ISSN: 3063-4938

CHATGPT VS DEEPSEEK: TRANSLATION TECHNIQUES AND READABILITY COMPARISON OF FOOTBALL NEWS HEADLINES

Vicky Vadila Muhti[1], Elitaria Bestri Agustina Siregar[2]
[1,2]Politeknik Negeri Jakarta, Depok, Indonesia
✉e-mail: vicky.vadila.muhti.an22@mhsw.pnj.ac.id[1]
elitaria.agustina@elektro.pnj.ac.id[2]

ABSTRACT
The rapid development of artificial intelligence (AI) technology has brought significant changes to various aspects, including translation. AI systems such as ChatGPT and DeepSeek are now capable of translating various linguistic units, including headlines particularly in sports news headlines such as football, where the language is typically concise, dense, and often incorporates idiomatic expressions and metaphors. However, the quality of these translations specifically their readability, still needs to be evaluated. This study aims to analyze translation techniques according to Molina & Albir (2002), and to evaluate and compare the readability of football news headlines translations generated by ChatGPT and DeepSeek according to Nababan et al. (2012). This study employs a qualitative descriptive method with a comparative approach. The data used consists of 100 football news headlines from the bbc.com. The results show that the most frequently used translation technique by ChatGPT is amplification (38.5%), while DeepSeek's is borrowing (30%). In terms of readability, ChatGPT outperforms DeepSeek with an average score of 2.26 compared to DeepSeek's average score of 2.21. In conclusion, ChatGPT excels in terms of clarity by adding additional information not present in the target language, while DeepSeek excels in terms of conciseness because it adheres to the characteristics of short, concise headlines.

Keywords: *ChatGPT, DeepSeek, Translation Techniques, Readability, Headline*

INTRODUCTION
The development of artificial intelligence (AI) has significantly transformed various fields, including translation. AI-based translation systems such as ChatGPT and DeepSeek are now capable of producing translations quickly and efficiently by utilizing Natural Language Processing (NLP) technologies (Peng et al., 2023; Saehudin et al., 2025). The presence of these technologies not only facilitates cross-language communication but is also increasingly used in media content production, including the translation of international news (Zou et al., 2025). This trend highlights how AI has become an integral part of the modern translation industry.

In the context of translation, quality is not solely determined by accuracy but also by readability. This aspect plays a crucial role, as it determines how easily the translated text can be understood by the target audience in a clear and natural way, without ambiguity (Nababan et al., 2012). They also emphasize that readability is closely related to sentence structure, word choice, and the naturalness of expression in the target language. In AI-based translation, a major challenge lies in how these systems can produce translations that are not only accurate in meaning but also highly readable and aligned with the norms of the target language (Wang, 2023). This issue becomes increasingly significant when translation involves texts that demand both brevity and stylistic nuance, such as news headlines.

Beyond readability, translation quality is also shaped by the techniques employed in rendering the source text into the target language. Translation techniques refer to specific procedures used to transfer meaning at the word, phrase, or sentence level (Molina & Albir, 2002). The selection of translation technique directly determines whether a translation preserves the cultural nuance, stylistic tone, and pragmatic intent of the source text (Hasibuan, 2025). However, in the domain of AI-generated translation, the application of these techniques remains largely unexamined, particularly when applied to linguistically dense text types such as news headlines.

This aspect becomes even more important in translating news headlines, particularly in sports news such as football, where the language is typically concise, dense, and often incorporates idiomatic expressions, metaphors, and distinctive journalistic styles (Bell, 1991; Anusitviwat et al., 2025). As the entry point of information, headlines are designed to capture readers' attention while conveying the core message (Sartika & Badri, 2019). News headlines serve as gateways to information, so errors in translation can lead to misunderstandings among readers (Dewi et al., 2021).

