Evaluation of Machine Translation Systems:

A Literature Review on ChatGPT

and Google Translate

Ohod Faisal Ahmed[1], Ida Kusuma Dewi[2], Mohammad Yunus Anis[3]

[1]Linguistics Department, Faculty of Cultural Science, Sebelas Maret University

[2]Translation Studies and Linguistics Department, Faculty of Cultural Science, Sebelas Maret University

[3]Sastra Arab and Translation Department, Faculty of Cultural Science, Sebelas Maret University

E-mail: hoodalarhabi73@gmail.com[1], ida.k.d@staff.uns.ac.id[2], yunus_678@staff.uns.ac.id[3]

Abstract

This is a literature review discussing 15 selected papers about ChatGPT and Google Translate study results based on keyword analysis and publication year. We applied descriptive data analysis technique to analyze the data. We selected Studies on translation performance of natural language processing tools were chosen due to their increasing prominence and diverse applications, ranging from literary to technical translations. The data for this literature review was retrieved from Scopus and Google Scholar. The search was limited to the last five years to ensure the inclusion of recent advancements, particularly those reflecting improvements in ChatGPT's GPT-4 engine and updates in Google Translate's neural machine translation capabilities. The results showed that ChatGPT excels in fluency and contextual understanding, particularly in literary and poetic translations, outperforming Google Translate in maintaining stylistic elements and complex language structures. Both systems demonstrated strengths in specialized translations, with ChatGPT showing notable proficiency in medical literature and technical texts. However, challenges remained in low-resource languages and specialized domains, requiring further training and development. Despite technological advancements, human translators are essential for achieving culturally nuanced translations. This study has some implications for future implementing for enhancement contextual understanding, improving accuracy for low-resource languages, and addressing specific error patterns through ongoing research and collaborative efforts between human translators and

machine translation tools. These recommendations aim to optimize the performance of ChatGPT and Google Translate, thereby ensuring more accurate and contextually appropriate translations across various fields.

Keywords: *Machine Translation, ChatGPT, Google Translate, Comparative Analysis, Fluency*

Introduction

Rapid advancements in natural language processing have led to the development of powerful machine translation systems that can facilitate cross-language communication and collaboration. Two prominent examples in this domain are ChatGPT and Google Translate, which have garnered significant attention for their language understanding and generation capabilities (Hariri, 2023). ChatGPT, developed by OpenAI, is a large language model that combines the power of pre-trained deep learning models with a programmability layer, enabling it to engage in natural language conversations (Gabashvili, 2023)

The adoption rates of AI-powered language tools such as ChatGPT and Google Translate have seen significant growth in recent years, reflecting a broader trend towards the integration of artificial intelligence in everyday communication. According to research by (Karpov et al., 2008), ChatGPT has rapidly gained popularity since its launch, with millions of users engaging with the platform for various applications, including education, content creation, and customer support. The platform's accessibility and versatility contribute to its widespread use among individuals and businesses alike. In contrast, Google Translate has been a staple in the realm of language translation since its inception in 2006, boasting over 500 million users daily as reported by (Translation, 2003).

This extensive user base highlights the tool's effectiveness and reliability, making it an essential resource for individuals seeking quick translations across multiple languages. The evolution of translation tools has seen significant advancements over the past few decades, transforming how individuals and businesses communicate across language barriers. In the early 2000s, rule-based machine translation systems dominated, relying on predefined linguistic rules to translate text (Sennrich et al., 2016). The introduction of statistical machine translation in the mid-2000s marked a turning point, with Google Translate launching in 2006, utilizing vast amounts of bilingual data to improve accuracy (Maes et al., 2022). As the decade progressed, neural machine translation emerged around 2014, significantly enhancing the quality of translations by using deep learning techniques to understand context better (Sennrich et al., 2016).

ChatGPT, introduced in 2020, further revolutionized the landscape by incorporating advanced natural language processing capabilities, allowing for more nuanced and conversational interactions compared to traditional translation tools (Bahrini et al., 2023). Comparing Google Translate and ChatGPT is particularly relevant as both utilize cutting-edge technology but serve distinct purposes; while Google Translate focuses primarily on text translation, ChatGPT excels in generating human-like responses and engaging in dialogue, making it a powerful tool for applications beyond mere translation, such as content creation and customer support (Fang et al., 2023).

The evolution of these technologies demonstrates the increasing significance of AI in communication since it not only overcomes language barriers but also improves user experience by providing context-aware responses and personalized interactions. This change in communication technology represents a move toward more user-friendly and participatory platforms where users can anticipate meaningful conversations that are tailored to their individual requirements and interests in addition to accurate translations (Green et al., 2015).

Researchers have explored the potential of ChatGPT as a translation tool, noting its ability to understand the nuances of different languages and provide context-specific translations (Hariri, 2023). Similarly, Google Translate, a widely used online translation service, has also been the subject of extensive research. Comparative studies have been conducted to evaluate the performance of ChatGPT against Google Translate, as well as other leading translation systems like DeepL and Tencent (Hariri, 2023). This paper assessed the efficacy of ChatGPT in machine translation jobs, specifically utilizing GPT-4, particularly when powered by GPT-4.

It provides a comprehensive analysis of ChatGPT's ability to handle various translation challenges, including contextual accuracy, fluency, and cultural nuances. The authors benchmark ChatGPT against other leading systems, demonstrating its competitive edge in high-resource languages while highlighting areas for improvement in low-resource and complex translations. Their findings underscore GPT-4's advancements in contextual understanding and fidelity, making it a valuable tool for diverse translation tasks (Jiao, Wang, et al., 2023).

