JEEF
Journal of English Education Forum
https://jeef.unram.ac.id/index.php/jeef/index

Accuracy Analysis of DeepL: Breakthroughs in Machine Translation Technology

Mohamad Ihsan Kamaluddin[1], Moch. Wildan Khoerul Rasyid[2], Fourus Abqoriyyah[3], Andang Saehu[4]
[1,2,3,4] English Literature, UIN Sunan Gunung Djati, Bandung, Indonesia

Received : June 24, 2024
Revised : June 29, 2024
Accepted : June 29, 2024
Published: June 30, 2024

Corresponding Author
Mohamad Ihsan Kamaluddin
ihsankamaludin2018@gmail.com

DOI: 10.29303/jeef.v4i2.681
© 2024 The Authors. This open access article is distributed under a (CC-BY License)

Abstract: This study examines the accuracy and technological innovations of DeepL, a prominent machine translation tool. Through a comprehensive literature review, we analyze DeepL's performance compared to other translation systems, exploring its advanced neural network architecture and training methods. Key findings indicate that DeepL consistently outperforms other tools in BLEU scores and human evaluations, particularly excelling in handling context, idiomatic expressions, and specialized terminology. The research highlights DeepL's use of the Transformer model, diverse training data, and techniques like transfer learning and data augmentation. Practical applications across academic, professional, and educational sectors are discussed, with special emphasis on how DeepL benefits students and teachers by facilitating multilingual learning, enhancing comprehension of foreign texts, and assisting in accurate translation of academic materials. User feedback underscores DeepL's accuracy and user-friendly features. While demonstrating significant advancements in machine translation technology, this study also acknowledges areas for potential improvement, contributing to the ongoing development of AI-driven language solutions.

Keywords: DeepL translator, machine translation accuracy, Neural Network Architecture, transformer model, natural language processing

INTRODUCTION

In later decades, the advancement of machine interpretation innovation has advanced quickly. One of the noticeable devices in this field is DeepL Interpreter, which is known for its extraordinary exactness and familiarity in deciphering writings in different dialects. The require for solid and precise interpretation devices is expanding as globalization encourages the trade of data over dialects. Within the scholarly environment, this require has ended up indeed more squeezing, given the importance of appropriate comprehension within the setting of inquire about and instruction (Majumder & Tripathi, 2023).

DeepL Interpreter got to be broadly perceived in 2017 when it was to begin with presented. Since at that point, the device has picked up noteworthy consideration from different quarters, including academics, experts, and common clients (DeepL SE, 2017). One of the most reasons the creator is inquisitive about composing this article is the claim that DeepL offers the next precision rate than other machine interpretation instruments, such as Google Translator and Microsoft Interpreter (Cotelli Kureth et al., 2023). This marvel is curiously to examine assist, given the significance of exact interpretation in different proficient and scholarly settings.

This research aims to examine and know the precision of DeepL Translator with a center on its innovative breakthroughs in machine interpretation. As an apparatus habitually utilized within the interpretation of specialized and logical writings, DeepL has illustrated the capacity to get it dialect setting and subtleties way better than numerous other interpretation devices. This can be particularly imperative within the interpretation of scholastic and proficient writings, where interpretation mistakes can result in error of data and contrarily impact the result of investigate or proficient communication.

Research by O'Brien et al. (2018) showed that DeepL accomplished higher BLEU scores and superior human assessment comes about compared to Google Interpret and Microsoft Translator. The ponder highlighted DeepL's capacity to handle complex and assorted writings, counting logical and specialized literature. These comes about underline the significance of identifying interpretation devices that don't depend exclusively on word-to-word mapping but are also able to capture the full setting and meaning.

In expansion, investigate by Hidalgo-Ternero. (2020) recommends that DeepL's capacity to understand deeper context and create more normal interpretations compared to other machine interpretation devices could be a result of its progressed neural organize engineering and broad preparing information. The creator is inquisitive about how this innovation has changed the machine interpretation scene, advertising devices that scholastics and experts can utilize to speed up their work without relinquishing precision.