This challenge can be more clearly observed through concrete examples of AI-generated translations. The linguistic character of football news headlines makes this challenge particularly concrete. For example, the headline "Man City smell blood - why Arsenal should fear title rivals" a headline in which the phrase 'smell blood' carries a metaphorical meaning describing a situation where a team senses its opponent is weak and the opportunity to win the match is wide open. Using the prompt "*Terjemahkan headline di atas ke dalam Bahasa Indonesia,"* ChatGPT produces *"Manchester City mencium peluang — mengapa Arsenal harus mewaspadai rival perebutan gelar*", while DeepSeek translates it as "*Man City mencium bau darah - mengapa Arsenal harus takut pada saingan gelar*". From a translation technique perspective, ChatGPT appears to use modulation, shifting the figurative expression 'smell blood' into a more contextually adapted equivalent '*mencium peluang*', while DeepSeek use a more literal technique by retaining the original imagery '*mencium bau darah*'. These differing technical choices consequently produce translations that diverge not only in stylistic tone but also in their degree of readability for Indonesian readers. The translations produced by the two AI systems for the phrase "smell blood" show quite obvious differences in terms of structure, readability, and naturalness, raising the question of which translation is more readily understood by Indonesian readers, and by how much. This contrast shows the kind of readability gap that AI-based translation of football news headlines can produce. Therefore, this example illustrates how differences in linguistic choices can directly affect the readability of AI-generated translations.

Given this issue, a number of studies have attempted to evaluate the readability of AI-generated translations, particularly those produced ChatGPT and DeepSeek. A study by Yu et al. (2026) found that both models show relatively similar levels of accuracy. However, ChatGPT tends to produce longer and more complex sentences than DeepSeek, which lead to lower readability. Meanwhile, Sari et al. (2025) reported significant differences in readability levels between ChatGPT and DeepSeek, measured using the Flesch Reading Ease index, with the results varying depending on the structure and purpose of the text. Furthermore, Wang and Yu (2025) reported that the readability of outputs generated by ChatGPT and DeepSeek varies depending on language complexity and the domain of the text.

In the context of news headlines, a study by Roselynn et al. (2025), which analyzed machine-translated headlines using systems such as Bing Translator and DeepL, found that readability is strongly influenced by the system's ability to maintain clarity and natural expression within a limited and concise structure. The findings also indicate that although machine translation can produce translatios that are generally informative, readability remains a major challenge. This is largely due to the linguistic characteristics of headlines, which are typically dense, brief, and often contain implicit meanings.

Most previous studies have focused on evaluating translation quality in general terms such as accuracy, acceptability, or readability without specifically comparing the performance of different AI platforms like ChatGPT and DeepSeek within particular text types. In addition, relatively few studies have examined translation at the level of news headlines, which have distinct linguistic characteristics: they are brief, compact, and often rely on metaphorical or idiomatic expressions. Another limitation is the lack of research addressing specific domains such as sports especially football which has its own terminology and writing style.

Based on these gaps, there is a clear need for a study that specifically compares the readability of translations produced by ChatGPT and DeepSeek in the context of football news headlines. Therefore, this study aims to provide a more in-depth analysis of the readability levels of sport news headlines translations generated by both ChatGPT and DeepSeek as well as contributes to the development of translation studies, particularly in the evaluation of AI-based translation quality. The findings are expected to provide new insights into the relative performance of ChatGPT and DeepSeek in producing translations that are easy for Indonesian readers to understand. In addition, this study may serve as a useful reference for media practitioners and translators in making more effective use of AI technologies.

METHODS

This study employs a qualitative descriptive method with a comparative approach to examine translation technique and readability of English-to-Indonesian translations of football news headlines produced by two AI technologies, ChatGPT and DeepSeek. The data consist of 100 English-language football headlines from *bbc.com*, published between March 24 and April 17, 2026, selected through purposive sampling based on recency of publication and high reader engagement (approximately 400 - 1,000 comments per article).The topic of football was chosen for this study because Indonesia is one of the largest football fan base in the world (Annur, 2022; Medy, 2025). Both AI models were given the same prompt to translate each headline, producing 100 pairs of translations per model, which were then analyzed using Krippendorff's (2004) content analysis method, drawing on Molina and Albir's translation technique framework and Nababan et al.'s translation quality (readability) assessment model.