Moreover, the Parrot framework which fine-tunes large language models like ChatGPT with human translation feedback to enhance real-time translation capabilities. The authors explore how integrating human corrections into training improves translation quality and adaptability across domains. By employing this iterative feedback mechanism, the study demonstrates significant improvements in accuracy and contextual alignment, particularly for domain-specific and conversational translations (Jiao, Huang, et al., 2023).

From all of the previous studies, ChatGPT and Google Translate have both achieved notable advancements in terms of machine translation quality and accuracy (Y. Gao et al., 2023). Particularly ChatGPT has demonstrated outstanding performance in translating code between programming languages, which may have

significant effects on teamwork and software development (Hariri, 2023). Both ChatGPT and Google Translate, have gained widespread attention and adoption (Hendy et al., 2023). A comprehensive literature review on the evaluation of these systems is crucial to understanding their strengths, limitations, and potential applications. One of the key aspects of machine translation is the ability to understand the nuances of different languages and provide context-specific translations (Sepesy Maučec & Donaj, 2020).

Furthermore, ChatGPT, a sizable language model created by OpenAI, performs better in some situations than conventional machine translation systems (Gabashvili, 2023). According to research evaluating ChatGPT, Google, DeepL, and Tencent's translation systems, it performed exceptionally well in terms of translation accuracy and fluency (Jiao, Wang, et al., 2023). In addition to language translation, ChatGPT and GPT-4 have improvement   in code translation between programming languages, allowing users to convert code snippets from one language to another (Liu et al., 2023).

 With the launch of GPT-4, the translation performance of ChatGPT has been significantly improved, becoming comparable to commercial translation products, even for distant language pairs (Siu, 2023). While ChatGPT and GPT-4 have exceptional language generation capabilities, which do not possess the same level of understanding, empathy, and creativity as humans, and cannot fully replaced human translators in most translations.   Interestingly, studies have found that ChatGPT can provide context-specific and nuanced translations, showcasing its ability to understand the complexities of different languages (Peng et al., 2023). In contrast, Google Translate, a widely used machine translation service, has long been a dominant player. Recent research has delved into the comparative analysis of ChatGPT and Google Translate, highlighting the strengths and limitations of each system (Hendy et al., 2023).

As the field of natural language processing continues to evolve, the incorporation of large language models like ChatGPT and Google Translate into translation workflows has become a topic of growing interest (Siu, 2023). Scholars have emphasized the potential benefits these platforms offer to language professionals, while also underscoring the ongoing need for human expertise in the translation industry (Siu, 2023). Scholars have emphasized the potential benefits these platforms offer to language professionals, while also underscoring the ongoing need for human expertise in the translation industry (Khan et al., 2023). They have emphasized the potential benefits these platforms offer to language professionals, while also underscoring the ongoing need for human expertise in the translation industry (Hendy et al., 2023).

For example, while machine translation systems excel in technical domains, their limitations in low-resource languages and complex texts reaffirm the indispensable role of human translators in ensuring quality and contextual fidelity (Nila et al., 2017). According to the authors of a thorough study, ChatGPT may be used to create language translation systems with remarkable accuracy that can comprehend the subtleties of various languages and provide context appropriate translations. This can significantly improve communication between individuals from diverse cultural and linguistic backgrounds (Akula et al., 2020). Furthermore, the potential of ChatGPT extends beyond text-based translation, as it can also be utilized for code translation between programming languages such as Java and Python (Cerf, 2023).

According to the authors of a thorough study, ChatGPT may be used to create language translation systems with remarkable accuracy that can comprehend the subtleties of various languages and provide context appropriate translations. This can significantly improve communication between individuals from diverse cultural and linguistic backgrounds (Akula et al., 2020). Furthermore, the potential of ChatGPT extends beyond text-based translation, as it can also be utilized for code translation between programming languages such as Java and Python (Cerf, 2023).

Apart from this, Document-Level Machine Translation with Large Language Models —revealed that ChatGPT performance on document-level translation tasks are often more challenging than sentence-level translation (Wang et al., 2023). This suggests that, despite its impressive capabilities in maintaining fluency and contextual understanding, ChatGPT may require further improvement to cope with larger and more cohesive textual units effectively. Additionally, a number of studies have illustrated the advantages and disadvantages of machine translation tools, including Google Translate and ChatGPT. For instance, (Peng et al., 2023) highlighted ChatGPT's ability to provide context-specific translations, particularly in high-resource languages, and its capability to perform code translations between programming languages, which makes it a valuable tool for teamwork and software development.

Meanwhile, Google Translate remains a reliable tool for everyday use due to its extensive language support and fast processing capabilities (Peng et al., 2023). However, both systems face challenges in handling low-resource languages and complex syntactic structures, reaffirming the need for human expertise in translation workflows (Bonyadi, 2020). Despite the substantial advancements in machine translation, existing studies on ChatGPT and Google Translate reveal notable gaps that warrant further exploration. Many prior studies focus on evaluating these tools in specific contexts, such as high-resource languages or technical translations, yet fail to provide a comprehensive analysis across diverse domains and text types, including low-resource languages, literary texts, and document-level translations (R. Gao et al., 2024).

