A Comparative Study on The Performance of ChatGPT-4 and Claude 3 in Translating Engineering Specialized Course Textbooks

Yiqin Jiang
Institute of International and Comparative Education, East China Normal University, Shanghai, 200062, China

Keywords: AIGC, Machine Learning, Translation Applications.

Abstract: This study examines the advantages and disadvantages of Claude 3 and ChatGPT-4 for translating engineering texts. Both AI technologies analyze scientific texts from four engineering textbooks for accuracy, fluency, audience appropriateness and logicality using the CO-STAR framework. The translation outcomes of AI tools were contrasted and examined with those of rule-based machine translation and human translation of the four paragraphs. Overall, Claude 3 performs better than ChatGPT-4, according to the results, and ChatGPT-4 has to be enhanced in terms of audience adaption and fluency. Nevertheless, the overall quality of the translations is inferior to that of human translators, and both AI systems have representational flaws. This study emphasizes the need for improvement in the distribution of educational resources across languages, as well as the potential of AI technologies for translating engineering textbooks.

1 INTRODUCTION

The sharing of educational resources across languages has become a significant force to promote education and has drawn the attention of many nations as a result of the internationalization of education and the ongoing advancement of communication technologies (Ahmadova, 2024). The process of teaching knowledge and concepts for the practice of engineering professions is known as engineering education (Kadhim & Hussein, 2024). One of the key issues facing engineering education is globalization, which requires engineers to possess a variety of cultural and communication abilities (Ahrenberg, 2017). Thus, there is an urgent need for learning materials for engineering education to be shared and distributed throughout the world. Nonetheless, there are numerous issues with the way specialized engineering course textbooks are now translated, and when translating scientific literature, it's critical to consider linguistic and cultural variations (Balas et al,2024). Engineering textbooks' precise terminology and theoretical explanations make it possible for manual translation or rule-based machine translation to highlight its shortcomings, such as inefficiency and low accuracy (Simpson & Weiner,1989; Duderstadt, 2007). This makes it difficult for engineering education learning resources to be disseminated internationally. As generative artificial intelligence advances, it opens up new avenues for translating engineering textbooks and offers a revolutionary solution to the aforementioned issues.

It has been discovered that Artificial Intelligence Generated Content (AIGC) technology is extensively utilized in the translation industry in addition to being utilized in the domains of cross-linguistic communication and cultural exchange (Ferrag & Bentounsi, 2024). In addition to having higher translation quality and efficiency than human translation, Artificial Intelligence (AI) translation is superior to machine translation (Haq et al, 2024; Li et al, 2024). To encourage global scientific and technical communication, AIGC technology can raise the translation quality in the interim (Li, 2024). AI translators do have some limitations, though: first, they are not as good at handling logical expressions and textual connotations (Li, 2024); second, they are not as good at adapting to different fields of specialization (Li,2022), and third, while the quality of translations is on par with intermediate human translators, they still fall short of advanced translators (Lommel et al, 2024). Numerous research have also been conducted on the efficacy of AI translation applications in several specialized domains. Artificial intelligence (AI) does well in the literary domain when it comes to word choice and narrative, but it struggles to handle cultural and subtle emotional distinctions (Mohsen, 2024). AI translations are

Jiang, Y.
A Comparative Study on the Performance of ChatGPT-4 and Claude 3 in Translating Engineering Specialized Course Textbooks.
DOI: 10.5220/0013999300004912

In *Proceedings of the 1st International Conference on Innovative Education and Social Development (IESD 2025)*, pages 454-460
ISBN: 978-989-758-779-5

typically less accurate than human translators in the legal domain (Quigley et al,2011). Studies in the medical field have also assessed and contrasted the performance of various AI tools in translating educational materials for neurological patients and ophthalmic terminology, with ChatGPT performing better in the latter case and Claude performing better in the former (Shokoohifar, 2024; Wang, 2023). Thus, it is yet unclear and worthwhile to investigate if these two AI techniques can translate engineering texts. It is evident that the majority of current research is concentrated on AI translation in the fields of law, medical, and literature, whereas engineering research is at a lower level. Although the scientific text is primarily based on objective facts, it is still influenced by cultural differences in terms of expression, according to the author, who believes that the use of AI tools to translate engineering textbooks will face many challenges. First, the text is extremely specialized, and arbitrary additions and omissions could cause it to lose its original meaning; second, even though the scientific text is primarily grounded in objective facts, cultural differences in expression still have an impact on it (Wang,2023). In order to provide a higher-quality, more accurate translation, this research will investigate the benefits and drawbacks of AI technologies in engineering textbook translation.