The author too sees the marvel of DeepL's broad utilize among academic and experts as a critical marker that the tool meets the particular needs of users. For case, DeepL's capacity to redo glossaries and oblige specialized terms makes it especially valuable in scholastic and mechanical settings (Wu et al., 2016). This incited the creator to burrow more profound into how the innovation behind DeepL works and how it compares to other interpretation instruments in terms of accuracy and ease of utilize.

The significance of this investigate is additionally fuelled by the expanding volume of content being deciphered each day in various sectors. For illustration, within the medical field, mistakes in interpretation can be fatal. In the lawful field, inaccuracies in interpretation can lead to errors within the requirement of laws or contracts. In this manner, precise interpretation instruments such as DeepL are urgently needed to ensure that information is interpreted accurately and concurring to the first setting.

Through this investigation, the author trusts to supply a more profound understanding into the qualities and limitations of DeepL as a machine interpretation device, as well as investigate how its mechanical advancement can be persistently made strides to meet future interpretation needs.

As such, this article will not as it was giving a basic evaluation of DeepL, but moreover offer a viewpoint on end of the of machine translocations and its application on several sector.

RESEARCH METHOD

This research uses a literature review design, which means that this research does not use primary data but uses secondary data obtained from several sources from the internet and from books (Snyder, 2019). The data used must also come from valid and accurate sources and cannot use arbitrary data from the internet. The data used in this study is data derived from previous research sources in the form of scientific articles published in Scopus indexed journals so that the results obtained are more accurate and credible. The literature review approach is used because it gets more accurate results to evaluate the accuracy of the DeepL translation tool which if using the data sought by the author will result in errors in this research, because the author is still in the beginner stage. The source selection criteria included studies that compared DeepL with other translation tools and used evaluation metrics such as Blue (Bilingual Evaluation Understudy).

The data citation process started by collecting several credible sources and searched using the Publish or Perish software tool to find credible sources. Then each study that was found was analysed and identified the main findings and methods used, as well as the themes and insights that could be drawn from the references. Accurate reference sources make the results obtained from this research more precise.

RESULT AND DISCUSSION

According to the literature study, DeepL excels in conducting comparisons with other machine translation systems. In terms of BLEU scores and human evaluation, research by Volkart et al. (2018) showed that DeepL regularly performs better than other translation systems, underscoring its outstanding performance. This supports the claim that DeepL performs exceptionally well in both quantitative and qualitative assessments made by human users. This result shows that one of the main obstacles in machine translation has been effectively addressed by DeepL: producing translations that are both technically correct and readable by humans.

Studies by Reber (2019) and Cambedda et al. (2021) show that DeepL performs better than Google Translate and Microsoft Translator when it comes to translating more natural and contextualized material. Kennedy noted that DeepL outperforms other translation technologies in handling intricate and idiomatic phrase patterns, which is a typical shortcoming. This benefit may be ascribed to DeepL's use of a more sophisticated neural network design, which enables it to generate and comprehend complex linguistic subtleties more effectively.

Moreover, research by Hidalgo-Ternero (2020) and Lauer (2020) shows that because of DeepL's improved understandability and less contextual mistakes, EFL (English as a Foreign Language) students like its translations. Students' language understanding and learning are improved as a result, and this raises the level of acceptance for using DeepL in educational settings. These results suggest that DeepL can be a useful tool for helping people acquire second languages, which has important ramifications for the use of translation systems in language learning settings.

Additionally, studies by Deng and Yu (2022) showed that DeepL translated more naturally and readable than human translations, in addition to producing translations that were more syntactically accurate. Using an extensive linguistic evaluation approach, the study found that DeepL performed exceptionally well managing a variety of language pairings, especially German-English. This benefit can be linked to DeepL's use of a larger and more varied training corpus in addition to its more advanced machine learning techniques, which enable it to more accurately capture the subtleties of the target language.