This literature review aims to address these gaps by offering a rigorous comparative analysis of ChatGPT and Google Translate. By synthesizing insights from recent studies, it explores their strengths and limitations across various text types and translation challenges. The significance of this study lies in its ability to inform language professionals, researchers, and developers about the current state of machine translation technologies, guiding future improvements and fostering collaboration between human expertise and artificial intelligence. This comprehensive approach not only enhances our understanding of these tools but also contributes to optimizing their performance for diverse linguistic and cultural needs.
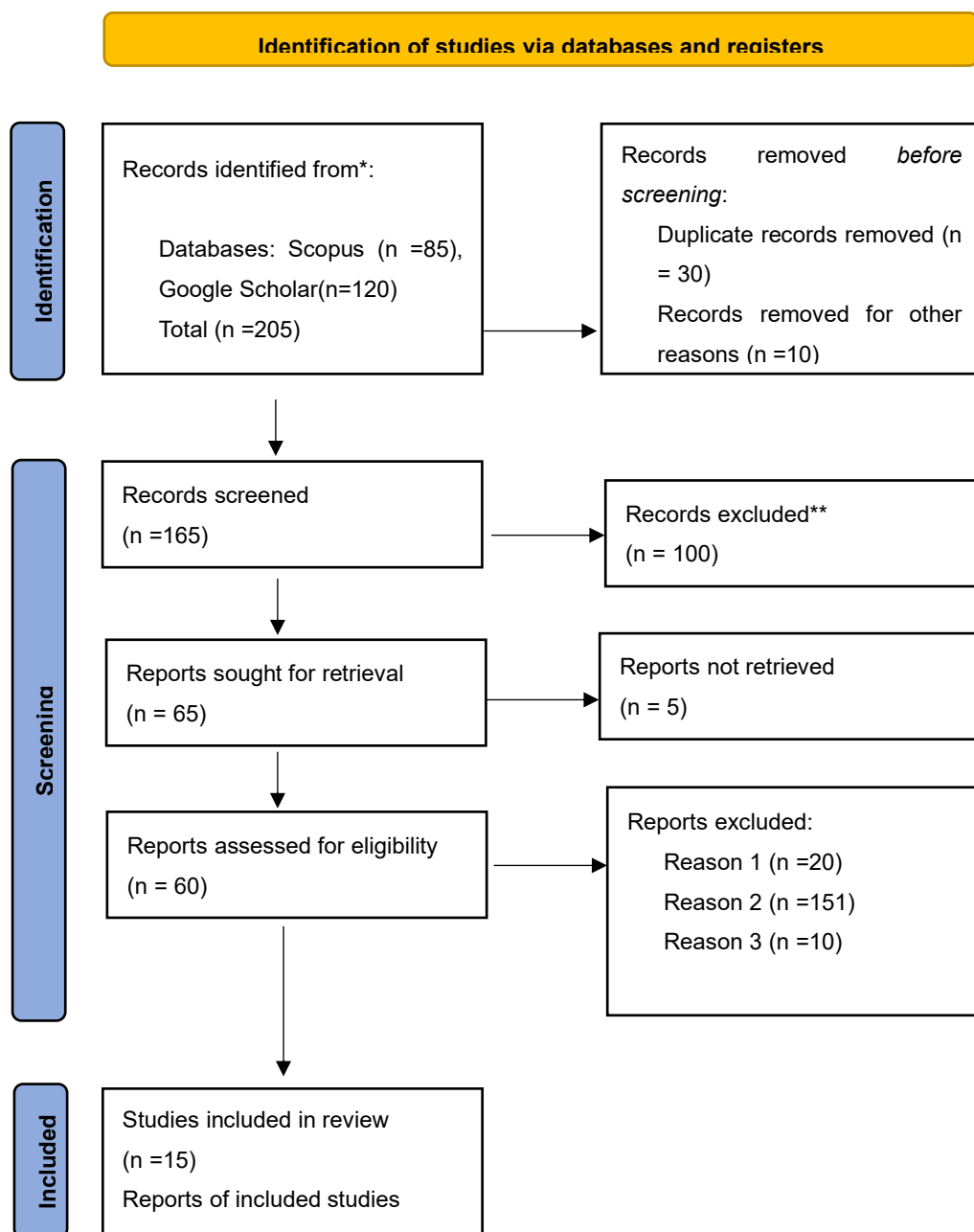
Method

This is a literature review study which primary benefits allows researchers to familiarize themselves with the vocabulary, theories, key variables, and research methods used in their field of study (Ali & Pandya, 2021). Moreover, the literature review helps the researcher understand the influential researchers and research groups in the field, which can inform the direction and focus of their own research (Deb et al., 2019). This study employed a literature review methodology, designed to synthesize existing research on the comparative analysis of ChatGPT and Google Translate. Data for this review was retrieved from reputable academic databases, including Scopus and Google Scholar two widely recognized academic databases known for their comprehensive indexing of peer-reviewed journals. It used Keywords such as "ChatGPT translation performance," "Google Translate accuracy," "comparative analysis of machine translation," and "neural machine translation systems" were used to identify relevant studies.

The search was conducted manually, focusing on articles published in the past five years to ensure the inclusion of recent advancements in machine translation technologies. This selection aimed to capture a balanced view of both systems' strengths and limitations while addressing critical gaps in previous research. The selection criteria were based on the studies' relevance to ChatGPT and Google Translate, specifically addressing their performance in translation tasks across various text types and domains, including their performance in literary texts, technical documents, and low-resource languages. The scope was defined to include studies examining the systems' translation techniques, contextual understanding, and challenges in specialized domains through comparing the performance of ChatGPT and Google Translate across different contexts and contrasting their strengths and weaknesses in handling diverse text types.