Among the various AI tools used in this study, ChatGPT-4 and Claude 3 were selected for two reasons: first, they are the most popular in the translation field when compared to other AI tools like Kimi; second, ChatGPT-4 does well when translating scientific texts (Yan et al, 2024), and Claude does better than ChatGPT when translating into English (Yan & Zhao, 2023). As a result, both are equally capable of translating technical texts. In this study, the translation skills and outcomes of these two AI systems for specialized engineering course texts will be evaluated and compared for similarities and differences. This study is innovative in two ways: first, it closes the knowledge gap regarding the potential and constraints of AI tools for translating specialized engineering science texts; second, it optimizes the research methodology by using the Multidimensional Quality Metrics (MQM) evaluation framework as a foundation for screening relevant evaluation indicators. The significance of this study is threefold: 1. To encourage the exchange of educational resources across languages and to compare the strengths and shortcomings of Claude 3 and ChatGPT-4 in translating engineering textbooks in order to enhance their quality. 2. Research findings have the potential to foster creativity and collaboration in international education. 3. This study can also offer fresh perspectives on the investigation of engineering talent development in Science, Technology, Engineering, Mathematics (STEM) education within an interdisciplinary setting.

2 RESEARCH METHOD

This study compares the translation capabilities of Claude 3 and ChatGPT-4 for engineering texts. For this study, two textbooks that are fairly representative of the Chinese-US Mechanical Design and Manufacturing and Automation joint engineering curriculum at Shanghai Normal University were selected. These two texts are Fluid Mechanics Fundamentals and Applications (2nd edition) and Thermodynamics: An Engineering Approach (7th edition). The author chose two scientific texts from each engineering textbook that addressed specialized terminology and the background of specialized domain knowledge for additional AI translations in order to assess whether these two AI tools could successfully negotiate the numerous challenges associated with translating engineering textbooks. To assess the quality of the AI translations, the author included the translation results in a questionnaire that was given to Shanghai Normal University engineering majors, ranging from freshmen to seniors, along with their professors. The author then contrasted and examined the rule-based machine translation, artificial intelligence translation, and human-translated materials after using the DeepL program and hiring a qualified English translator to translate the textbook simultaneously.

2.1 Textbooks Chosen for the Analysis

Thermodynamics: An Engineering Approach (7th edition), written by Yunus A. Cengel and Michael A. Boles, is the first textbook. Important topics like energy, the second law of thermodynamics, entropy, the gas power cycle, chemistry, and phase equilibrium are covered in this book, which presents the fundamental ideas and laws of engineering thermodynamics and their applications. The book also includes numerous instances that are explored together with technical applications. Undergraduate and graduate students studying energy and power engineering, architecture, machinery, and other relevant fields at universities can use this book as a textbook.

Fluid Mechanics Fundamentals and Applications (2nd edition), written by Yunus A. Cengel and John

M. Cimbala, is the second book. This book provides many examples of engineering while introducing the fundamental ideas and formulas of fluid mechanics. It discusses crucial topics such fluid kinematics, pressure and hydrostatics, Bernoulli equations, and energy equations. Undergraduate and graduate students studying physical engineering, chemical engineering, aerospace, and other relevant fields can use this book as a university textbook.

These two volumes, which are standard textbooks for Chinese-foreign cooperative education in engineering, cover a lot of information in several engineering disciplines and are appropriate for use in this study.