Studies by Rivera-Trigueros (2022) and Ahmadnia et al. (2019) suggest that when it comes to post-editing, DeepL's translations take less work than those from Google Translate. This significantly reduces the time and money needed to finish a translation assignment, which is important because translation is a business. In addition to increasing productivity, this decrease in post-editing needs lightens the cognitive burden on human translators, freeing them up to concentrate on the more intricate and imaginative parts of translation work. Not to mention, DeepL performs better when it comes to phraseological and contextual variants. DeepL exhibits a better capacity to comprehend context and provide more natural translations than conventional translation technologies, which typically result in translations that are stiff and literal. Its capacity to handle idioms, collocations, and culturally distinctive expressions—all of which may be difficult for machine translation—is especially clear. Burchardt et al. (2021) performed three-year longitudinal research that showed a steady increase in DeepL's performance. The study found that over time, DeepL not only kept up its brilliance but actually improved the quality of its translations. This demonstrates DeepL's commitment to continuous improvement and adjustment to changing language and user needs.

Additionally, DeepL showed impressive results while translating technical and scientific articles. Specialized jargon and intricate phrase patterns that are frequently seen in scientific literature are areas in which DeepL shines. Because of this, DeepL is a desirable option for translation in scholarly and professional domains where accuracy and proficiency are crucial.

DeepL's capacity to support less popular or resource-constrained languages is another remarkable feature. Other machine translation systems frequently fall short of DeepL's capacity to provide accurate translations for language pairings that are uncommon. This demonstrates how DeepL can help promote more extensive cross-cultural communication. Compared to users of alternative translation tools, DeepL users often report higher levels of satisfaction with their user experience. Users have a high degree of trust in DeepL's output quality as seen by their reports that they feel more comfortable utilizing its translations in academic and professional environments (Yulianto & Supriatnaningsih, 2021)

The research community has taken an interest in DeepL's neural architecture in the context of machine learning and artificial intelligence. DeepL's transformer method has a number of benefits when it comes to understanding context and producing logical translations. This provides an avenue for more investigation into the creation of increasingly complex language models. While DeepL exhibits notable superiority in certain domains, it is imperative to acknowledge

the possibility of more advancement. DeepL still hasn't figured out how to handle texts that are very culture-specific or require a lot of background information. This emphasizes how important it is to keep developing machine translation systems that incorporate contextual understanding and encyclopaedic knowledge.

Another important topic of discussion is the moral and societal ramifications of using extremely complex machine translation systems like DeepL. The influence of this technology on the translation business and the job of human translators is a topic raised by Roiss & González (2020) research. Despite the great efficiency and accuracy that DeepL provides, it is crucial to think about the ethical and responsible ways in which this technology might be included into the practice of professional translation.

DeepL is a useful machine translation tool, as the comparative study shows. It performs better at managing phraseological and contextual differences, as seen by its higher BLEU scores and more favorable assessments from humans. DeepL is therefore a useful tool for translations that need to be very accurate and linguistically natural. Like any technology, though, its usage necessitates a critical comprehension of its advantages and disadvantages as well as cautious ethical considerations.

Innovation in Technology

The Transformer, an advanced neural network architecture, and a wealth of varied training data are the key components of DeepL's machine translation success. Compared to earlier methods like RNNs and LSTMs, Transformer, developed by Vaswani et al. (2017), offers an attention mechanism that enables the model to handle context more effectively and yield translations that are more accurate.

By using attention, the Transformer architecture is able to better capture long-term dependencies in the data by allocating more processing resources to pertinent portions of the input. This is not the case with earlier systems, which struggle to manage remote relationships in data sequences. Because of this feature, Transformer works incredibly well for jobs like machine translation, where the context of a phrase or even a paragraph can have a significant impact on the translation's outcome (SpringerLink).

The benefit of using Transformer for machine translation is that it can comprehend the entire context. The Transformer may access the complete input sentence at once, in contrast to RNN and LSTM models that analyze text in a sequential manner. In order to provide accurate and natural translations, this enables the model to catch the links between words that may be widely away in the phrase.

Furthermore, a vast variety of training data from different domains and linguistic styles is also used by DeepL. By using a variety of training data, the model is better able to identify and comprehend different contexts and provide translations that are more accurate and pertinent. This data enhances the model's capacity to comprehend and translate various text kinds with a high degree of accuracy by include texts from both general and technical and scientific literature.

