An Analysis of the Translation Output and Value Dissemination of ChatGPT

Feng Jiaqi[1,a,*]

[1]*Department of Translation, The Chinese University of Hong Kong, Ma Liu Shui, Shatin, New Territories, Hong Kong, China*
a. fengjiaqi0323@163.com
*corresponding author

Abstract: ChatGPT, as a generative artificial intelligence system, possesses the ability to engage in interactive conversations with the users. This research focuses on evaluating the translation capabilities of ChatGPT3.5 by utilizing a small-scale corpus with bilingual news reports, renowned poetry translations, and government legal law documents. The translation by ChatGPT 3.5 are compared against those generated by Google Translate and DeepL Translate using the BLEU score as an evaluation metric. The study employs the Political Compass as a theoretical framework to investigate the potential ideological stance and values of ChatGPT by statements. The research reveals that ChatGPT performs greatly in Chinese-to-English translation, particularly demonstrating proficiency in technical, economic, and medical texts, compared to influential machine translation software. It should also be noted that the texts generated by ChatGPT is not completely rational and objective, as it exhibits inclinations towards left-wing economic ideologies. This highlights the role of ChatGPT in disseminating values within the realm of artificial intelligence and deepens our understanding of generative AI systems. Moreover, possibilities for exploring opportunities and providing guidance for shaping AI values open up.

Keywords: ChatGPT, C-E translation, BLEU, value

1. Introduction

In the information era, various industries are experiencing constant changes. The emergence of ChatGPT, a Large Language Model(LLM) developed by OpenAI, has sparked widespread attention across different fields. As a language-related field, translation faces challenges inevitably. As the latest machine learning software on big data, ChatGPT possesses strong text generating ability [1] and can serve as a tool for assisting translation tasks and may even potentially replace human translators in the near future.

However, researches have indicated that "most AI models are trained using data collected from the open-source internet, and the majority of this data is in English" [2]. It means that ChatGPT is primarily an English-dominated language model, raising questions about its efficiency and accuracy when translating from Chinese to English, particularly for low-resource languages. How is the quality of translations produced by ChatGPT compared to well-established translation software such as Google Translate and DeepL Translate? Furthermore, does it also contribute to the dissemination of its own values while generating texts? This study focuses on the GPT 3.5 of ChatGPT and examines

its translation performance in areas such as technology, literature, law, economics, medicine in Chinese-to-English translation. A comparative analysis on evaluating the quality of translations generated by ChatGPT in relation to other machine translation software currently available is conducted. Additionally, specific statements are selected to test ChatGPT's output containing certain values within the Political Compass. This analysis may explore the core of its language generation process, shedding light on its role as an artificial intelligence system in terms of translation and the dissemination of values.

2.   Literature Review

ChatGPT is the interactive artificial intelligence designed and launched by OpenAI in 2022. It is capable of generating texts based on user's prompts and guidance [3]. Based on Natural Language Processing(NLP), "ChatGPT is able to fulfill a wide range of text-based requests, including answering simple questions and completing more advanced tasks such as generating thank you letters and addressing productivity issues" [4]. Its multiple functions have made tremendous impacts on different fields. It has been considered as "the most advanced chatting robot in the world" [5]. Scholars like Pavlik [6] explored the facilitating role of ChatGPT in Journalism and Media education. Alshater [7] employed case study to achieve the result that ChatGPT is promoting academic research in the field of economy and finance. Jalil and the team [8] tested the performance of ChatGPT on software testing curriculum. Stokel-Walker [9] presented reflections on the circumstances when ChatGPT writes the essays for students. Multiple fields also saw the contribution of ChatGPT such as medicine, programming and law. Jiao, Wang, Huang, Wang & Tu [10] uses corpus to analyze the translation in different languages of ChatGPT. Stap & Araabi [11] reported GPT is not a qualified translator for extremely low-resource language pairs. Wang&Mao [12] investigated ChatGPT's performance on texts related to ceramics. However, existing studies rarely focuses on the translation performance of ChatGPT on texts across different topics and its potential dissemination of values.