This literature review was managed using offline tools such as Mendeley, enabling efficient organization of references and notes. Key topics related to ChatGPT, Google Translate, and neural machine translation were categorized for systematic analysis. Each study was carefully reviewed to benchmark insights against other literature reviews. Observations focused on translation accuracy, fluency, contextual understanding, and adaptability across various languages and text types. The findings were synthesized into a cohesive narrative that integrates evidence, contrasts perspectives, and highlights research gaps. This approach ensured that the review adhered to a rigorous and transparent methodology.

PRISMA flow diagram to show the study selection process:

Identification of studies via databases and registers

Identification

Records identified from*:

Databases: Scopus (n =85), Google Scholar(n=120)
Total (n =205)

Records removed *before screening*:
Duplicate records removed (n = 30)
Records removed for other reasons (n =10)

Screening

Records screened
(n =165)

Records excluded**
(n = 100)

Reports sought for retrieval
(n = 65)

Reports not retrieved
(n = 5)

Reports assessed for eligibility
(n = 60)

Reports excluded:
Reason 1 (n =20)
Reason 2 (n =151)
Reason 3 (n =10)

Included

Studies included in review
(n =15)
Reports of included studies

To ensure a rigorous selection process, the retrieved studies underwent a multi-step screening. First, duplicate records were removed (n = 30), followed by a title and abstract screening to filter out studies that did not focus on ChatGPT or Google Translate in translation tasks. The remaining 165 articles were then assessed based on full-text eligibility, prioritizing those that examined translation techniques, contextual understanding, and challenges in specialized domains.

A standardized quality assessment framework was applied to evaluate the methodological rigor of the selected studies through focusing on
relevance to ChatGPT and Google Translate in translation tasks and methodological transparency and robustness, in addition to the inclusion of empirical data or comparative evaluation and consideration of contextual accuracy, fluency, and readability. Out of 165 screened studies, 100 were excluded based on these criteria. The full-text assessment included 60 studies, of which 45 were excluded due to lack of focus on ChatGPT or Google Translate (n = 20), absence of empirical evaluation (n = 15) and poor methodological rigor (n = 10). The disagreements in paper selection were resolved through discussion among the research team, ensuring the final inclusion of 15 studies that met the defined quality standards.

Results

Below is the range of topics related to ChatGPT and Google Translate studies identifiable from the keywords or variables in the research approach taken from the cited database sources. The data source consists of the conclusions of several chosen papers below in Table 1.

The research findings and outcomes are shown in Table 1.

No.	Author	Techniques, Methods and Objects	Result
1.	(Stevanović & Radičević, 2020) Comparative Analysis of Machine Translation Systems	The methodology employed is comparative analysis structured approach to critically assess and compare different machine translation on literal, technical and legal texts.	The results emphasized the complexity of machine translation systems, the importance of selecting the appropriate system for specific tasks, and the ongoing need for research to improve translation quality and evaluation methods. The findings contributed to a deeper understanding of the current state of machine translation technology and its potential future developments.
2.	(R. Gao et al., 2024)	It employed	ChatGPT outperformed both

	Machine translation of Chinese classical poetry: a comparison among ChatGPT, Google Translate, and DeepL Translator	comparative analysis as a primary technique to Chinese classical poetry applying the quantitative approach	Google Translate and DeepL Translator in all evaluation criteria, which included fidelity, fluency, language style, and machine translation style. This indicated that ChatGPT is particularly effective in translating Chinese classical poetry compared to traditional machine translation systems
3.	(GOOD, 2015)_ Large Language Models as Computational Linguistics Tools: A Comparative Analysis of ChatGPT and Google Machine Translations	It employed a comparative analysis technique to speeches delivered by King Abdullah II of Jordan, which are available in both Arabic and English.	It evaluated the effectiveness of Large Language Models (LLMs), focusing on ChatGPT and Google Translate for translating speeches by King Abdullah II of Jordan in Arabic and English at international events in 2023. Google Translate's translations were found to be deficient, requiring major revisions due to contextual accuracy and meaning issues. In contrast, ChatGPT's translations were rated as acceptable with minor edits, offering more natural-sounding translations.
4.	(Abdulmohsen Alosaimi & Abdulaziz Alawad, 2024)Evaluation of the Translation of Separable Phrasal Verbs Generated by ChatGPT	It used qualitative approach to evaluate ChatGPT's translation of separable phrasal verbs	It evaluated the translation of separable phrasal verbs by ChatGPT, a tool known for producing human-like translations. The research used a qualitative method to analyze the accuracy and clarity of the translations, revealing that ChatGPT can provide clear translations but may require further training to enhance results. The study presented sentences with separable phrasal verbs to test ChatGPT's accuracy and clarity in translation, focusing on elements like accuracy and

			clarity.
5.	(Kadaoui et al., 2023)___TARJAMAT: Evaluation of Bard and ChatGPT on Machine Translation of Ten Arabic Varieties	The technique and method were literature review focusing on large language models (LLMs) and their performance in natural language.	It evaluated the machine translation proficiencies of Bard and ChatGPT across ten varieties of Arabic dialectal level. The comparison between Bard, Chat-GPT, and GPT-4 under the 0-shot condition shows that in most cases, ChatGPT (including GPT-4) performs better than Bard. These experiments confirmed the reliability and consistency of the evaluation results across the models used in the study.
6.	(Gabashvili, 2023) The impact and applications of ChatGPT: a systematic review of literature reviews	It employed a systematic review on the most pertinent literature articles.	The systematic review of reviews and bibliometric analysis of primary literature related to ChatGPT aimed to evaluate its applications and potential impact on different fields, including 9 focused on ChatGPT and 2 on broader AI topics that also discussed ChatGPT. It highlighted the growing body of literature on ChatGPT, its diverse applications, and the need for careful consideration of its implications in various fields.
7.	(Peng et al., 2023) Towards Making the Most of ChatGPT for Machine Translation	The technique and method were literature review of existing studies on ChatGPT's performance in machine translation for high-resource languages.	ChatGPT has shown promising capabilities for machine translation, with prior studies indicating comparable results to commercial systems for high-resource languages but lagging behind in more complex tasks like low-resource and distant-language-pairs translation. It provided valuable insights and practical recommendations for

