4. Liang, J., & Liu, Y. (2023). The advantage of human intelligence in translation: A corpus-based comparative study of human and machine translation. *Foreign Languages ​​and Their Teaching*, (3), 74–84, 147–148. Doi:10.13458/j.cnki.flatt.004946.
5. Merx, R., Phillips, C., & Suominen, H. (2024). Machine translation technology in health: A scoping review. *Stud Health Technol Inform*, 318, 78-83.
6. Noll, R., Frischen, L. S., Boeker, M., Storf, H., & Schaaf, J. (2023). Machine translation of standardized medical terminology using natural language processing: A scoping review. *New Biotechnology*, 77, 120-129. https://doi.org/10.1016/j.nbt.2023.08.004
7. Richard Noll, Lena S. Frischen, Martin Boeker, Holger Storf, & Jannik Schaaf. (2023). Machine translation of standardized medical terminology using natural language processing: *A scoping review. New Biotechnology*, 77, 120-129. https://doi.org/10.1016/j.nbt.2023.08.004
8. Schulz, S., Bernhardt-Melischnig, J., Kreuzthaler, M., Daumke, P., & Boeker, M. (2013). Machine vs. human translation of SNOMED CT terms. In C. U. Lehmann et al. (Eds.), *Studies in Health Technology and Informatics* (pp. 581–584). IOS Press.
9. Siu, S. C. (2023). ChatGPT and GPT-4 for professional translators: Exploring the potential of large language models in translation. Available at SSRN 4448091.
10. Song, X., Zhou, J., & Liu, W. (2024). Study on the Construction of Intelligent Machine Translation Model for TCM Ancient Books. *Chinese Journal of Library and Information Science for Traditional Chinese Medicine*, 48 (6), 130–135. DOI: 10.3969/j.issn.2095-5707.202312075
11. Tian, F. J.(2024). A comparative study of human and machine translation quality: A case study of public sign translation in District A [人工与机器翻译质量对比研究——以A区公示语英译为例]. *English Square*, (19), 45–49. DOI:10.16723/j.cnki.yygc.2024.19.027
12. Wang E.L. & Yang Y. (2022) *Chinese-English Dictionary of Common Terms in Huang Di Nei Jing (Yellow Emperor’s Internal Classic)*. Shanghai Scientific & Technical Publishers.
13. Wang, H.S., & Xie, Y. (2023). The development of translation technology in the Era of ChatGPT era and its implications. *Foreign Languages and Cultures*, 7(4), 80–89. DOI: 10.19967/j.cnki.flc.2023.04.008
14. Wang, Q. (2024). The Contemporary Value of The Yellow Emperor's Inner Classic. *Journal of Nanjing University of Traditional Chinese Medicine*,40 (10), 993–1003. DOI:10. 14148 / j. issn. 1672-0482. 2024. 0993
15. Yang, B., & Leng, B. (2025). An empirical analysis of machine translation error types in specialized texts based on the MQM quality evaluation model [An empirical analysis of machine translation error types in specialized texts based on the MQM quality evaluation model]. *Journal of University of Shanghai for Science and Technology (Social Science Edition)*, 1–7.
16. Zhao, H.F., & Shi, C.R. (2025). A Study on the English Translation of Huangdi Neijing Lingshu from the Perspective of Translator’s Voice. C*hinese Journal of Basic Medicine in Traditional Chinese Medicine*, 31(01), 120-123. Doi:10.19945/j.cnki.issn.1006-3250.2025.01.031.
17. Zheng, L. (2013). *English reading and writing tutorial for traditional Chinese medicine [中医英语读写教程]*. Ancient Books Publishing House of Traditional Chinese Medicine
18. Zhu, J., Jiang, J., Yang, M., & Ren, Z. J. (2023). ChatGPT and Environmental Research. ChatGPT and environmental research. *Environmental Science & Technology*, 57(46), 17667–17670. https://doi.org/10.1021/acs.est.3c01818
19. Zhu, Y. (2024). ChatGPT enables translation quality assessment: opportunities, challenges and countermeasures.*Journal of Zhejiang Wanli University*, 37(05):88-94. DOI:10.13777/j.cnki.issn1671-2250.2024.05.007.

1. https://www.nltk.org/\_modules/nltk/translate/bleu\_score.html [↑](#footnote-ref-0)
2. https://pypi.org/project/pyter3/ [↑](#footnote-ref-1)