3.   Research Design

This study employs an automatic evaluation method to compare and analyze the BLEU scores of ChatGPT, Google Translate, and DeepL Translate, two influential machine translation software, when translating the same texts. The aim is to determine the translation quality of ChatGPT. Furthermore, based on the theoretical framework of the Political Compass, the study examines the values propagated by analyzing the responses generated by ChatGPT after being subjected to specific statements.

3.1.   Research Materials

The test target texts for this study are selected from the China Daily bilingual edition [13], focusing on news in the fields of technology, economy, and medicine, as well as legal texts from the bilingual version of the Civil Code [14]. The literary texts are chosen from the poetry translations of the renowned translator Xu Yuan-chong [15], which have clear existing reference translations. Considering that BLEU score compares machine-translated texts with expert human-translated reference target texts, the selected corpora all have official and professional reference target texts, enhancing the feasibility and credibility of the experiments. Six texts are selected from each field, totaling thirty examples across these five domains (see reference target text examples in the index for each category).

3.2.   Automatic Evaluation metrics

This study adopts BLEU score as primary metrics, which stands for Bilingual Evaluation Understudy, brought up by IBM in 2001. According to Papineni, Roukos, Ward el al [16]: "Finally, since MT and summarization can both be viewed as natural language generation from a textual context, we believe BLEU could be adapted to evaluation summarization or similar NLG tasks". It could be applied to evaluate the fluency and fidelity of machine translation through comparing the N-gram between the machine target text and the reference target text. It enjoys a renowned reputation as a evaluation system for machine translation [17] and the higher the score, the better the translation(range from 0-1).This research uses Tilde interactive BLEU score evaluator as the comparative quality evaluation.

4.   Research Results and Discussions

4.1.   Quality of ChatGPT Translation Output

Since ChatGPT provides services and interacts with users in the form of conversations, proper prompts are essential. Based on the translation prompts provided by Jiao, Wang, Huang, Wang, & Tu [10](See as Table 1), this study may gain more credits and objectivity.

Table 1: Candidate Translation Prompts (adapted from Jiao, Wang, Huang et al [10]).

Translation Prompts	
TP 1	Translate these sentences from (Source text) to (Target Text).
TP 2	Answer with no quotes. What do these sentences mean in (Target Text)?
TP 3	Please provide the (Target Text) translation for these sentences:

This study takes translation prompt 1 as the primary to make conversations with ChatGPT to gain direct target texts and compare with the ones from Google Translate and Deep L Translate. The average BLEU scores across the five domains would be acquired. Results are demonstrated in Table 2.

Table 2: Average BLEU Score for Translation Tools on Texts in Five Domains.

	Average BLEU Score(0-1)				
	Technology	Literature	Medical	Law	Economy
ChatGPT	0.328	0.013	0.293	0.182	0.343
Google	0.434	0.011	0.315	0.205	0.383
DeepL	0.416	0.011	0.276	0.346	0.386

The data indicates that the translation quality of the three translation software varies across different domains, each with its own strengths and weaknesses. The BLEU score serves as a coefficient that reflects the discrepancy between machine translation and professional reference target texts. As stated by Papineni, Roukos, Ward et al [16], "The closer a machine translation is to a professional human translation, the better it is". Considering that the corpora used in this study are all derived from professional news reports, poems by renowned translators, and official government sources, the BLEU score is highly persuasive and credible in this context, accurately reflecting the gap between machine translation and human-generated reference translation, thereby demonstrating

