

## Bridging Language Learning and SDG 4 through AI: A Comparative Study of Google Translate and DeepL in Arabic Text Understanding

Alysa Salsabila Ramadhan<sup>1</sup>

<sup>1</sup>Universitas Sultan Ageng Tirtayasa, Banten, Indonesia



### ABSTRACT

**Keywords:**

AI Translation,  
Arabic Language,  
Google Translate,  
DeepL Translate,  
SDG 4

*This study aims to analyze the contribution of artificial intelligence (AI)-based translation tools, Google Translate (GT) and DeepL Translate (DT), in supporting Arabic language learning and the achievement of Sustainable Development Goal (SDG) 4: Quality Education. This study uses a comparative qualitative approach by measuring three language components, namely meaning accuracy, grammatical structure, and comprehensibility. Data were obtained through a sample of Arabic narrative text translations assessed by educators and students using a 1–5 Likert scale, then analyzed descriptively. The results show that DeepL Translate is superior to Google Translate in terms of meaning accuracy and structural clarity, making it more effective for learning that requires contextual understanding of meaning. Meanwhile, Google Translate remains useful for quick and easily accessible general translation. The novelty of this study lies in the combination of AI translation analysis with the SDG 4 perspective, which emphasizes that AI translation tools not only function as linguistic tools but also as supporters of inclusive, equitable, and technology-based learning, while highlighting the importance of AI literacy for educators and learners.*

### INTRODUCTION

Digital transformation has become a major driving force in bringing about fundamental changes in the world of education, including language learning. (Nuralan, 2025) AI-based applications such as automatic translation tools are now increasingly being used in educational contexts to facilitate communication, text comprehension, and crosslanguage interaction. (Woo & Choi, 2021) This era of change has significantly impacted the field of translation, moving away from conventional practices and toward innovative technological advancement. (Maulidiya et al., 2025). Through language learning automation, AI enables interactive and adaptive learning, where materials and feedback are tailored to individual needs.

Arabic text, as a language that is morphologically rich and grammatically complex, presents its own challenges in the translation process. (Zafar & Komilovich, 2024) Many learners find it difficult to understand Arabic texts because of their complex sentence structure and deep meaning. In this context, translation becomes an important bridge to help learners understand the content and message of Arabic texts. As stated by (Rosita et al., 2024), translation is a means of communication that allows individuals to share information and society to gain knowledge or information through translation. The development of digital technology has made language translation easier, with many automatic translation applications (machine translation tools) now available to aid in the process of understanding texts across languages. Among the popular artificial intelligence (AI)-based translation tools are Google Translate (GT) and DeepL Translate

(DT). These two tools are practical suggestions for students in understanding Arabic texts, although the quality of the translation results varies, especially in terms of accuracy of meaning, grammatical structure, and comprehensibility.

Google Translate (GT) is a very popular translation application due to its speed, ease of access, and support for more than 100 languages worldwide. GT utilizes Neutrak Machine Translation (NMT) technology to produce translations that take sentence context into account. (Rao Majji et al., 2024) However, when translating, GT produces literal translations and sometimes has difficulty capturing connotations and hidden cultural meanings. Meanwhile, DeepL is an advanced machine translation tool that uses artificial neural networks (ANN) to provide highly accurate translations in various languages. DeepL is known for its ability to produce translations that appear more natural than other translation tools. (Üyesi et al., 2024) However, DeepL also has difficulty handling Arabic with its complex morphological and contextual structure, such as Arabic words with polysemy (one word has many meanings) and idioms that depend on context. Both tools, GT and DeepL, play a major role in supporting Arabic language learning in the digital age, particularly in efforts to improve access, effectiveness, and quality of education in line with Sustainable Development Goal 4 (SDG 4) on quality and inclusive education.

Research by (Muthilillah et al., 2025) examining the translation of expressive Arabic speech acts in the Masameer animation shows that DeepL is more accurate in capturing literal meaning and religious expressions, while Google Translate produces translations that are more communicative but less semantically accurate. Both tools still face challenges in translating the connotative meaning and pragmatic context of Arabic, which is rich in cultural nuances.

Research by (Telaumbanua et al., 2024) shows that DeepL is superior in terms of meaning accuracy and grammatical structure, while Google Translate is faster and more accessible, but both are still limited in capturing cultural context and idiomatic meanings. Research by Paul Sebo and Sylvain de Lucia (2023) shows that Google Translate and DeepL both have high denotative equivalence and are able to preserve the connotative meaning of the source text. DeepL is more accurate semantically, while Google Translate is more communicative pragmatically.