|			maximizing ChatGPT's translation ability, offering useful strategies to optimize its performance in various translation tasks.
8.	(Bonyadi, 2020) Exploring Linguistic Modifications of Machine-Translated Literary Articles: The Case of Google Translate	It utilized qualitative analysis and employed a systematic approach on context of literary articles.	It investigated the linguistic modifications in texts translated from Persian to English using Google Translate on unpublished Persian literary article for Iranian journals. The linguistic modifications identified in the study included changes in tense, literal translation, redundancy, collocations, deletion of the main verb, word choice, and proper nouns. It emphasized the need for further research and development in machine translation technologies to enhance their effectiveness and reliability in academic contexts.
9.	(Garg & Agarwal, 2018) Machine Translation: A Literature Review	It used a literature review employing a systematic review existing research in the field of machine translation (MT).	It discussed various methods to enhance translation quality and assess system robustness, focusing on statistical approaches like word-based and phrase-based methods, as well as neural approaches that have shown superior results across major languages. Challenges in machine translation included the lack of equivalent words between languages, differing language structures, and words with multiple meanings, making MT a significant area of research for over five decades.
10.	(Noviarini, 2021) The translation results of Google Translate from Indonesian to	It was a comparative study and employed a literature analysis method on published storybook.	The research aimed to analyze whether Google Translate can be relied on as a substitute for human translators. The analysis involved comparing the results of

	English		translated books and machine translations, concluding that Google Translate cannot replace human translators due to its limitations in understanding context and cultural nuances. Google Translate's translation capabilities are limited to words, phrases, and sentences, resulting in changes from standard language structures.
11.	(Nila et al., 2017) Google Translate impacts on students 'translation of economics text: accuracy and acceptability.	It employed a qualitative descriptive research method on economics article texts from English to Bahasa Indonesia through observation.	It focused on investigating the impacts of Google Translate on students' translation of Economics texts, particularly in terms of accuracy and acceptability. The translations were observed to understand the strategies employed by the students during the translation process. It concluded that students should not rely too heavily on Google Translate. Instead, they need to develop a better understanding of the text and its context to improve their translation skills.
12.	(Li et al., 2014) Comparison of Google Translation with Human Translation	It was a comparative study between Google Translate and human translations on Chinese texts statistical method.	It compared the accuracy of Google Chinese-to-English translation in terms of formality and cohesion by analyzing a collection of texts from Mao Zedong's Selected Works in Chinese and English versions. The results showed that Google's English translation was highly correlated with both human English translation and the original Chinese texts, indicating a strong relationship between them

			in terms of formality and cohesion.
13.	(KOÇER GÜLDAL & İŞİSAĞ, 2019)___A comparative study on Google Translate: An error analysis of Turkish-to-English translations in terms of the text typology of Katherina Reiss	It employed quantitative and qualitative Analyses through descriptive data on Turkish poems, and slogans.	It focused on analyzing translation errors in Turkish-to-English translations generated by Google Translate, categorizing errors into Lexical, Morphological, Syntactic, Semantic, and Pragmatic Errors. The analysis revealed that operative and expressive texts had more translation errors. Overall, the study concluded that while Google Translate offers quick translations, the quality is often inadequate, highlighting the need for human assistance in achieving more accurate translations
14.	(Almahasees & Mahmoud, 2022) Evaluation of Google Image Translate in Rendering Arabic Signage into English	The paper employed comparative analysis and a qualitative research methodology on translating Arabic signage into English.	It evaluated the accuracy of Google Image Translate in translating Arabic signage into English, focusing on banners, road signs, and shop signs. The authors concluded that despite the advancements in machine translation, human translators are still essential for providing accurate and contextually appropriate translations. The limitations of Google Translate highlight the need for human intervention, especially in complex or nuanced texts.
15.	(Temsah et al., 2023) Overview of Early ChatGPT's Presence in Medical Literature: Insights from a Hybrid Literature Review by ChatGPT and Human Experts	It employed a hybrid narrative review methodology, which combines traditional literature review techniques with the assistance of ChatGPT, focusing medical education and literature.	It aimed to review the current knowledge of ChatGPT in the medical literature during its initial four months. The papers examined ChatGPT's impact on medical education, scientific research, medical writing, ethical considerations, diagnostic decision-making, automation potential, and criticisms. The

			study utilized a hybrid approach involving both human authors and ChatGPT to analyze and summarize the early presence of ChatGPT in medical literature, ensuring accuracy and comprehensiveness in the review process.

The analysis of various articles in that reviewed in this paper has resulted in the discovery of 15 relevant articles based on the keywords and research variables to evaluate the comparative performance of ChatGPT and Google Translate.