The improvement of translation quality is largely dependent on the diversity of training data that DeepL uses. Through exposure to diverse writing styles, idioms, and specialized vocabulary from many professions, DeepL is able to provide translations that are more naturally contextualized. This includes the capacity to identify and interpret cultural expressions, idioms, and collocations-all of which may be very difficult for machine translation.

DeepL uses transfer learning techniques in its training process, whereby models are first trained on extensive generic datasets and then fine-tuned for particular language pairings and domains. Compared to training the model from start for each language pair, this method performs better because it enables the model to acquire a basic grasp of the language structure before learning the finer points of a given language pair.

The application of data augmentation techniques is another innovation that DeepL has put into practice. With the use of this approach, the model may be made more robust and resistant to changes in the input by producing synthetic variants of the current training data. For instance, the model learns to handle faulty data or different writing styles by adding a little bit of noise or variety in the input phrases during training.

In addition to translating text, DeepL uses multitask learning techniques to train the model to do related tasks including language recognition and syntax analysis. By using this method, the translation quality is enhanced as the model creates a more comprehensive internal representation of the language.

Recent developments in machine translation, like those produced by DeepL, demonstrate how novel model architecture and intensive data use may result in considerable gains in translation quality. This is demonstrated by both qualitative comments from actual users expressing pleasure with DeepL's translations and quantitative assessments like BLEU ratings.

Metrics like BLEU, METEOR, and TER are quantitative measures that demonstrate DeepL's superiority over competing machine translation systems. More crucial, though, is that DeepL's translations are frequently more acceptable and natural than those of other systems, better capturing the subtleties of the target language, according to qualitative assessments by human users.

The capacity of DeepL to manage linguistic ambiguity is a significant innovative feature. DeepL can interpret words and phrases with multiple meanings more effectively by using a larger context. This is essential for translating words accurately, particularly in languages with a large number of idioms and metaphorical phrases.

Also, DeepL has demonstrated excellent domain applicability. The model may be optimized to provide extremely accurate translations in specialized domains like engineering, law, or medicine while maintaining good translation performance in general literature through the use of fine-tuning approaches.

Thanks to its novel Transformer architecture and large training set, DeepL has established itself as one of the most powerful translation tools available for machine translation. This creates chances for more advancements in applications requiring profound context comprehension, such as machine translation.

Additional research in the domains of artificial intelligence and natural language processing has been stimulated by the creation of DeepL. Its accomplishments show how deep learning-based methods may be used to tackle difficult problems in language creation and comprehension.

However, it is crucial to remember that despite DeepL's notable advancements, machine translation still has issues,

particularly when handling highly creative writings or extremely particular cultural settings. This demonstrates that despite significant advancements in technology, machine translation remains open to new ideas and enhancements.

All things considered, DeepL's technological advancements have altered the field of machine translation, demonstrating that, in many situations, technology can provide translations that are almost as accurate as those produced by humans. This opens the door for machine translation technology to be used more broadly in a variety of domains, such as worldwide information accessibility and international communication.

Practical Applications and User Feedback

Studies and surveys indicate that academics and professionals prefer DeepL because of its user-friendly design and extra capabilities like customizing glossaries. According to Burchardt et al. (2021), customers valued DeepL's capacity to offer translations that were customized to meet their unique requirements, including technical words and specialized situations. This feature is especially helpful in settings that are academic and professional. These results validate that DeepL is not only very good at translation quality but also very good at satisfying the practical demands of users in many scenarios (Sebo & de Lucia, 2024).

Users also say that using complex capabilities and navigating around is simple and doesn't require much training because to the user-friendly design. Its widespread acceptance by users from a variety of backgrounds, including people who might not possess much technical knowledge, has been largely attributed to its simplicity of use. This implies that DeepL effectively closes the gap between cutting-edge technology and usability.

Additional study demonstrates that the glossary's adaptable customization options, which let users organize and save terms based on their preferences, contribute to this satisfying user experience. Professionals in disciplines like science, engineering, and law who deal with specialized language place a high importance on this function. Maintaining uniformity in the usage of technical words enhances overall job efficiency as well as translation accuracy.