Based on these findings, this study aims to analyze and compare the quality of Google Translate and DeepL translations in understanding Arabic narrative texts, focusing on three main linguistic aspects: meaning accuracy, grammatical structure, and comprehensibility. This study is expected to contribute to the use of AI technology in language learning, as well as support the achievement of Sustainable Development Goal 4 (SDG 4) through increased access to inclusive and innovative quality education.

## RESEARCH METHOD

This study uses a comparative qualitative approach to analyze the quality of translations produced by two artificial intelligence (AI)-based translation tools, namely Google Translate (GT) and DeepL Translate (DT), in understanding Arabic narrative texts. This

approach was chosen because it is capable of exploring linguistic and semantic phenomena in terms of translation quality and learners' and educators' perceptions of text meaning. In line with (Abdussamad, 2021) view, qualitative research focuses on understanding the meaning, context, and interpretation that arise from the participants' experiences, not merely on numerical representations. Thus, this approach is relevant for analyzing translation results that are rich in linguistic and cultural nuances. Furthermore, according to (Mohajan & Mohajan, 2022) qualitative comparative research allows researchers to examine differences in characteristics between research objects in depth and contextually, resulting in a more complete understanding of linguistic phenomena and translation technology.

The participants in this study consisted of 20 people, including 10 Arabic teachers and 10 Arabic students of varying levels. They were selected using purposive sampling, based on their expertise and experience in understanding Arabic. The selection of a combination of teachers and students was intended to obtain a balanced view between the perspectives of teachers and learners in assessing the quality of translation results. As explained by (Negou et al., 2023) the selection of participants in qualitative research is aimed at obtaining rich, meaningful, and contextual data so that researchers can understand the phenomenon more deeply from the participants' perspectives. This instrument consists of three components: (1) Arabic narrative texts taken from an eBook entitled "*Kisah-kisah Inspiratif*," (2) translations from Google Translate and DeepL Translate, (3) assessment sheets using a 1-5 Likert scale to assess three main linguistic aspects, namely accuracy of meaning, grammatical structure, and comprehensibility. These criteria refer to previous studies by (Telaumbanua et al., 2024) and (Muthilillah et al., 2025) which also used semantic, syntactic, and pragmatic aspects as the basis for assessing the quality of AI-based translations.

The research procedure was carried out in one week, with the following stages: participants were given an Arabic text along with GT and DT translations, then asked to assess both independently using the G-form assessment. After that, data analysis was carried out using a qualitative descriptive approach from the assessment sheets, and the average was calculated to see the assessment trends for each translation tool. Next, qualitative comments from participants were analyzed using content analysis techniques to identify and support arguments from previous research results. Thus, this method provides a comprehensive picture of the quality of AI translation, while supporting the Sustainable Development Goal 4 (SDG 4) in increasing access to inclusive and technology-based language education.

## RESULTS AND DISCUSSION

Before conducting further analysis, it is important to present the raw data from each participant as a basis for research transparency. The following table presents the individual assessment scores of all participants, both teachers and students, of the Arabic

narrative text translations by Google Translate and DeepL. Assessments were given using a 1–5 Likert scale for three main aspects, namely accuracy of meaning, sentence structure, and comprehensibility. Presenting the raw data allows readers to see the variation in ratings between participants, while also providing a basis for calculating average scores and conducting further analysis.

**Table 1.** Participant assessment scores from teachers

Participant (teacher)	Accuracy of Meaning		Grammatical Structure		Comprehensibility	
	GT	DT	GT	DT	GT	DT
1	4	5	4	4	4	5
2	3	4	3	4	2	4
3	4	4	4	4	4	4
4	5	5	5	5	5	5
5	3	4	3	4	3	4
6	3	4	4	4	3	5
7	3	3	2	4	3	4
8	4	3	4	3	4	4
9	3	4	4	3	4	5
10	4	4	4	4	4	5

**Table 2.** Participant assessment scores from students

Participant (student)	Accuracy of Meaning		Grammatical Structure		Comprehensibility	
	GT	DT	GT	DT	GT	DT
1	2	5	3	5	2	5
2	5	5	5	5	5	5
3	4	5	4	5	4	5
4	1	5	1	5	1	5
5	1	5	1	5	1	5
6	4	5	5	4	4	4
7	4	4	4	3	4	3
8	3	4	3	4	4	5

9	4	5	4	4	4	5
10	3	4	4	4	3	5

Tables 1 and 2 show the variation in scores between teachers and students, as well as differences in perceptions of GT and DT. From this individual data, the average score per component was calculated for each participant group.