Data collection proceeded through several steps: gathering recent football headlines from *bbc.com*, compiling them into a table, translating them via ChatGPT and DeepSeek using the same prompt, analyzing the translation techniques applied, and evaluating readability through three raters. The raters were purposively selected based on strong English and Indonesian reading comprehension (two held TOEFL ITP scores above 550), familiarity with football terminology as regular readers of bbc.com football news, and one being a practitioner lecturer specializing in journalistic translation. Each rater independently scored the translations using Nababan et al.'s scale: 3 (high readability), 2 (medium readability), and 1 (low readability), after which a Focus Group Discussion (FGD) was held to align the raters' perceptions and strengthen the reliability of the findings.

To ensure data validity, the study applied data source triangulation which combining primary data from headline analysis with secondary data from rater assessments and prior studies and theoretical triangulation (integrating Molina and Albir's translation technique theory with Nababan et al.'s readability assessment theory). Data analysis followed Miles and Huberman's (2014) three-stage model: data reduction (selecting the 100 headlines with the highest reader engagement), data display (organizing translation techniques and readability scores from both AI systems into tables), and conclusion drawing (interpreting the resulting patterns to determine the readability category (high, medium, or low) of the translations produced by ChatGPT and DeepSeek).

RESULT AND DISCUSSION

Result

Based on the analysis of 100 football news headlines from bbc.com, seven out of the eighteen translation techniques proposed by Molina & Albir (2002) were found in ChatGPT's translations, while nine out of eighteen were found in DeepSeek.

Table 1. Result of Translation Techniques (ChatGPT)

Translation Technique	Frequency	Percentage
Amplification	77	38.5%
Established Equivalent	63	31.5%
Modulation	24	12%
Borrowing	20	10%
Transposition	8	4%
Adaptation	4	2%
Literal	4	2%
Total	200	100%

As shown in Table 1, ChatGPT produced 200 instances of technique use in total, dominated by amplification (77 data, 38.5%) and established equivalent (63 data, 31.5%), followed by modulation (24 data, 12%), borrowing (20 data, 10%), transposition (8 data, 4%), and adaptation and literal, each with the lowest frequency of 4 data (2%).

Table 2. Result of Translation Technique (DeepSeek)

Translation Technique	Frequency	Percentage
Borrowing	60	30%
Established Equivalent	54	27%
Literal	33	16.5%
Transposition	25	12.5%
Modulation	22	11%
Amplification	3	1.5%
Calque	2	1%
Generalization	1	0.5%
Description	1	0.5%
Total	200	100%

In contrast, Table 2 shows that DeepSeek's 200 instances were led by borrowing (60 data, 30%) and established equivalent (54 data, 27%), followed by literal (33 data, 16.5%), transposition (25 data, 12.5%), modulation (22 data, 11%), amplification (3 data, 1.5%), calque (2 data, 1%), and generalization and description, each with the lowest frequency of 1 data (0.5%). These frequencies and percentages, calculated using the formula *(total occurrences of a technique ÷ total occurrences) × 100%*, indicate that ChatGPT consistently

favors techniques that expand and clarify meaning, whereas DeepSeek favors techniques that preserve conciseness and the similarity of the headline's style.

Table 3. Result of Readability Score of ChatGPT & DeepSeek

AI	Rater 1	Rater 2	Rater 3	Final Average Score
ChatGPT	2.63	2.65	1.51	2.26
DeepSeek	2.50	2.45	1.69	2.21

Regarding readability, table 3 presents the average final scores given by three raters on a 1–3 scale proposed by Nababan et al. (2012). The table shows the average final scores for both AI models. ChatGPT achieved an average total score of 2.26, while DeepSeek achieved an average total score of 2.21. Both scores fall within the medium range on the scale used, indicating that in general, the translation results from both AI models have not yet reached the highest level of readability. The difference between the two is very small, only 0.05 points, so the readability quality of ChatGPT and DeepSeek translations is at a nearly equivalent level.