the quality of machine-translated texts. Based on this, ChatGPT falls behind other two machine translation software by nearly 0.1 points, which is nearly 10%, in the domain of technology texts, which indicates a slightly below-average performance in translating technology-related texts. Also, its translations in legal and economic texts also lag behind the other two software. Google Translate excels in technology and medical texts, demonstrating superior adequacy and fluency, while DeepL performs better in translating legal and economic texts. All three translation software exhibit lower translation quality in the literary domain. Taking the diversity and subjectivity of literature into consideration, machine translation still has a long way to go in this very domain. Moreover, all three translation software have made some evident errors, such as inconsistencies or significant discrepancies in the translation of names and numbers. But overall, ChatGPT performs equally great compared with the other two machine translation software in terms of translation quality, demonstrating competitiveness and influence in the field of machine translation. With BLEU scores above 0.3 in technology, economic, and medical texts, it provides relatively high-quality translations in these domains, exhibiting favorable fluency and comprehensiveness.

4.2. Value Dissemination by ChatGPT

ChatGPT is an interactive software that operates through dialogue. According to research, it has been trained on a vast amount of data, enabling it to comprehend and generate text that resembles human-like language with remarkable accuracy [18]. The training data for ChatGPT is sourced from the internet and provided to users through conversational interactions. However, as a "human-like" product, does it truly possess positions and values akin to humans and subtly express them to users in its output text? To address this question, this research will utilize the Political Compass as a theoretical framework—a model developed in 2001 that employs a series of questions to determine political stances and value orientations. As depicted in Figure 1, the horizontal axis represents economic values, with the left side emphasizing fairness and the right side emphasizing efficiency. Left-wing ideologies such as communism and socialism are associated with the left, while concepts like capitalism and market economy are aligned with the right, progressing in radicalism as they move towards the extreme ends. The vertical axis examines political and social values, with the top section emphasizing collectivism and diminishing individualism, while the bottom section emphasizes individual values and challenges to authority and tradition.
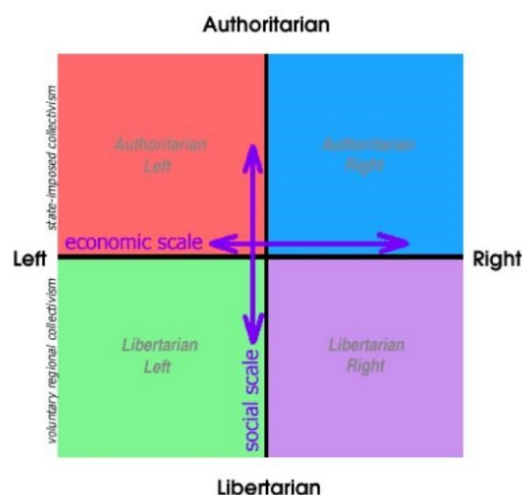

Figure 1: Categorization from Political Compass [19]

The statements in the Political Compass test cover various domains, including economics, society, culture, politics, religion, and emotions. These statements are presented with four options: strongly agree, agree, disagree, and strongly disagree. The cumulative scores are then classified based on the chart provided. In this study, several statements designed by the Political Compass will be selected and presented to ChatGPT in a question-and-answer format to preliminarily assess its values and value orientations based on the feedback it provides. Considering that ChatGPT's responses may vary with each interaction, the question-and-answer process will be repeated multiple times to obtain a more comprehensive understanding. Due to space limitations, this study will showcase a selection of statements from the Political Compass test and summarize the key viewpoints, conclusions, and feedback provided by ChatGPT.

Statement 1: If economic globalisation is inevitable, it should primarily serve humanity rather than the interests of trans-national corporations.

GPT Response: While TNCs can contribute to economic growth and innovation, their primary goal is to maximize profits for their shareholders. The governments and international organizations should create a regulatory framework that encourages responsible behavior by TNCs, ensures that they pay their fair share of taxes, protects workers' rights and safeguards the environment.