2.2 Procedures

The author intends to start by providing ChatGPT-4 with the same CO-STAR command framework as Claude 3 (with “C” for Context, “O” for Objective, “S” for Style, “T” for Tone, “A” for Audience, and “R” for Response). These two AI systems ought to be able to more precisely and expertly translate technical materials from English into Chinese with the help of this framework. Two factors led to the selection of the CO-STAR education framework: 1. It is thorough, producing translated text that is accurate and readable. 2. It is focused, making translations in specialized fields easier.

Second, upon the completion of the translation, the writers will ask professionals in the field who have undergone these specialized courses to assess the outcomes. A framework for assessing the quality of translations produced by humans, machines, and artificial intelligence is called MQM (Multi-dimensional Quality Metrics) (Zhang & Li, 2009). Although there are seven assessment indicators in the MQM system, not all of them can be used when translating scientific materials in engineering textbooks. Therefore, three variables from the MQM framework — accuracy, fluency, and audience adaptability—were selected for this study in order to reduce redundancy and increase evaluation efficiency. In addition, the inclusion of the indication and logic makes it easier to assess how well the scientific text's reasoning and argumentation process is supported in the translation. Table 1 displays the four indicators' definitions and evaluation goals. As indicated in Table 2, a Likert scale was used to assess each indicator from 1 to 5. In this study, a five-level scale questionnaire with four assessment indicators was designed using a quantitative-qualitative and qualitative-mixed research methodology. This allowed for both a qualitative analysis of the differences

Table 1: Explanation of indicators.

Indicator	Definition	Objectives
Accuracy	The translation faithfully captures the original text's content, message, and intent.	Evaluate whether the translation faithfully captures the original text's meaning and make sure that terminology is not mistranslated when translating scientific texts.
Fluency	The translation is adequately fluid, exhibiting natural expression, a range of language and phrase structures, and cohesion between sentences.	Evaluate the translation's vocabulary, syntax, and grammar to see if it is readable, fluid, and natural.
Audience Appropriateness	The translation is suitable for the intended audience's reading level, comprehension, and interests.	Evaluate whether the translation has taken into consideration the target audience's expected level of knowledge, which in this case was undergraduate students.
Logicality	The translation's procedures for reasoning and arguments are sound, understandable, and consistent.	Evaluate whether the translation adheres to the scientific text's rigor and steers clear of ambiguous or nonsensical reasoning.

Table 2: Likert scale for all assessments.

Metric	1	2	3	4	5
Accuracy	Very inaccurate	Inaccurate	Moderately accurate	Accurate	Very accurate
Fluency	Not fluent	Not very fluent	Moderately fluent	Fluent	Very fluent
Audience Appropriateness	Very inappropriate	Inappropriate	Moderately appropriate	Appropriate	Very appropriate
Logicality	Very illogical	Illogical	Moderately logical	Logical	Very logical

between the AI translation results and those of the human and rule-based machine translations, as well as a quantitative analysis and interpretation of the research data. In order to provide a higher-quality, more accurate translation of engineering textbooks, this study ultimately analyzes the benefits and drawbacks of ChatGPT-4 and Claude 3.

2.3 Data Collection

In all, 71 questionnaires were gathered for this study, with 32 male and 39 female participants. The samples comprised professors and freshman to senior engineering students at Shanghai Normal University, including 13 teachers, with almost 60% of the samples being “fourth-year university students”.

3 RESULTS & DISCUSSION

3.1 Quantitative Research Results

The scale sample in the questionnaire is dependable and trustworthy since the Cronbach alpha coefficient, which is larger than 0.9, was 0.972 when the gathered questionnaire was examined for validity and reliability in this study. The study data is very acceptable for information extraction since the KMO value was 0.858, which is more than 0.8. The research sample has strong validity and reliability.

Since the discrete degree of evaluation of Claude 3 and ChatGPT-4 is essentially the same, students and teachers of different engineering majors are in the same divergence of evaluation of these two AI tools, and there is never a situation in which the evaluation of one AI tool is more controversial. According to the

Table 3: A table comparing the advertising language ability of ChatGPT-4o and Kimi in the translation beauty industry.