Substantive differences were actually observed at the rater level: ChatGPT received more "Readable" ratings (Raters 1 and 2), whereas DeepSeek was more consistent in the "Fairly Readable" category across all raters except Rater 1. This is due to several factors. Rater 1 and Rater 2 focused more on the ease of understanding the meaning. Neither of them was overly concerned about ChatGPT's lengthy translations or the literal translations from both AIs as long as the headline's content remained easy to understand, so they tended to give scores of 2 or 3. Conversely, Rater 3 paid more attention to alignment with the characteristics of news headlines which are short, concise, and natural. Therefore, ChatGPT's overly long translations and the overly literal translations from both AI systems are considered less effective and less comfortable to read, resulting in lower scores. Consequently, the differences in scores among raters in this study can be understood as a consequence of each rater's differing evaluation focus regarding the readability of the headlines.

Discussion

Based on the results of the translation technique analysis above, it can be concluded that the dominant technique used by ChatGPT in translating football news headlines is the amplification technique. This is consistent with research conducted by Framesthia et al. (2024) which showed that ChatGPT uses the amplification technique when translating Arabic articles into Indonesian. This means that for both headlines. Meanwhile, DeepSeek's dominant technique, borrowing, has not been explicitly identified in previous research, which means this finding extends the existing discussion on AI translation techniques at large. or articles, ChatGPT consistently uses the amplification technique when translating these units.

ChatGPT's reliance on amplification is most visible in its tendency to clarified abbreviations, expanded names for the target reader, and added information that was not present in the source language, for instance expanding "PSG" into "Paris Saint-Germain" and "Dembele" into "Ousmane Dembélé", or "Romero" into "Cristian Romero" "Tottenham" into "Tottenham Hotspur", "double" into "*dua gol*". Raters judged this pattern as making ChatGPT's translations clear and informative, since readers no longer need to guess the identities or context referred to in the headline. However, this same technique is also the source of ChatGPT's main weakness: by adding explanatory words, ChatGPT's translations often become longer than a headline should be, reducing their conciseness and, in some cases, their readability. On the other hand, On the other hand, DeepSeek more often maintains a concise style in its translations and borrows words or retains player and club names from the source

language, such as “Iran”, “Infantino”, “Romero”, and others so that the translations fit the short and concise style of headline. However, its translations sometimes sacrifice clarity and ambigious such as “double” into “*ganda*” in the context of football.

These patterns directly explain the readability results reported in Table 3. ChatGPT’s explicit and detailed translations reduce the interpretive burden on readers. This is why ChatGPT received a higher overall score of 2.26 compared to DeepSeek’s 2.21. Thus, it can be concluded that in aspects of readability, ChatGPT outperforms DeepSeek in translating football news headlines from English to Indonesian. This aligns with study by Hidayatullah & Aliurridha (2025), whose findings indicate that ChatGPT outperforms other AI models such as Gemini in aspect of readability across genres. Therefore, ChatGPT and DeepSeek can be utilized as tools to assist in translating English-language football news headlines into Indonesian, but both AI still require human translators to refine the translations so they can be easily understood by readers.

Overall, the translation techniques chosen by ChatGPT and DeepSeek affect the readability of headline translations. This aligns with the study by Normalita & Nugroho (2023) which states that the use of translation techniques affects the readability of headline translations. Although the study by Normalita & Nugroho (2023) examined headline translation techniques based on human translations, it can be understood that for both human and AI translations, the selection of specific techniques can affect readability. Therefore, the analysis of translation techniques from this study can serve as a basis for reviewing and improving AI translation algorithms to produce headline translations that are more effective and easier for the target audience to understand.