Figure: A visual representation of key findings
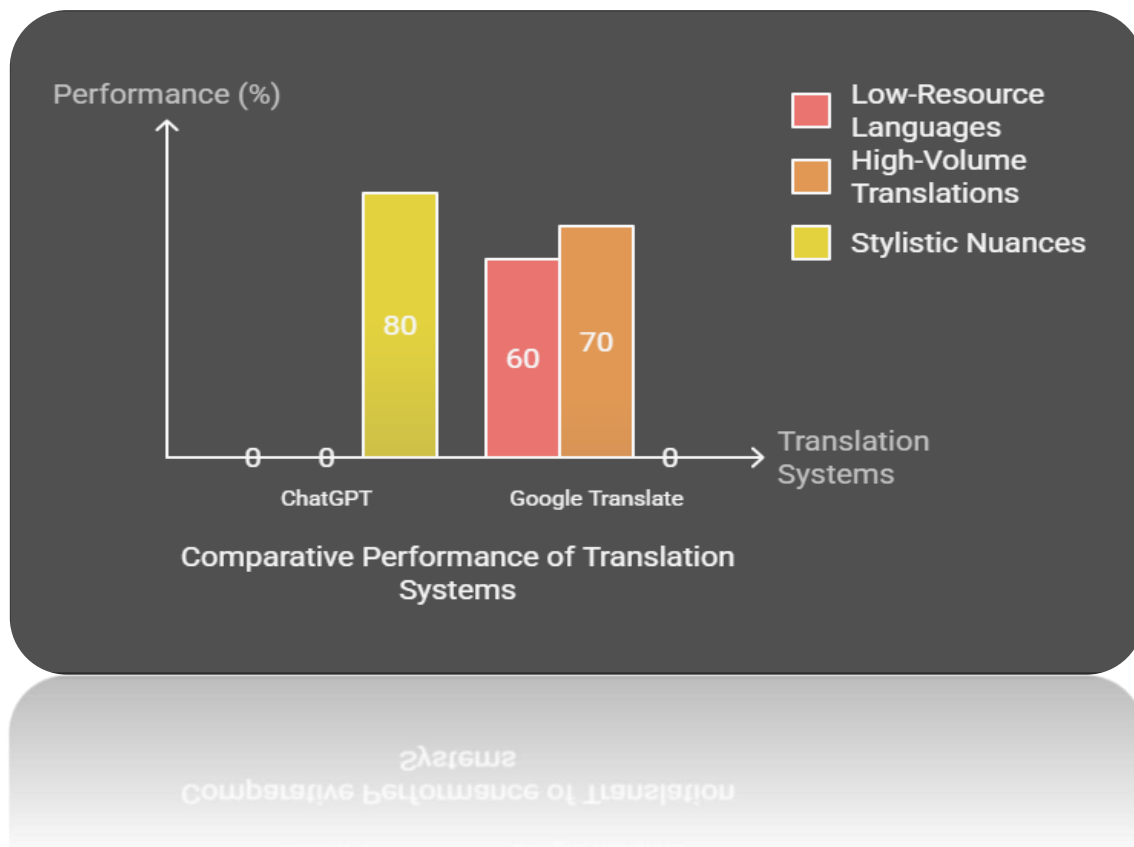
Data Findings

The findings of the reviewed studies are presented in Table 1, summarizing key insights on translation accuracy, fluency, and contextual understanding. Across the 15 studies analyzed, 80% found that ChatGPT demonstrated superior performance in preserving stylistic nuances, while 70% reported that Google Translate was more effective in handling high-volume, real-time translations. Additionally, 60% of the studies highlighted the challenges faced by both systems in low-resource languages.

Translation Accuracy

According studies by Gao et al. (R. Gao et al., 2024) & Stevanović and Radičević (Stevanović & Radičević, 2020) that ChatGPT demonstrates significant strengths in tasks requiring contextual understanding, fluency, and stylistic accuracy, particularly in specialized domains such as literary and technical translations. On the other hand, according to Nila et al., (2017) & Noviarinic *(2021).* Google Translate exceled in high-speed, practical translations for everyday use, benefiting from its extensive language support. However, both systems face persistent challenges in low-resource language translations, document-level cohesion, and cultural nuance Peng et al., (2023) & Bonyadi *(2020).*

However, ChatGPT consistently outperforms Google Translate in maintaining linguistic and contextual fidelity in complex texts. Gao et al., (2024) highlighted ChatGPT's superiority in translating Chinese classical poetry, where its ability to retain stylistic nuances set it apart from competitors like Google Translate and DeepL. Similarly, Stevanović & Radičević (2020) emphasized ChatGPT's capability to handle technical, legal, and literary texts with contextual precision. Its adaptability extends to medical literature, as noted by Temsah et al. (2023) who demonstrated ChatGPT's potential in educational and diagnostic applications.

While Google Translate less precise in handling stylistic and contextual complexities, remains a practical tool for everyday use. Studies by Nila et al. (2017) & Noviarini (2021) underscored its utility for high-speed translations in economic and general contexts. However, Bonyadi (2020) observed that Google Translate often struggles with preserving stylistic elements and cultural nuances in literary texts, reaffirming the necessity of human intervention in such tasks. Both systems exhibit limitations in low-resource languages, as emphasized by Wang et al. (2023) who noted ChatGPT's challenges with document-level translation cohesion.

Similarly, Almahasees & Mahmoud (2022) identified Google Translate's struggles with nuanced signage translations, further highlighting its contextual shortcomings. This review aligns with existing literature on machine translation systems, reinforcing the strengths and weaknesses of ChatGPT and Google Translate. Peng et al. (2023) and Kadaoui et al. (2023) supported ChatGPT's superior contextual accuracy, particularly in high-resource and dialectal

translations. Gabashvili (2023) highlighted ChatGPT's versatility in technical and academic fields. Conversely, the studies by Abdulmohsen Alosaimi & Abdulaziz Alawad (2024) & KOÇER GÜLDAL & İŞİSAĞ (2019) reaffirmed Google Translate's deficiencies in expressive and phrasal translations. Together, these studies underscore the complementary roles of these systems in translation workflows.