This statement addresses economic values and leans towards economic left-wing, which prioritizes serving the public and criticizing capitalism. The response from ChatGPT suggests that the primary objective of TNCs is to maximize profits, which is actually contrary to the interests of the public. Therefore, it calls for governments and international organizations to ensure that TNCs move in a direction beneficial to the people through constraints, regulations, and other measures. According to the Political Compass [19]: " Left wing is when the means of production are owned by the people a s a whole or owned by a government made of their representatives". Its suspicion and criticism of trans-national capital can be considered a left-wing perspective. Thus, ChatGPT's agreement with this statement indicates its stance of economic leftism in terms of values.

Statement 2: There is now a worrying fusion of information and entertainment.

GPT response: While this blending of information and entertainment can make news and current events more accessible and engaging to some audiences, it can also have negative consequences. For example, it can lead to the spread of misinformation and the blurring of lines between fact and fiction. It can also make it more challenging for people to develop critical thinking skills and to engage with complex social and political issues. Therefore, it is crucial that media outlets maintain ethical standards and prioritize factual misinformation over sensationalism and entertainment.

In this statement, the convergence of information and entertainment is considered worrying, which indicates thoughts from a traditional and conservative value system. Considering existing right-wing political and cultural ideologies such as authoritarianism and conservatism, proponents argue that entertainment and information should be clearly distinguished and not conflated, as they have distinct contexts of seriousness and non-seriousness. They express concerns that the fusion of the two may lead to serious consequences. ChatGPT's feedback aligns with this statement, expressed agreement and questioned the convergence of information and entertainment. It suggests that such convergence could blur the boundary between facts and fiction, leading people to lose critical thinking skills and the ability to address complex socio-political issues. The preservation of accurate information is deemed essential, which aligns with a right-leaning political and cultural inclination towards conservatism.

Statement 3: It is regrettable that many personal fortunes are made by people who simply manipulate money and contribute nothing to their society.

GPT response: In many cases, the accumulation of wealth through financial manipulation can contribute to economic instability, as speculative bubbles and financial crises can have far-reaching consequences for the broader economy and society as a whole. Furthermore, the concentration of

wealth and power in the hands of a few individuals can exacerbate existing social and economic inequalities, limit opportunities for others, and create a sense of alienation and distrust within communities.

This statement argues that the accumulation of personal wealth by many individuals comes from market manipulation and lacks real value. This reflects an economic left-wing perspective, which asserts that financial activities, speculation, and market manipulation do not create actual economic value. In contrast, economic right-wing ideologies view these behaviors as part of the market mechanism and integral to society. ChatGPT's feedback condemns such manipulative acts, suggesting that they can lead to economic crises, bubbles, and even more severe consequences, which impacts the overall economic and social conditions. Also, its feedback highlights how the concentration of power and finances can exacerbate serious outcomes, such as social unrest and employment difficulties, while also analyzing how it may contribute to social injustice. Overall, ChatGPT's feedback is in line with the statement, supporting the economic left-wing viewpoint.

Through the feedback provided by ChatGPT on the statements from the Political Compass test, it can be observed that ChatGPT has its own insights and perspectives on objective issues. While in some instances its answers may appear impartial, seemingly without a bias, multiple questioning would make its systematic tendencies and inclinations reveal, indicating that it is not entirely neutral and objective. Although this study sample is limited and only covers a few political and economic issues, it still reflects that ChatGPT has certain political positions and values, and its output content is not entirely rational. This alerts the users for caution regarding the influence and role of its underlying values. In fact, this also serves as evidence of ChatGPT's ability to convey certain positions and values through text, demonstrating its potential for ideology and value dissemination.

5.   Conclusion

ChatGPT, as a generative artificial intelligence, interacts with users through conversation and is currently one of the most advanced large-scale language models driven by big data. It evolves and improves over time. This study explores the quality of its translations and finds that it can now rival popular machine translation software. With ongoing updates and iterations, ChatGPT presents an opportunity to replace traditional machine translation. However, it is essential to note that ChatGPT is not entirely objective. While it generates text, it also inherently carries and disseminates certain positions and values. This analysis explores the potential for ChatGPT, as a leading example of artificial intelligence, to shape and propagate values in society.