CONCLUSION

Based on an analysis of 100 football news headlines from *bbc.com*, it can be concluded that ChatGPT employ seven out of eighteen translation techniques with the dominant techniques is amplification. This dominant technique is ChatGPT’s key characteristic in translating football news headlines. On the other hand, DeepSeek employ nine out of eighten translation techniques with the dominant techniques is borrowing. This dominant technique is DeepSeek’s key characteristic in translating football news headlines. Overall, it can be conclude that ChatGPT tends to produce translations that are more explicit and informative by adding details. On the other hand, DeepSeek produces translations that are more concise, succinct, and closer to the style of the original headline. In aspects of readability, ChatGPT exceled DeepSeek in translating football news headlines from English to Indonesian. Nevertheless, both AI models have their own strengths: ChatGPT performs better in terms of clarity, while DeepSeek performs better in terms of conciseness and the similarity of the headline’s style. However, ChatGPT and DeepSeek’s ability to translate football news headlines is still not perfect, so a human translator is needed to refine the translations so that they are easy for readers to understand.

REFERENCE

Anusitviwat, C., Suwannaphisit, S., Bvonpanttarananon, J., & Tangtrakulwanich, B. (2025). Comparing ChatGPT and DeepSeek for Assessment of Multiple-Choice Questions in Orthopedic Medical Education: Cross-Sectional Study. *JMIR Formative Research*, *9*, 1–7. https://doi.org/10.2196/75607

Bell, A. (1991). *The Language of News Media*. https://smoschon.ntlab.gr/archives/courses/mda0405/notes/Bell_Media_and_Language.pdf

Dewi, N. P. A. K. S., Martini, N. K., & Suardana, I. wayan. (2021). Pilihan Kata Pada Penerjemahan Judul Berita. *Prosiding Seminar Nasional Linguistik Dan Sastra (SEMNALISA)*, 171–177. https://e-journal.unmas.ac.id/index.php/semnalisa/article/view/2340

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*. http://arxiv.org/abs/2301.08745

Krippendorff, klaus. (2004). *Content analysis : an introduction to its methodology*. https://archive.org/details/contentanalysisi0000krip_y7k5

Matthew B. Miles, A. Michael Huberman, J. S. (2014). Qualitative Data Analysis A Methods Sourcebook. *Experiencing Citizenship: Concepts and Models for Service-Learning in Political Science*, 109–118. https://www.metodos.work/wp-content/uploads/2024/01/Qualitative-Data-Analysis.pdf

Molina, L., & Albir, A. H. (2002). Translation techniques revisited: A dynamic and functionalist approach. *Meta*, *47*(4), 498–512. https://doi.org/10.7202/008033ar

Nababan, Mangatur Nuraieni, Ardiani, S. (2012). *Pengembangan Model Penilaian Kualitas Penerjemahan* (Vol. 24). https://doi.org/10.2307/416501

Normalita, A., & Nugroho, R. A. (2023). *Changing the "Body" of BBC News: a Study of News Headlines Translation Techniques*. *Colalite*, 58–65. https://doi.org/10.2991/978-2-38476-140-1_7

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/10.18653/v1/2023.findings-emnlp.373

Roselynn, V. A., Bowo, T. A., & Sigalingging, B. M. (2025). Comparative study of news headlines using Bing translator and DeepL. *Journal of English Literature, Language, and Culture*, *7*(1), 1–7. https://journal.unpak.ac.id/index.php/albion/article/viewFile/11571/5685

Saehudin, A., Syahputri, A., Maharani, A. A., & Julia, D. (2025). *Kualitas Terjemahan Artificial Intelligence : Studi Komparatif Keberterimaan Terjemahan Deepseek AI dan Claude AI pada Teks Akademik Berbahasa Arab*. 349–368. https://conferences.uinsaid.ac.id/iccl/paper/view/768