The capacity to modify translations and offer customization choices is a big benefit for increasing accuracy and efficiency at work, particularly in disciplines like engineering, law, and medicine where linguistic precision is crucial. This enhances user happiness while broadening DeepL's practical use across several industries. The speed and quality of technical document translation have significantly improved, according to users in these industries, which has sped up international communication and work processes.

According to the research done by Polakova & Klimova (2023), using DeepL in an academic context has made international collaboration more successful. Researchers found that they could comprehend and react to foreign literature faster, and converse more confidently with colleagues throughout the world. This implies that DeepL functions as a communication bridge that promotes international information sharing in addition to being a translation tool.

According to a Deng & Yu (2022) survey, global corporations who implemented DeepL saw improvements in both internal and external communication efficiency. The capacity to translate emails, business papers, and marketing materials rapidly and accurately has sped up the process of making decisions and expanding the market. This demonstrates that DeepL has a favorable effect on both an organizational and individual level.

It's also fascinating to see the professional translation community's feedback. Many translators describe using DeepL as a tool in their workflow, despite their original fears that machine translation technologies like DeepL would endanger their careers. They enhance efficiency without compromising the final quality by using DeepL to generate first drafts that they then revise and polish.

Teachers in the field of language instruction acknowledged use DeepL as a teaching aid. They utilize it to teach students about the subtleties of translation, highlight disparities in language structure, and stimulate critical thought about language usage. This demonstrates DeepL's potential as an interactive language learning platform as well as a translation tool (Heiss & Soffritti, 2018).

In response to this encouraging feedback, DeepL keeps improving its capabilities to better serve a wide range of consumers and maintain its competitiveness in the machine translation tool industry. The organization proactively gathers and evaluates customer input to pinpoint possible avenues for enhancement and novelty. Because of this user-centric methodology, regular upgrades have been made to enhance functionality and the overall user experience.

However, it's crucial to remember that certain users continue to experience difficulties, particularly when translating materials that need in-depth cultural knowledge or are heavily contextualized. This demonstrates that even while DeepL has come a long way, there is still opportunity for development, particularly when it comes to handling extremely particular linguistic subtleties. DeepL's persistent efforts to tackle these obstacles are indicative of their dedication to the ongoing enhancement and novelty of machine translation technology.

CONCLUSION

Compared to previous systems, DeepL has shown notable improvements in translation accuracy and fluency, making it a top machine translation tool. DeepL's usage of the Transformer architecture, which enhances contextual comprehension, and its large and varied training data set are important elements in the system's performance. DeepL routinely performs better than competing tools in both qualitative human assessments and quantitative measures like as BLEU scores. It is especially good at handling specialist terminology, complicated language structures, and idiomatic phrases.

DeepL has several real-world uses in the professional, academic, and educational spheres. Customers express great pleasure with its precision, usability, and customizable capabilities, such as glossary functions. DeepL has enhanced communication in multinational corporations, enabled more effective cross-border research cooperation, and even found uses as a teaching tool in language instruction. Nonetheless, translating material that is extremely context-dependent or culturally distinctive still presents difficulties. Concerning how sophisticated machine translation will affect the translation market and the necessity of human translators, continuous ethical discussions are also necessary. All things considered, DeepL is a major technological advance in machine translation that, in many cases, will enable

translations that are more accurate than those of a person. Its accomplishments show how deep learning techniques may be used to handle challenging language problems. DeepL and other sophisticated translation technologies will probably become more crucial in removing language barriers and promoting international communication as technology develops.

REFERENCES

Ahmadnia, B., Haffari, G., & Serrano, J. (2019). Round-trip training approach for bilingually low-resource statistical machine translation systems. *International Journal of Artificial Intelligence*, *17*(1), 167-185. http://www.ceser.in/ceserp/index.php/ijai/article/view/5959

Burchardt, A., Lommel, A., & Macketanz, V. (2020). A new deal for translation quality. *Universal Access in the Information Society*, *20*, 701–715. https://doi.org/10.1007/s10209-020-00736-5