**Table 3.** Average Scores Per Aspect and Participant Group

Component	GT - Teacher (Average)	DT - Teacher (Average)	GT - Student (Average)	DT -Student (Average)
Accuracy of meaning	3,6	4,0	3,1	4,7
Structure	3,7	3,9	3,4	4,4
Comprehensibility	3,6	4,5	3,2	4,7

Table 3 shows a comparison of average scores between Google Translate and DeepL, from both the perspective of teachers and students. This average presentation helps to see general patterns of preference for AI translation tools in terms of meaning accuracy, sentence structure, and comprehensibility. The results of the components tested are as follows: first, the average meaning accuracy of teacher participants on GT was 3.6 and DT 4.0, while student participants scored 3.1 and DT 4.7. From these assessments, teachers and students rated DT as slightly more accurate in meaning than GT. Furthermore, the average grammatical structure of teacher participants on GT was 3.7 and DT 3.9, while student participants on GT was 3.4 and DT 4.4, indicating that DT was better than GT. As for the average comprehensibility of teacher participants, GT scored 3.6 and DT scored 4.5, while student participants scored GT 3.2 and DT 4.7. Both scores indicate that DT is much easier to understand according to teachers and students.

**Table 4.** Sentence-By-Sentence Translation Analysis of Narrative Text

No	Arabic Text	Google Translate (GT)	DeepL (DT)	Comparative Analysis
1	كَانَ وَلَدٌ يَنْظُرُ الْفِيلَ فِي جُنَيْبَةِ الْحَيَوَانَاتِ	There was a boy looking at an elephant in the zoo.	A boy was watching an elephant in the zoo.	DT is more literal and formal in structure, while GT is more concise. Both translations are clear, but DT is more accurate in reflecting the original sentence structure and meaning.
2	فَمَدَّ يَدَهُ إِلَيْهِ بِالتُّفَّاحَةِ	He reached out to him with an apple.	He stretched out his hand toward it playfully.	DT retains the playful nuance and original subject matter more effectively. GT is simpler and more communicative but loses descriptive detail.
3	وَلَمَّا هَمَّ الْفِيلُ أَنْ يَأْخُذَهَا، فَبَضَّ الصَّبِيُّ يَدَهُ	When the elephant tried to take it, the boy pulled his hand back.	When the elephant was about to take it, the boy held his hand back.	DT is more formal and accurate in word choice. GT is more concise and easier to read. The differences are minor, but DT is more faithful to the original text.

4	ثُمَّ غَادَ وَمَدَّ يَدَهُ بِالتُّفَاحَةِ مَرَّةً ثَانِيَةً	Then he reached out again with the apple.	Then he came back and reached out his hand with the apple a second time.	DT maintains the sequence of actions more clearly. GT is more concise, while DT is more accurate in preserving narrative progression.
5	وَعَمِلَ كَمَا عَمِلَ أَوَّلَ مَرَّةٍ	He did the same thing he did the first time.	Doing the same thing the first time.	DT emphasizes sequence and literalness. GT simplifies the expression but remains clear. DT is more accurate overall.
6	فَغَضِبَ الْفِيلُ	The elephant got angry.	The elephant was angry.	Minor lexical differences appear. DT sounds slightly more formal and natural, while GT is communicative. DT is marginally more accurate in meaning.
7	وَلَكِنَّهُ صَبَرَ عَلَى الصَّبِيِّ حَتَّى سَهَا عَنْهُ	But he was patient with the boy until he forgot about it.	But he was patient with the boy until he forgot about him.	DT is more literal and consistent with the original reference to the boy. GT simplifies the structure slightly.
8	وَمَدَّ خُرْطُومَهُ وَخَطَفَ طَرْبُوشَهُ	He stretched out his trunk and snatched his tarboosh.	He stretched out his trunk and grabbed his turban.	DT uses a more familiar term ("turban"), making it accessible to wider readerships. GT retains the original cultural term. Both remain accurate.