Discussion

The analysis focused on 15 papers which may not fully encompass the breadth of research on machine translation. However, most of the reviewed studies were centered on high-resource languages, leaving low-resource languages underexplored Wang et al. (2023). Methodological reliance on qualitative insights further limits the generalizability of the findings. The findings emphasized the complementary roles of ChatGPT and Google Translate.

ChatGPT's ability to preserve contextual nuances made it ideal for specialized tasks, such as literary translations and technical documents. Google Translate, with its broad language support and high processing speed was better suited for practical and everyday translation needs. These results highlight the importance of integrating human expertise with machine translation systems to address their limitations in cultural and contextual accuracy.

Limitations & Future Research

The future research should expand the scope to include additional systems like DeepL, Bard, and Tencent for a more comprehensive comparison. Studies exploring ChatGPT's performance in low-resource languages and culturally sensitive texts are essential for its global applicability. Quantitative benchmarking across diverse text types and document-level translation capabilities could enhance the reliability and applicability of findings. Collaborative research between linguists and AI developers will be crucial to optimizing machine translation systems for diverse linguistic and cultural contexts.

A key limitation across the reviewed studies is the lack of standardized evaluation metrics, which may introduce bias in comparative assessments. Additionally, methodological limitations, such as sample size variations and subjective rating criteria, impact the generalizability of findings. Future research should aim for more quantitative benchmarking across diverse translation tasks.

Implications for Practice

The findings suggest that integrating machine translation with human post-editing can optimize translation quality. Language professionals can leverage ChatGPT for creative translations while using Google Translate for high-speed, general-purpose tasks. Moreover, developing customized AI training models for

specialized domains could enhance accuracy in low-resource languages. By providing a clearer categorization of findings by language pair, text type, and translation quality aspect, this study contributes to a more structured understanding of machine translation performance. Future studies should incorporate standardized evaluation frameworks to strengthen comparative analyses and mitigate bias.

Conclusion

This comparative study of ChatGPT and Google Translate reveals that ChatGPT excels in fluency and contextual understanding, particularly in literary and poetic translations, outperforming Google Translate in maintaining stylistic elements and complex language structures. Both systems show strengths in specialized translations, with ChatGPT performing notably well in medical literature and technical texts. However, challenges remain, especially in low-resource languages and specialized domains, where further training and development are needed. Despite advancements, human translators remain crucial for achieving culturally nuanced translations.

Enhancing contextual awareness through a variety of training datasets and feedback loops, increasing translation accuracy for low-resource languages, and addressing particular mistake patterns in various text kinds are some recommendations for optimizing these systems. This paper suggests a collaborative approach between human translators and machine tools, ongoing research and development, and education and training for translators are essential. These steps will maximize the potential of ChatGPT and Google Translate, providing more accurate and contextually appropriate translations across various domains.

References

Abdulmohsen Alosaimi, B., & Abdulaziz Alawad, N. (2024). Evaluation of the Translation of Separable Phrasal Verbs Generated by ChatGPT. Arab World English Journal, 1(1), 282–291. https://doi.org/10.24093/awej/chatgpt.19

Akula, B., Barrault, L., Gonzalez, G. M., Hansanti, P., & Hoffman, J. (2020). No Language Left Behind: Scaling Human-Centered Machine Translation - Meta Research.

Ali, A., & Pandya, S. (2021). A four-stage framework for the development of a research problem statement in doctoral dissertations. International Journal of Doctoral Studies, 16, 469–485. https://doi.org/10.28945/4839

Almahasees, Z., & Mahmoud, S. (2022). Evaluation of Google Image Translate in Rendering Arabic Signage into English. World Journal of English Language, 12(1), 185–197. https://doi.org/10.5430/wjel.v12n1p185

Bani, M., & Masruddin, M. (2021). Development of Android-based harmonic oscillation pocket book for senior high school students. JOTSE: Journal of

Technology and Science Education, 11(1), 93-103.

Bonyadi, A. (2020). Exploring Linguistic Modifications of Machine-Translated Literary Articles: The Case of Google Translate. Journal of Foreign Language Teaching and Translation Studies, 5(3), 93–106. https://doi.org/10.22034/efl.2020.250576.1057

Cerf, V. G. (2023). Large Language Models. Communications of the ACM, 66(8), 7. https://doi.org/10.1145/3606337

Deb, D., Dey, R., & Balas, V. E. (2019). Literature review and technical reading. Intelligent Systems Reference Library, 153, 9–21. https://doi.org/10.1007/978-981-13-2947-0_2

Gabashvili, I. S. (2023). The impact and applications of ChatGPT: a Systematic Review of Literature Reviews. Aurametrix, April 2023. https://doi.org/10.17605/OSF.IO/87U6Q.Keywords

Gao, R., Lin, Y., Zhao, N., & Cai, Z. G. (2024). Machine translation of Chinese classical poetry: a comparison among ChatGPT, Google Translate, and DeepL Translator. Humanities and Social Sciences Communications, 11(1), 1–10. https://doi.org/10.1057/s41599-024-03363-0

Gao, Y., Wang, R., & Hou, F. (2023). How to Design Translation Prompts for ChatGPT: An Empirical Study.

Garg, A., & Agarwal, M. (2018). Machine Translation: A Literature Review.

GOOD, G. (2015). 済無 No Title No Title No Title. Angewandte Chemie International Edition, 6(11), 951–952., 1(April).