Passage	Metric	AI	Mean	Standard Deviation	Mean Difference	t	p
1	Accuracy	ChatGPT-4	3.92	0.65	-0.07	-0.672	0.504
		Claude 3	3.99	0.80			
	Fluency	ChatGPT-4	3.90	0.74	-0.21	-2.200	0.031*
		Claude 3	4.11	0.67			
	Audience Appropriateness	ChatGPT-4	3.92	0.71	-0.18	-2.333	0.023*
		Claude 3	4.10	0.70			
	Logicality	ChatGPT-4	3.97	0.68	-0.04	-0.445	0.658
		Claude 3	4.01	0.80			
2	Accuracy	ChatGPT-4	3.94	0.81	-0.07	-0.698	0.488
		Claude 3	4.01	0.80			
	Fluency	ChatGPT-4	3.89	0.73	-0.15	-1.742	0.086
		Claude 3	4.04	0.78			
	Audience Appropriateness	ChatGPT-4	3.90	0.74	-0.14	-1.598	0.114
		Claude 3	4.04	0.76			
	Logicality	ChatGPT-4	3.96	0.73	-0.11	-1.526	0.132
		Claude 3	4.07	0.76			
3	Accuracy	ChatGPT-4	4.00	0.74	0.00	0.000	1.000
		Claude 3	4.00	0.72			
	Fluency	ChatGPT-4	3.90	0.78	-0.17	-2.044	0.045*
		Claude 3	4.07	0.72			
	Audience Appropriateness	ChatGPT-4	3.93	0.74	-0.08	-0.973	0.334
		Claude 3	4.01	0.73			
	Logicality	ChatGPT-4	3.96	0.75	-0.14	-1.926	0.058
		Claude 3	4.10	0.68			
4	Accuracy	ChatGPT-4	3.97	0.76	0.00	0.000	1.000
		Claude 3	3.97	0.77			
	Fluency	ChatGPT-4	3.99	0.80	0.04	0.445	0.658
		Claude 3	3.94	0.75			
	Audience Appropriateness	ChatGPT-4	3.87	0.77	-0.21	-2.154	0.035*
		Claude 3	4.08	0.71			
	Logicality	ChatGPT-4	4.03	0.74	-0.10	-1.355	0.180
		Claude 3	4.13	0.65			

* $p<0.05$ ** $p<0.01$

results of the questionnaire data survey, Claude 3 has a slightly higher average value of each index than ChatGPT-4, which is more advantageous in user evaluation. In contrast, the assessments of the four indicators do not change ($p>0.05$) when ChatGPT-4 translates four sets of distinct engineering textbook sections, and the translation capability remains constant. The stability of Claude 3's translation of the engineering textbook can still be improved, though, as one of the four groups of distinct passages in the textbook exhibits a significance at the 0.05 level of fluency with the other two passages (p is 0.038 vs. 0.022, respectively).

Furthermore, Table 3 displays the survey results of the comparison analysis of the translations of Claude 3 and ChatGPT-4. The table shows that the first translated passage, Claude 3, has better audience adaptation and fluency than ChatGPT-4; the second translated passage's results show no difference ($p>0.05$); the third translated passage, Claude 3, has better fluency than ChatGPT-4; and the fourth translated passage, Claude 3, has better audience adaptation than ChatGPT-4.

In summary, the research indicates that both ChatGPT-4 and Claude 3 perform well when translating engineering textbooks; both exhibit good accuracy and logic, but ChatGPT-4 performs worse than Claude 3 in terms of fluency and audience adaptability. Claude 3's translation stability still requires work.

3.2 Comparative Result Analysis of Four Translation Methods

After comparing and evaluating four selections translated with rule-based machine translation, human translation, ChatGPT-4 translation and Claude 3, the author came to the following conclusions.