Sari, Fulden. Celik, Z. M. Y. (2025). *ChatGPT-4 vs. DeepSeek-V3: a comparative study of response quality, reliability, usefulness, and readability for exercise and rehabilitation strategies in patients with ankylosing spondylitis*. https://pubmed.ncbi.nlm.nih.gov/41207975/

Sartika, P. L., & Badri, M. (2019). Proses Penentuan Headline Di Halaman Metropolis Riau Pos. *Jurnal Riset Mahasiswa Dakwah Dan Komunikasi(JRMDK)*, *1*(5), 301–312. https://ejournal.uin-suska.ac.id/index.php/jrmdk/article/view/8744

Wang, S., & Yu, Y. (2025). A Comparative Analysis of the Readability and Information Quality of the Chinese and English Versions of Educational Materials for Thoracic Surgery Patients Generated by DeepSeek, Grok-3 and ChatGPT. *Asia Pacific Journal of Clinical Medical Research*, *1*(4), 1–7.

https://doi.org/10.62177/apjcmr.v1i4.731
Yu, C., Fan, J., Chen, Y., Shen, W., Zhang, Y., Li, L., Wen, J., & Chen, X. (2026). A comparative study of ChatGPT 4o and DeepSeek in addressing CIED infection-related questions: Accuracy and readability assessment. *Medicine*, *105*(5), e47493. https://doi.org/10.1097/MD.0000000000047493
Zhang, Z., Nurulakla Syed Abdullah, S., Alif Redzuan Abdullah, M., & Duan, W. (2025). Translation Quality in an Evolving Paradigm: Neural Machine Translation and Large Language Models in Technical Domains. *International Journal of English Language Education*, *13*(2), 52. https://doi.org/10.5296/ijele.v13i2.23057
Zou, L., Li, K., Lamerton, J., & Mirzapour, M. (2025). GenAIese - A Comprehensive Comparison of GPT-4o and DeepSeek-V3 for English-to-Chinese Academic Translation. *Proceedings of the Eleventh Workshop on Patent and Scientific Literature Translation (PSLT 2025)*, *Pslt*, 1–12. https://aclanthology.org/2025.pslt-1.1/
Framesthia, L. M., Khoerunnisa, S., Al Izzah, S., Al Farisi, M. Z., & Supriadi, R. (2024). Analisis perbandingan teknik penerjemahan Arab-Indonesia pada Google Translate dan ChatGPT (Kajian sintaksis: Kalimat verba dan nomina). Ihtimam: Jurnal Pendidikan Bahasa Arab, 7(2), 114–127. https://journal.staimsyk.ac.id/index.php/ihtimam/article/view/716/369
Hidayatullah, N. T., & Aliurridha. (2025). Evaluasi komparatif kualitas terjemahan AI: Gemini vs ChatGPT pada novel *The Lord of the Rings* (*The Two Towers*). *Jurnal Lisdaya*, *21*(1), 75–82. https://lisdaya.unram.ac.id/index.php/lisdaya/article/view/123
Hasibuan, Z. (2025). A study of accuracy and translation technique in narrative text by using Google Translate application. *JETLEE: Journal of English Language Teaching, Linguistics, and Literature*, *5*(1), 47–63. https://doi.org/10.47766/jetlee.v5i1.4613
Annur, M, C. (2022). *Survei Ipsos: Indonesia Punya Penggemar Sepak Bola Terbesar di Dunia*.databoks.katadata.co.id. https://databoks.katadata.co.id/olahraga/statistik/522de0f585f0915/survei-ipsos-indonesia-punya-penggemar-sepak-bola-terbesar-di-dunia
Medi, R. (2025). *Indonesia Jadi Negara dengan Jumlah Penggemar Sepak Bola Terbesar Ke-3 di Dunia*. goodstats.id. https://data.goodstats.id/statistic/indonesia-jadi-negara-dengan-jumlah-penggemar-sepak-bola-terbesar-ke-3-di-dunia-zhi21