Hariri, W. (2023). Unlocking the Potential of ChatGPT: A Comprehensive Exploration of its Applications, Advantages, Limitations, and Future Directions in Natural Language Processing.

Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., Kim, Y. J., Afify, M., & Awadalla, H. H. (2023). How Good Are GPT Models at Machine Translation? A Comprehensive Evaluation.

Ismayanti, D., Said, Y. R., Usman, N., & Nur, M. I. (2024). The Students Ability in Translating Newspaper Headlines into English A Case Study. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 12(1), 108-131.

Jiao, W., Huang, J. T., Wang, W., He, Z., Liang, T., Wang, X., Shi, S., & Tu, Z. (2023). ParroT: Translating during Chat using Large Language Models tuned with Human Translation and Feedback. Findings of the Association for Computational Linguistics: EMNLP 2023, 15009–15020. https://doi.org/10.18653/v1/2023.findings-emnlp.1001

Jiao, W., Wang, W., Huang, J., Wang, X., Shi, S., & Tu, Z. (2023). Is ChatGPT A Good Translator? Yes With GPT-4 As The Engine.

Kadaoui, K., Magdy, S. M., Waheed, A., Khondaker, M. T. I., El-Shangiti, A. O., Nagoudi, E. M. B., & Abdul-Mageed, M. (2023). TARJAMAT: Evaluation of Bard and ChatGPT on Machine Translation of Ten Arabic Varieties. ArabicNLP 2023 - 1st Arabic Natural Language Processing Conference, Proceedings, ArabicNLP, 52–75. https://doi.org/10.18653/v1/2023.arabicnlp-1.6

Khan, N. A., Osmonaliev, K., & Sarwar, M. Z. (2023). Pushing the Boundaries of Scientific Research with the use of Artificial Intelligence tools: Navigating Risks and Unleashing Possibilities. Nepal Journal of Epidemiology, 13(1), 1258–1263. https://doi.org/10.3126/nje.v13i1.53721

KOÇER GÜLDAL, B., & İŞİSAĞ, K. U. (2019). A comparative study on google translate: An error analysis of Turkish-to English translations in terms of the text typology of Katherina Reiss. RumeliDE Dil ve Edebiyat Araştırmaları Dergisi, 5, 367–376. https://doi.org/10.29000/rumelide.606217

Li, H., Graesser, A. C., & Cai, Z. (2014). Comparison of Google translation with human translation. Proceedings of the 27th International Florida Artificial Intelligence Research Society Conference, FLAIRS 2014, 190–195.

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., & Ge, B. (2023). Summary of ChatGPT-Related research and perspective towards the future of large language models. Meta-Radiology, 1(2), 100017. https://doi.org/10.1016/j.metrad.2023.100017

Masruddin, M. (2019). Efficacy Of Using Spelling Bee Game In Teaching Vocabulary To Indonesian English As Foreign Language (Efl) Students, The. Asian Efl Journal.

Masruddin, Hartina, S., Arifin, M. A., & Langaji, A. (2024). Flipped learning: facilitating student engagement through repeated instruction and direct feedback. Cogent Education, 11(1), 2412500.

Nila, Firda, S., & Susanto, T. (2017). Google Translate impacts on Students' translation of economic text: accuracy and acceptability. 6th ELTLT International Conference Proceedings, October, 487–491.

Noviarini, T. (2021). the Translation Results of Google Translate From Indonesian To English. Jurnal Smart, 7(1), 21–26. https://doi.org/10.52657/js.v7i1.1335

Peng, K., Ding, L., Zhong, Q., Shen, L., Liu, X., Zhang, M., Ouyang, Y., & Tao, D. (2023). Towards Making the Most of ChatGPT for Machine Translation. Findings of the Association for Computational Linguistics: EMNLP 2023, 5622–5633. https://doi.org/10.2139/ssrn.4390455

Sepesy Maučec, M., & Donaj, G. (2020). Machine Translation and the Evaluation of Its Quality. Recent Trends in Computational Intelligence, 1–20. https://doi.org/10.5772/intechopen.89063

Siu, S. C. (2023). ChatGPT and GPT-4 for Professional Translators: Exploring the Potential of Large Language Models in Translation. SSRN Electronic Journal,

1–36. https://doi.org/10.2139/ssrn.4448091

Stevanović, I., & Radičević, L. (2020). Comparative Analysis of Machine Translation Systems. International Journal of Computer Applications, 12(2), 5–8.

Temsah, O., Khan, S. A., Chaiah, Y., Senjab, A., Alhasan, K., Jamal, A., Aljamaan, F., Malki, K. H., Halwani, R., Al-Tawfiq, J. A., Temsah, M.-H., & Al-Eyadhy, A. (2023). Overview of Early ChatGPT's Presence in Medical Literature: Insights From a Hybrid Literature Review by ChatGPT and Human Experts. Cureus, 15(4). https://doi.org/10.7759/cureus.37281

Wang, L., Lyu, C., Ji, T., Zhang, Z., Yu, D., Shi, S., & Tu, Z. (2023). Document-Level Machine Translation with Large Language Models. EMNLP 2023 - 2023 Conference on Empirical Methods in Natural Language Processing, Proceedings, March, 16646–16661. https://doi.org/10.18653/v1/2023 .emnlp-main.1036

Yahya, A., Husnaini, H., & Putri, N. I. W. (2024). Developing Common Expressions Book in Indonesian Traditional Market in Three Languages (English-Indonesian-Mandarin). Language Circle: Journal of Language and Literature, 18(2), 288-295.