9	فَرَزَعَقَ الْوَلَدُ وَبَكَى	The boy screamed and cried.	The boy screamed and cried.	Both translations are identical, accurate, and easy to understand.
10	وَعَلِمَ أَنَّ الَّذِي يَفْعَلُ السَّرَّ يَلْقَى السَّرَّ	He knew that those who do evil will meet with evil.	And he knew that whoever does evil will encounter evil.	DT is more formal and precise in conveying the moral lesson. GT is communicative, but DT better fits the didactic tone of the narrative.

Table 4 illustrates the differences in translation quality more concretely. Table 4 displays several selected sentences from the narrative text, the translations produced by each tool, a brief analysis, and comparative comments from participants. The comparison shows that Google Translate (GT) tends to produce translations that are more natural, communicative, and easy to understand, while DeepL (DT) is more literal, formal, and structurally accurate. In simple sentences, the results of both are almost the same. Thus, GT is superior in terms of comprehensibility, while DT is stronger in terms of structural accuracy and fidelity to the source text.

Based on the results of the analysis in Tables 1-4, it can be seen that DeepL Translate (DT) is superior and shows that the translation quality is better in two components, namely in terms of structure and accuracy of meaning, while Google Translate (GT) is superior in terms of comprehensibility. Both tools have different advantages, namely DT is more consistent in its fidelity to the Arabic source text, while GT is more communicative for general readers.

These results are in line with previous research by (Yulianto & Supriatnaningsih, 2021), who said that DeepL produces more accurate, natural, and readable translations than Google Translate, even though the difference is not statistically significant. Therefore, DeepL can be said to be qualitatively superior, while GT is still competitive in terms of the closeness of the text structure to the source. Based on the results of the study (Desry Noviany et al., 2024), DeepL Translate performed better than Google Translate in translating legal documents, particularly in terms of accuracy and acceptability. However, neither translation engine is yet adequate for independent use in the context of legal translation, and should continue to be used as an additional tool for human translators.

In contrast, (Freskila & Jayantini, 2025) say that Google Translate excels at producing more pragmatic and connotative translations, making them more contextual and appealing to readers, especially in the field of tourism. Meanwhile, DeepL Translate maintains its strength in formal structure and grammatical accuracy. Thus, GT can be said

to be functionally superior, while DT is structurally superior. Research by (Sihombing et al., 2025) shows that Google Translate is superior in terms of translation speed and support for more languages, while DeepL offers better context understanding and produces more natural translations. DeepL achieved an average accuracy score of 2.8, which is slightly higher than Google Translate's 2.6, highlighting its strength in capturing linguistic nuances, especially in complex texts.

In the context of Quality Education (SDG 4), these findings have important implications. The use of AI-based translation tools such as DeepL and Google Translate can be an innovative learning medium that supports digital literacy skills and cross-language comprehension. Teachers and students can use both tools complementarily: DeepL for more academic or formal text analysis due to its structural precision, and Google Translate for quick comprehension and communicative exercises. (Lazuwardiyah & Setiawan, 2024). Thus, the use of this technology not only improves the effectiveness of language learning, but also supports the achievement of inclusive, adaptive, and technology-based educational goals as mandated by the SDGs.

## CONCLUSION

This study concludes that Artificial Intelligence (AI)-based translation tools, namely Google Translate (GT) and DeepL Translate (DT), contribute significantly to Arabic language learning and the achievement of Sustainable Development Goal 4 (Quality Education) by bridging linguistic barriers and promoting accessible learning. Based on the framework, the findings indicate that DeepL Translate outperforms Google Translate in terms of meaning accuracy and structural clarity, while Google Translate excels in understandability accuracy, making it more effective for educational and contextual translation tasks.

However, both tools remain valuable in supporting inclusive and innovative learning environments if used critically. Therefore, educators and students are encouraged to utilize AI translation tools as complementary learning aids, rather than as a complete replacement for human understanding. Educational institutions should also strengthen AI literacy to ensure responsible use, while developers and researchers are advised to improve translation accuracy for context-sensitive languages like Arabic in the future. Overall, integrating AI translation into education can advance linguistic competence and promote equitable, technology-based learning, in line with SDG 4.