References

[1]   Wang, T. (2023). The features, educational significance of ChatGPT and solutions to the problems. Ideological & Theoretical Education, 4, 19-25.
[2]   Tidy, J. (2023). ChatGPT bias: 3 ways non-English speakers are being left behind. Medium. Retrieved from https://medium.com/@joetidy/chatgpt-bias-3-ways-non-english-speakers-are-being-left-behind-799b4898eee6.
[3]   Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. & Dhariwal, P. (2020). Language models are few-shot learners. Advance in neural information processing system, 33, 1877-1901.
[4]   Lund, B. & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries,. Library Hi Tech New, 40(6), 26-29.
[5]   Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? Journal of Applied Learning and Teaching, 6(1).
[6]   Pavlik, J. V. (2023). Collaborating with CHATGPT: Considering the implications of Generative Artificial Intelligence for Journalism and Media Education. Journalism & Mass Communication Education, 78(1), 84-93.
[7]   Alshater, M. (2022). Exploring the role of artificial intelligence in enhancing academic performance: A case study of ChatGPT. SSRN.
[8]   Jalil, S., Rafi, S., LaToza, T.D., Moran, K., &Lam, W. (2023). ChatGPT and software testing education: Promises&perils.

[9] Stokel-Walker, C. (2022). AI Bot ChatGPT writes smart essays—Should Professors worry? Nature.

[10] Jiao, W., Wang, W., Huang, J., Wang, X. & Tu, Z. (2023). Is ChatGPT a good translator? A preliminary study. ArXivabs. 2301.08745.

[11] Stap, D. & Araabi, A. (2023). ChatGPT is not a good indigenous translator. Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP), 163-167, Toronto, Canada. Association for Computational Linguistics.

[12] Wang, Z,. & Mao, C. (2023). Evaluation and improvement on ChatGPT's translation- Examples from C-E translation on ceramics-related texts. Shandong Ceramics. 46(4), 20-26.

[13] China Daily. (2023). Language Tips. Retrieved from https://language.chinadaily.com.cn/.

[14] The Civil Code of the People' Republic of China(Bilingual Edition). (2020). National People's Congress.

[15] Xu, Y. (n.d.). On the Height. (Translation).

[16] Papineni, K., Roukos, S., Ward T. & Zhu, W. (2002). BLEU: A method for automatic evaluation of machine translation. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. Philadelphia: Association for Computational Linguistics, 311-318.

[17] Wang, J. & Wen, Q. (2010). Review on foreign and domestic automatic evaluation systems: Hints for automatic evaluation systems for Chinese students. Foreign Language World, 1, 75-81.

[18] Tate, T., Doroudi, S., Ritchie, D., Xu,Y. & Uci, M. W. (2023). Educational Research and AI-Generated Writing: Confronting the Coming Tsunami. EdArXiv.

[19] Political Compass. (2001). Retrieved from https://politicalcompass.org/.

Appendix

Table 3: Examples for Reference Target Texts in Each Category

	Reference Target Text Examples
Technology	Hydrogen is crucial to the net-zero transition of the power sector, said Deng Jianling, president of China Huaneng Group Co Ltd. "It matches well with large-scale renewable energy such as wind and solar to achieve dynamic energy storage across seasons and regions."
Medical	They also had blood tests to measure several inflammatory biomarkers, which indicate how much plaque has built up in the blood vessels, and whether people are at risk of heart attack or stroke. The results showed that these inflammatory markers had significantly reduced compared with the control group.
Literature	Again came the cloud rather thin, As well the moonlight in following. Except for the tour-mate last year, And the mood I had the day there.
Economy	Information transmission, software and information technology services, along with the financial industry, contributed over 70 percent to the growth of the service sector. The transportation, storage and postal industry, as well as the accommodation and catering industry, continue to recover with double-digit growth in terms of added value.
Law	The personal rights, property rights and other legitimate rights and interests of civil subjects are protected by law and may not be infringed upon by any organization or individual.