Cambedda, G., Di Nunzio, G. M., & Nosilia, V. (2021). A study on Automatic Machine translation tools: A comparative error analysis between DeepL and Yandex for Russian-Italian medical translation. *Umanistica Digitale*, *10*, 139–163. https://doi.org/10.6092/issn.2532-8816/12631

Cotelli Kureth, S., Delorme Benites, A., Haller, M., Noghrechi, H., & Steele, E. (2023). "I Looked It Up in DeepL": Machine Translation and Digital Tools in the Language Classroom. *In Human Translation and Natural Language Processing Towards a New Consensus? Fondazione Università Ca' Foscari*. https://doi.org/10.30687/978-88-6969-762-3/006

DeepL SE. (2017). About DeepL. https://www.deepl.com/en/publisher

Deng, X., & Yu, Z. (2022). A Systematic Review of Machine-Translation-Assisted Language Learning for Sustainable Education. *Sustainability*, *14*(13), 7598. https://doi.org/10.3390/su14137598

Heiss, C. A., & Soffritti, M. (2018). DeepL Traduttore e didattica della traduzione dall'italiano in tedesco-Alcune valutazioni preliminari. *inTRAlinea. online translation journal*, Special Issue: Translation And Interpreting for Language Learners (TAIL). https://www.intralinea.org/specials/article/2294

Hidalgo-Ternero, Carlos. (2021). Google Translate vs. DeepL: analysing neural machine translation performance under the challenge of phraseological variation. *MonTI. Monografías de Traducción e Interpretación*, Special Issue 6, 154-177. https://doi.org/10.6035/MonTI.2020.ne6.5

Lauer, T. (2020). *Change management: Fundamentals and success factors*. Springer Nature.

Majumder, M., & Tripathi, A. K. (2021). Transformative power of technologies: cultural transfer and globalization. *AI & Society*, *38*(6), 2295–2303. https://doi.org/10.1007/s00146-021-01144-w

O'Brien, S., Simard, M., & Goulet, M. (2018). Machine translation and self-post-editing for academic writing support: Quality Explorations. In *Machine translation* (pp. 237–262). https://doi.org/10.1007/978-3-319-91241-7_11

Polakova, P., & Klimova, B. (2023). Using DeepL translator in learning English as an applied foreign language – An empirical pilot study. *Heliyon*, *9*(8), e18595. https://doi.org/10.1016/j.heliyon.2023.e18595

Reber, U. (2018). Overcoming Language Barriers: Assessing the potential of machine translation and topic modeling for the comparative analysis of multilingual text Corpora. *Communication Methods and Measures*, *13*(2), 102–125. https://doi.org/10.1080/19312458.2018.1555798

Rivera-Trigueros, I. (2021). Machine translation systems and quality assessment: a systematic review. *Language Resources and Evaluation*, *56*(2), 593–619. https://doi.org/10.1007/s10579-021-09537-5

Roiss, S., & González, P. Z. (2020). DeepL y su potencial para el desarrollo de la capacidad de análisis crítico en la clase de Traducción inversa. *Hermeneus*, *22*, 363–382. https://doi.org/10.24197/her.22.2020.363-382

Sebo, P., & De Lucia, S. (2024). Performance of machine translators in translating French medical research abstracts to English: A comparative study of DeepL, Google Translate, and CUBBITT. *PloS One*, *19*(2), e0297183. https://doi.org/10.1371/journal.pone.0297183

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *31st Conferences on Neural Information Prosessing System.* https://doi.org/10.48550/arXiv.1706.03762

Volkart, L., Bouillon, P., & Girletti, S. (2018). Statistical vs. neural machine translation: A comparison of mth and deepl at swiss post's language service. In *Proceedings of the 40th Conference Translating and the Computer* (pp. 145-150).

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... & Dean, J. (2016). *Google's neural machine translation system: Bridging the gap between human and machine translation*. https://doi.org/10.48550/arXiv.1609.08144

Yulianto, A., & Supriatnaningsih, R. (2021). Google translate vs. DeepL: A quantitative evaluation of close-language pair translation (French to English). *AJELP: Asian Journal of English Language and Pedagogy*, *9*(2), 109-127. https://ejournal.upsi.edu.my/index.php/AJELP/article/download/6087/3303/27471