1. Rule-based machine translation is able to recognize and translate most specialized vocabulary, but omissions and mistranslations are more problematic. For instance, “Fluid statics is used to determine the forces acting on floating or submerged bodies and the forces developed by devices like hydraulic presses and car jacks.” Rule-based machine translation can identify and translate the majority of specialized vocabulary. The forces generated by machinery such as vehicle jacks and hydraulic presses, as well as the forces acting on submerged or floating things, are determined using fluid statics. While “floating or submerged bodies” was accurately translated but just a portion of the text was translated, leaving out important details. because the sentence's use of “floating or submerged bodies” is incorrectly translated, leaving out important details. In addition, it suffers from certain translation accents, such as slightly unnatural expressions like “the so-called” and “become very important”, which make it less fluent to read.

2. Certain specialist terms, such “isentropic” and “isentropic efficiencies,” are appropriately translated using ChatGPT-4. The translation is consistent with the Chinese expression pattern, yet the sentences are also fluid, natural, and logical, such as the sentence “Electrons at outer orbits have larger kinetic energies.” The phrase “the force relations developed naturally involve the gravitational acceleration” is one example of a term that has to be better improved since it is a little stiff. The phrase “the force relations” is too simple to translate.

3. Since there is no relative motion between the fluid and the solid surface, there are no shear forces acting parallel to the surface. Claude 3's translation guarantees accuracy, fluency, and logicalness while also excelling at handling lengthy and challenging sentences. Shear forces operating parallel to the surface are absent as there is no relative motion between the fluid and the solid surface. The sentence's translation is clear and faithfully captures the original text's logical flow. However, there are also some cases where the translation is too colloquial and the meaning is vague, such as in the sentence “the variation of pressure is due only to the weight of the fluid. For example, in the sentence “the variation of pressure is due only to the weight of the fluid.”, the translation of “is due to” lacks professionalism.

4. Compared to the other three translation processes, human translations are the slowest and least effective, but they are of the greatest quality and have almost evident errors.

The aforementioned investigation results indicate that the public is aware of ChatGPT-4 and Claude 3's capacity to translate engineering texts. Although both ChatGPT-4 and Claude 3 have presentation issues, Claude 3 is superior in this area in terms of translation. In the meanwhile, ChatGPT-4 and Claude 3 continue to struggle with difficult, informal, and ambiguous language, and their overall translation quality falls short of that of human translators. Thus, according to the author, when translating engineering textbooks, translators can use the AI tool to finish the first translation and then combine it with the benefits of human translation quality to add embellishments or rewrites to further improve the translation's wording and elaboration.

4 CONCLUSION

With advancements in science and technology, translation technology also brings about innovation as times change. According to the author, textbooks from all disciplines should aim to use the most advanced translation technology, such as AI translation, when they are translated, as the academic community is now placing a strong emphasis on the sharing of global educational materials across languages. This study employed a mixed research approach to identify the advantages and disadvantages of ChatGPT-4 and Claude 3 in the translation of engineering textbooks. The advantages of these two AI tools include their ability to effectively translate scientific texts in engineering textbooks in a way that is accurate, fluent, acceptable, and logical, as an alternative to rule-based machine translation. Furthermore, in this subject, Claude 3 is better suited for translation. The drawback is that, although scientific texts are objective, engineering textbooks must be presented carefully to be appropriate for this age range in order to serve as instructional resources for students. The presentation of these two AI tools is still difficult, informal, and unclear; translators who are knowledgeable about the variations in languages, cultures, and modes of expression among nations must manually alter and add to them. In order to produce more accurate and expert translations of top-notch engineering textbooks, the author suggests that the AI tool should be able to learn the languages and cultures of other nations and continually enhance its algorithm. However, this study has many drawbacks, such as a too limited selection of engineering textbooks; more engineering textbooks in other domains may be included. In future research, more in-depth studies can be conducted to continue exploring the complementary aspects of AI translation and human translation in textbooks for different specialties in engineering. Through this project, it is envisaged that translators would become proficient in integrating AI and manual touch-ups while translating engineering textbooks, fostering collaboration and innovation in international education while also fostering cross-linguistic communication.

REFERENCES

Ahmadova, S. 2024. A comparative quality assessment of ChatGPT - 4 and human translation of scientific texts (Master's thesis, Khazar University (Azerbaijan)).