## REFERENCES

- Abdussamad, Z. (2021). *Metode Penelitian Kualitatif* (P. Rappana, Ed.). CV. Syakir Media Press.
- Desry Noviany, H., Erlanda Edel, E., Nuruz Zaman, M., Negeri Jakarta, P., Siwabessy, J. G., & Baru, K. (2024). *Analisis Perbandingan Kesalahan Penerjemahan Google Translate dan Deepl dalam Kualitas Penerjemahan Teks Hukum*. <https://beritanegara.co.id/>

- Freskila, E. P., & Jayantini, I. G. A. S. R. (2025). Translation Equivalence of Tourism Website Content: A Comparison Between Google Translate and DeepL Translate. *Indonesian Journal of EFL and Linguistics*, 1-20.  
<https://doi.org/10.21462/ijefl.v10i1.835>
- Lazuwardiyah, F., & Setiawan, S. (2024). ENGLISH EDUCATION The Utilization of Artificial Intelligence (AI) Translators to Improve Writing in EFL Classroom. In *Journal for Teaching and Learning* (Vol. 12).  
<https://jurnal.uinsyahada.ac.id/index.php/>
- Maulidiya, A. R., Abdurrahman, M., & Saleh, N. (2025). Exploring AI Capabilities in Arabic Grammar: Comparative Analysis of ChatGPT and Gemini. *Arabiyat: Jurnal Pendidikan Bahasa Arab Dan Kebahasaaraban*, 11(2), 160-174.  
<https://doi.org/10.15408/a.v11i2.42671>
- Mohajan, D., & Mohajan, H. K. (2022). Exploration of Coding in Qualitative Data Analysis: Grounded Theory Perspective. *Research and Advances in Education*, 1(6), 50-60. <https://doi.org/10.56397/rae.2022.12.07>
- Muthilillah, A. Z., Sopian, A., & Supriadi, R. (2025). Comparing DeepL and Google Translate in Translating Arabic Expressive Speech Acts in Masameer: al-Film. *Anaphora: Journal of Language, Literary, and Cultural Studies*, 8(1), 48-66.  
<https://doi.org/10.30996/anaphora.v8i1.132082>
- Negou, E., Nkenganyi Fonkem, M., Suh Abenwi, J., & Ibrahima. (2023). Qualitative Research Methodology in Social Sciences. *International Journal of Scientific Research and Management (IJSRM)*, 11(09), 1431-1445.  
<https://doi.org/10.18535/ijstrm/v11i09.sh01>
- Nuralan, S. (2025). *Transformasi Pendidikan di Era Artificial Intelligence (AI)*.
- Rao Majji, E., Babu Ampalam, S., & Raju, C. (2024). Evaluating Google Translate as an AI-Powered Translation Tool: Strengths, Limitations, and Implications for Multilingual Communication. In *Journal of Information Systems Engineering and Management* (Vol. 2025, Issue 54s). <https://www.jisem-journal.com/>
- Rosita, Made Sujana, I., & Jaswadi Putera, L. (2024). A Comparative Analysis Of Accuracy Between Google Translate and Bard IN Translating Abstract Of Scientific Journal.  
<https://ejournal.uin->
- Sihombing, M., Permata, K., Manalu, S., Tampubolon, N., & Sinurat, B. (2025). A Literature-Based Comparative Analysis of DeepL and Google Translate: Strengths and Limitations in English Translation.
- Telaumbanua, Y. A., Marpaung, A., Putri, C., Gulo, D., Kardo, D., Waruwu, W., Zalukhu, E., & Zai, N. P. (2024). *Journal of Artificial Intelligence and Engineering Applications An Analysis of Two Translation Applications: Why is DeepL Translate more accurate than Google Translate?* (Vol. 4, Issue 1). <https://ioinformatic.org/>
- Üyesi, Ö., Çetin, Ö., Üniversitesi, A., & Duran, A. (2024). A Comparative Analysis Of The Performances Of ChatGPT, DeepL, Google Translate And A Human Translator in Community Based Settings.
- Woo, J. H., & Choi, H. (2021). *Systematic Review for AI-based Language Learning Tools*.

- Yulianto, A., & Supriatnaningsih, R. (2021). Google Translate vs. DeepL: A quantitative evaluation of close-language pair translation (French to English). *AJELP: The Asian Journal of English Language and Pedagogy*, 9(2), 109-127. <https://doi.org/10.37134/ajelp.vol9.2.9.2021>
- Zafar, T. Z., & Komilovich, E. U. (2024). Challenges of Arabic - English Translation. *Pubmedia Jurnal Pendidikan Bahasa Inggris*, 1(3), 9. <https://doi.org/10.47134/jpbi.v1i3.616>