Ahrenberg, L. 2017. Comparing machine translation and human translation: A case study. In RANLP 2017: The First Workshop on Human - Informed Translation and Interpreting Technology (HiT - IT) (pp. 21 - 28). *Association for Computational Linguistics*.

Al - Romany, T. A. H., Kadhim, & M. J. 2024. Artificial Intelligence Impact on Human Translation: Legal Texts as a Case Study. *International Journal of Linguistics, Literature and Translation* 7.5:89 - 95.

Balas, M., Kaplan, A. J., Esmail, K., Saleh, S., Sharma, R. A., Yan, P., & Arjmand, P. 2024. Translating ophthalmic medical jargon with artificial intelligence: a comparative comprehension study. *Canadian Journal of Ophthalmology*.

Dictionary, O. E. 1989. Oxford English dictionary. Simpson, Ja & Weiner, Esc, 3.

Duderstadt, J. J. 2007. Engineering for a changing road, a roadmap to the future of engineering practice, research, and education.

Ferrag, F., & Bentounsi, I. 2024. A. The Use of Artificial Intelligence in Academic Translation Tasks Case Study of Chat GPT, Claude and Gemini.

Haq, M., Mushhood Ur Rehman, M., Derhab, M., Saeed, R., & Kalia, J. 2024. Bridging Language Gaps in Neurology Patient Education Through Large Language Models: a Comparative Analysis of ChatGPT, Gemini, and Claude. *medRxiv* 2024 - 09.

Li, B., Yang, P., Sun, Y., Hu, Z., & Yi, M. 2024. Advances and challenges in artificial intelligence text generation. *Frontiers of Information Technology & Electronic Engineering* 25(1): 64 - 83.

Li, Q. 2024. Bridging Languages: The Potential and Limitations of AI in Literary Translation: A Case Study of the English Translation of A Pair of Peacocks Southeast Fly. *Advances in Humanities Research* 8.1: 1 - 7.

Li, S. Q., 2024. The Role and Challenges of Generative Artificial Intelligence (AIGC) for Science and Technology Communication. *China Science and Technology Industry* (05):50-52.

Lommel, A., Gladkoff, S., Melby, A., Wright, S. E., Strandvik, I., Gasova, K. & Nenadic, G. 2024. The multi-range theory of translation quality measurement: Mqm scoring models and statistical quality control. *arXiv preprint arXiv* 2405.16969.

Mohsen, M. 2024. Artificial Intelligence in Academic Translation: A Comparative Study of Large Language Models and Google Translate. *PSYCHOLINGUISTICS* 35(2): 134 - 156.

Quigley, C., Oliviera, A. W., Curry, A., & Buck, G. 2011. Issues and techniques in translating scientific terms from English to Khmer for a university-level text in Cambodia. *Language, Culture and Curriculum* 24(2): 159 - 177.

Shokoohifar, M. 2024. Artificial Intelligence-Assisted Translation: A Study of Cognitive Load and Time Through EEG (Doctoral dissertation, Allameh Tabataba'i University).

Wang, L. 2023. The Impacts and Challenges of Artificial Intelligence Translation Tool on Translation Professionals. *SHS Web of Conferences* 163.

Wang, Y. 2023. Artificial Intelligence Technologies in College English Translation Teaching. *Journal of Psycholinguistic Research*.

Yan, J., Yan, P., Chen, Y., Li, J., Zhu, X., & Zhang, Y. 2024. Benchmarking GPT - 4 against Human Translators: A Comprehensive Evaluation Across Languages, Domains, and Expertise Levels. *arXiv preprint arXiv*:2411.13775.

Yan Y. Q., & Zhao Y., 2023. Research on the Application of AIGC in Traditional Artificial Intelligence Systems. *Guangdong Communication Technology* (12):14 - 17+68.

Zhang, H. B., & Li, Y. S., 2009. A study on the development status of educational resource sharing environment and sharing mechanism. *China Electronic Education* (11):68-73.

Zhao, L. H., 2022. The Relationship between Machine Translation and Human Translation under the Influence of Artificial Intelligence Machine Translation. *Mobile Information Systems* (1): 9121636.